

Information Technology  
Solutions

---

---

---

---

---

# Managing LANs

---

DATAPRO

# Managing LANs

---

President and Publisher  
**Carl G. Tobiasen**

Vice President  
Information Services  
**Patricia A. Lauletta**

Managing Analyst  
**Joseph F. Kelly**

Analyst  
**Robert W. McCombe**

Product Manager  
**Gerald J. Arcuri**

Director  
Customer Service  
**Kathleen C. Patto**

---

## DATAPRO

Information Services  
Group

600 Delran Parkway  
P.O. Box 1066  
Delran, New Jersey 08075  
USA

Tel: (800) 328-2776  
Tel: (609) 764-0100  
Fax: (609) 764-2814

### Australia:

Datapro International  
Level 5  
3 Spring Street  
Sydney NSW 2000  
Australia  
Tel: (61) 2 2522311  
Fax: (61) 2 2477992

### Canada:

Datapro International  
270 Yorkland Boulevard  
North York, Ontario  
M2J 1R8 Canada  
Tel: (800) 668-9308  
Tel: (416) 496-3131  
Fax: (416) 496-3160

### Frankfurt:

Datapro International  
Liebigstrasse 19  
D-6000 Frankfurt 1  
Germany  
Tel: (49) 69 714070  
Fax: (49) 69 71407146

### Hong Kong:

Datapro International  
16/F Unit C  
Cindic Tower  
128 Gloucester Road  
Wanchai, Hong Kong  
Tel: (852) 8022040  
Fax: (852) 8360697

### Japan:

Nikkei Business Publications  
1-14-6, Uchikanda  
Chiyoda-ku, Tokyo 101  
Japan  
Tel: (81) 03-233-8081  
Fax: (81) 03-233-3048

### Paris:

Datapro International  
128 rue du Faubourg Saint-Honore  
F-75008 Paris  
France  
Tel: (33) 1 42890381  
Fax: (33) 1 42890400

### Singapore:

Datapro International  
Unit 05-03 Dapenso Building  
158 Cecil Street  
Singapore 0106  
Tel: (65) 2225091  
Fax: (65) 2213321

### South America:

MIPS  
Rua Camargo 12  
Ossio-Sao Paulo-SP  
Brazil  
Tel: (55) 11 813 7492  
Fax: (55) 11 813 7476  
814 4363

### United Kingdom:

Datapro International  
McGraw-Hill House  
Shoppenhangers Road  
Maidenhead  
Berkshire SL6 2QL Great Britain  
Tel: (44) 628 773277  
Fax: (44) 628 773628

---

Copyright © 1992 by McGraw-Hill, Incorporated. All rights reserved. Printed in the United States of America. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a data base retrieval system, without the prior written permission of the publisher.

Information has been obtained by Datapro Information Services Group from sources believed to be reliable. However, because of the possibility of human or mechanical error by our sources, Datapro Information Services Group, or others, Datapro Information Services Group does not guarantee the accuracy, adequacy, or completeness of any information and is not responsible for any errors or omissions or for the results obtained from use of such information.

## Managing LANs

(Formerly Datapro Management of Microcomputer Systems)

---

Report Number	Report
---------------	--------

---

### Selection & Acquisition

- |      |   |
|------|---|
| 5010 | Criteria for LAN Selection                    |
| 5020 | Choosing a NOS: Banyan, Microsoft, and Novell |
| 5034 | Selecting a 486                               |
| 5036 | Selecting Mass Storage Devices                |
| 5050 | Contracting Fundamentals                      |
| 5054 | Evaluating Potential Consultants              |
| 5058 | Buying Third-Party Network Support            |

### Installation & Maintenance

- |      |  |
|------|--|
| 5410 | Installing a LAN                                 |
| 5412 | Your First LAN: Do It Yourself?                  |
| 5413 | Ethernet Wiring Made Simple                      |
| 5415 | A Guide to Installing Windows on a NetWare LAN   |
| 5420 | Installing Modems and Software                   |
| 5445 | Turning an Older PC into a 386 CPU-Based Machine |
| 5450 | PC Maintenance Guidelines                        |

### Operations Management

- |      |  |
|------|--|
| 5610 | A Guide to Keeping LANs Running Smoothly                       |
| 5611 | Control Change From the Ground Up                              |
| 5612 | LAN Administration: A New Job Function                         |
| 5613 | LANs by the Book   |
| 5615 | Training Network Users   |
| 5618 | LAN Training Sources   |
| 5620 | LAN Management Issues  |
| 5622 | Establishing a LAN Dossier                                     |
| 5625 | Network Analysis: Ten Steps to Fine-Tuning Network Performance |
| 5630 | Fault-Tolerant LANs  |
| 5640 | The LAN Diagnostic Process                                     |
| 5650 | Managing LAN-Based Services                                    |
| 5660 | Server Backup  |

### Security

- |      |  |
|------|--|
| 5810 | Planning for Microcomputer Security                |
| 5820 | Developing Microcomputer Security Awareness        |
| 5822 | A Sample Microcomputer Security Policy             |
| 5830 | Microcomputer Data Security Solutions              |
| 5835 | Security and Integrity in Micro-Mainframe Networks |
| 5840 | Protecting Against Computer Viruses                |
| 5842 | LAN Security                                       |
| 5845 | Developing a LAN Virus Protection Strategy         |
| 5850 | Application Layer Network Security                 |
| 5860 | Microcomputer Encryption and Access Control        |

---

### Technology Overviews

#### Local Area Networks

- |      |   |
|------|---|
| 7210 | An Overview of Local Area Networks            |
| 7220 | An Overview of Bridges, Routers, and Gateways |
| 7230 | An Overview of Communications Servers         |
| 7240 | An Overview of Network Management             |

#### Computer Systems

- |      |  |
|------|--|
| 7410 | An Overview of Microcomputers          |
| 7420 | An Overview of Portable Microcomputers |
| 7422 | Notebook Computers                     |
| 7430 | An Overview of UNIX Microprocessors    |
| 7440 | An Overview of Superservers            |
| 7450 | An Overview of Database Servers        |
| 7460 | An Overview of Technical Workstations  |

#### Software

- |      |   |
|------|---|
| 7610 | The Evolution of LAN Operating Systems              |
| 7620 | An Overview of PC-to-Host Communications Products   |
| 7625 | An Overview of Asynchronous Communications Software |
| 7627 | Document Management Software for Networks           |

---

**Report  
Number    Report**

---

- 7640      An Overview of Microcomputer Operating Systems
  - 7645      Multiprocessing Network Operating Systems
  - 7646      An Overview of Multiuser DOS Systems
  - 7648      An Overview of Object-Oriented Programming
  - 7658      An Introduction to Windows
  - 7660      An Overview of Environments and GUIs
  - 7662      An Overview of Utility Software
  - 7668      An Overview of Language Compilers and Interpreters
  - 7680      An Overview of Electronic Mail
  - Peripherals**
  - 7810      An Overview of Mass Storage
  - 7812      An Overview of Optical Storage
  - 7820      An Overview of Displays
  - 7825      An Overview of Scanners
  - 7830      An Overview of Expansion Cards
  - 7832      An Overview of Modems
  - 7835      An Overview of PC-to-Fax Boards
  - 7840      Microcomputer Sound Capabilities
  - 7842      PC Mice
  - 7850      An Overview of Laser Printers
  - 7852      An Overview of Dot Matrix and Ink Jet Printers
- 

**Architectures & Standards**

- Architectures**
- 8210      Microcomputer Bus Architecture
- 8215      Extended Industry Standard Architecture (EISA)
- Standards**
- 8410      Overview of Data Communications Standards
- 8420      IEEE 802 Standards for Local Area Networking
- 8430      ANSI Fiber Distributed Data Interface (FDDI) Standards
- 8432      Introduction to FDDI-II
- 8440      Simple Network Management Protocol (SNMP)
- 8442      Dueling Protocols: SNMP vs. CMIP
- 8460      Modem Standards

# Criteria for LAN Selection

## In this report:

General Selection Criteria .....	4
Specific LAN Requirements .....	5
Type of Vendor .....	8

## Datapro Summary

All LAN vendors claim to offer the perfect solution for your networking needs. The range of hardware and software offered is nearly mind-boggling. However, by following a structured set of LAN guidelines, the processes of planning a LAN installation and selecting the proper equipment become somewhat clearer.

## Planning a System

In this section,<sup>1</sup> we look at the process of translating the information needs of your organization into an overall plan for an office automation system. The process has three steps: (1) Define objectives, (2) Describe the system, and (3) Determine communications needs.

As we proceed, you will see that the discussion is based on determining your data processing equipment needs. The focus of this report is on local area networks—purchasing and incorporating them in your organization. Because the principles are general, however, they are also useful in the broader context that includes devices that hook into the network.

### Define Objectives

The objectives for the installation or modification of an office automation system must relate to the information products of the organization. These products are the items used to gather, store, present, and analyze information needed for the operation of the organization. Table 1 provides a representative list of such products.

By considering how these products are produced and used, you can probably see

several ways to improve the productivity of your information workers. But first, you must be more specific. You must be able to answer in quantitative terms the question: What is the information movement and processing system supposed to do?

It is best to begin by defining some productivity goals. The goals must be quantitative and must relate to information creation, storage, transfer, and reporting. General categories include timeliness, responsiveness, convenience, and workload. The goals you set must be based on the situation you have today and where you think improvements are needed. For example, suppose it is important to get reports or documents quickly; also suppose that these documents are usually revised because of an internal review cycle. A productivity goal would be to reduce revision typing time. Another example: Only certain key professionals can produce briefing materials or reports, and this is draining too much of their time. A corresponding productivity goal would be to reduce the effort required of the professional staff during the information creation phase.

These goals should be as quantitative as possible. Think in terms of how many documents need to be produced by how many people, the time it should take to prepare charts, and so forth.

At this stage, don't limit the number of things you look at. Consider all your information products. Table 1 can be used as a checklist. Consider the feasibility and significance of your goals. Propose goals in

This Datapro report is a reprint of Chapter 9, "LAN Selection," pp. 291-309, from *The Business Guide to Local Area Networks* by William Stallings, Ph.D. Copyright © 1990 by Howard W. Sams & Company. Reprinted with permission.

**Table 1. Typical List of Information Products**

Correspondence	Letter Memorandum Message
Reports	Budget initiatives Case study Fiscal Management Material deficiency Personnel Project status Technical Trip Weekly activities
Documents	A76 cost comparison Action item list Administrative notice Change order Configuration change status report Contract funds status report (CFSR) Contract management systems checklist Cost estimate Data management report Delivery order Engineering change proposal (ECP) Environmental assessment Independent cost analysis (ICA) Invitation for bid Integrated logistics support plan (ILSP) Life-cycle cost study Military construction program reporting Procurement directive Procurement plan Program management directive Program management plan Quarterly resources report Request for proposal (RFP) Sole source justification Staff meeting agenda (and report) Statement of work (SOW) System safety program plan Technical evaluation and report Training plan

practical areas where there is room for significant improvement. If certain activities are already accomplished efficiently, leave them be.

One particularly effective way of formulating goals is to survey the concerns of senior management. This has the added benefit of creating a climate in which reasonable and effective improvements will more readily receive management approval. For certain key information products, determine the concerns that management considers important. Table 2 presents a list of possible management concerns and goals. For example, if timeliness was identified as a primary concern for a key product, the suggested product goal may be to reduce delays in the preparation and distribution of information. If responsiveness was cited as a primary concern, the suggested product goal may be to reduce the amount of "telephone tag" among professionals.

Forms	Data item description Inspection and acceptance document Military order Personnel action request Position description Printing request Purchase request Report of survey Security classification guide Time card Travel request Work order request
Services	Electronic bulletin board Information tracking Response to inquiry Technical assistance
Reviews and Briefings	Business strategy panel meeting Command or senior officer briefing Division advisory group (DAG) review EEO review Executive management review (EMR) Financial management board review Internal management review Periodic program review Program management review (PMR) Quarterly financial review Scientific advisory board (SAB) meeting
Audiovisual Aids	Briefing board Briefing text Graphic aid Vugraph 35mm slide

From these product goals, it is possible to derive more specific system requirements that will guide product selection. Table 3 lists examples of system requirements for the product goals cited in Table 2.

### Describe the System

Now you have some specific goals. The next thing to do is to list some options for moving closer to achieving those goals. As a manager, you know that not all improvements come from buying equipment. Consider three interrelated methods of achieving improvement: organizational, procedural, and technological. Perhaps things are being done by the wrong group (organizational), or in the wrong way (procedural), or with inadequate equipment (technological). Table 4 gives some idea of how these three factors can be used to improve productivity.

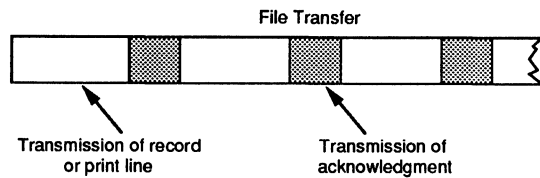
These factors affect each other. A technological change may dictate an organizational change. For example, if it becomes quick and easy to produce charts using presentation software on a personal computer, perhaps the charts should be produced by the professional staff rather than a support group.

Our focus is technological. We need to determine what additional equipment can be beneficial. We can group our options into four categories: input, production, output, and distribution. Table 5 lists representative equipment in each category, together with a description of benefits and drawbacks and an estimate of productivity improvement. This will allow you to assess the impact of various pieces of equipment.

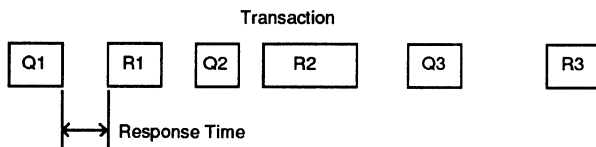
### Determine Communications Needs

You know what information equipment your organization has. You have tentatively decided what the creation, processing, and storage equipment should achieve in terms of

Figure 1. Types of Information Exchange



$$\text{Throughput} = \frac{\text{Total records, print lines, etc.}}{\text{Total transmission time}}$$



productivity. The last element is the information movement equipment—the local area network. To be truly effective and efficient, your collection of equipment must become a system capable of moving information as well as creating, processing, and storing it. The process of selecting a LAN or a set of LANs begins with a specification of requirements. These boil down to two considerations: compatibility and capacity.

*Compatibility* is the easier of the two to specify, but the harder to achieve. You simply need to determine which devices need to exchange data. For example, if you intend to combine data from an accounting system with text from a word processor, there must be an electrical path from the accounting system to the word processor, and the word processor must be able to accept and process the received data.

The second part, *capacity*, is more difficult to pin down but, paradoxically, easier to satisfy. In fact, you don't need to be very accurate in estimating capacity. LANs have tremendous capacity; this is one of their chief advantages. For example, the National Institute of Standards and Technology has a LAN that uses a baseband cable, with a 1-Mbps data rate and CSMA/CD protocol. By today's standards, this is a system with modest capacity and performance characteristics. Over 200 devices in twenty buildings are connected, with an overall extent of 1.5 kilometers. This is not a prototype or experimental system; it is in daily use by a staff with data processing needs as least as great as those of the average office. Yet, the utilization of the network's information movement capacity is under 2 percent!

So, for this part, don't worry about making precise estimates. To determine the required LAN capacity, all that a network designer needs is a general idea of the workload and performance demands of the network. Table 6 summarizes the information needed to design a network. Again, you do not need exact numbers for this checklist; rough estimates will be sufficient.

First, estimate the workload that the various devices on the LAN will generate. Table 7 will give you an idea of the

kind of workload generated by various devices on a local area network. These estimates were prepared by the IEEE 802 standards committee.

It is useful at this point to know the nature of the information exchanges that will take place. We can group these exchanges into two broad categories: file transfer and transaction (Figure 1). A file transfer is primarily a one-way movement of data, such as the transfer of a word processing file to a printer or an image to a facsimile machine. Transactions are typically responses to inquiries or requests for data. In the first case, you are more concerned with *throughput*: the amount of data to be transferred in a given time. In the second case, you are more concerned with *response time*: how long it takes to get an item after it is requested.

Next, you need to look at the total capacity demand. On average, how many information exchanges will take place at a time? What is the peak? For file transfers, you can now add up the individual loads to get an idea of the overall throughput demand on the network. For transactions, it is better to express your needs in terms of a performance goal. For example, for a certain type of transaction, you might specify that you want 90 percent of the transaction to have a response within three seconds.

Finally, you need to look at future growth. If your organization is growing, someday you will want to buy more equipment. Your total file size will grow. You may add new functions to the network. The network should be designed to accommodate growth. Given the capabilities of today's LANs, you should be able to install a network now that will not need to be replaced in the near future. Furthermore, with tiered LANs, any future growth can be absorbed partly by an additional load on existing LANs and partly by the addition of new LANs.

Table 2. Key Product Goals

Management Concern	Desired Goals
Timeliness	Reduce preparation delays Reduce distribution delays
Responsiveness	Reduce "telephone tag" Improve query response time Reduce float
Convenience	Improve information input/output methods
Efficient use of resources	Increase office automation system usage Improve training Reduce user resistance
Organizational effectiveness	Reduce need for reprocessing Reduce duplication of effort Provide access to organizational database
Managerial effectiveness	Improve decision-making process Improve quality of presentation
Cost of labor and overhead	Reduce costs, waste, and overtime

**Table 3. Examples of System Requirements**


---

Produce documents locally
Access or download data from organizational databases
Edit and revise documents easily
Transfer documents among offices within and outside the organization
Perform interactive file queries by office users
Send and receive messages electronically
Produce compound documents (text, spreadsheets, and graphics)
Perform "what if" scenarios quickly
Conferencing (by computer and telephone)
Output presentation and letter quality documents
Transfer word processing skills across offices
Send and receive documents by facsimile
Use electronic calendars and scheduling tools
Integrated, easy-to-use office automation systems
Train users to realize full capabilities of systems
Develop policies and procedures for system usage, information storage and disposition, security, and system operations
Streamline paper handling
Enable the staff to be responsible for administrative and operational needs of office information resources

---

**General Selection Criteria**

Having gone through the process of planning a system, you can translate your needs into a statement of network requirements. The preceding section developed your requirements from a user or application point of view. That should help you answer the question: What is the system supposed to do? The next step is to answer the question: What features must the local area network have to meet my system requirements?

Keep in mind two points as you go through the rest of this report. First, we are not trying to teach you how to design your own network; we just want you to be able to hold your own in dealing with experts and would-be experts. Second, the emphasis is still on what the network will do, not how the network will do it. The *what* is the customer's responsibility; the *how* is the vendor's responsibility. Don't make the mistake of precluding a cost-effective solution by telling the vendor how to do his or her job. In this section, we give you some general selection criteria to keep in mind as you consider various vendors (Table 8). These are the features you should look for and compare.

The first criterion is cost. If the cost exceeds what you perceive as the expected savings in increased productivity, the system isn't worth the price.

Second, the system has to meet the requirements you have established. Although there are always compromises, the system must match the basic requirements. Related to this is the concept that the network should be expandable with only incremental cost. That is, expanding the system later should not be overly expensive or complicated. This allows you to start small, at low risk, and gradually expand the network to meet more of your requirements.

Along with expandability, your system needs reliability. The system should be designed to prevent total network

**Table 4. System Design Model**

Organizational	Procedural	Technological
Work group A will be responsible for initial compilation of budget data	All budget computations will be performed using spreadsheet software	Personal computers
Work group B will be responsible for researching program data	All intraoffice memos will be sent by electronic mail	Software applications including: word processing, spreadsheet, communications, and database management systems
Work group C will replace work group A in the review of key products relating to grant applications	Key products will be developed on personal computers and electronically transferred to support personnel for final processing	Laser printers and letter quality printers
Professional staff in work group C who are responsible for budget compilation will be reassigned to work group A	Support staff will be designated as product distributors	Electronic typewriter with OCR fonts
	Final drafts will be microfilmed rather than filed	Minicomputer
	Training classes will be available to all personnel	Microfilm reader-printer
	Policies on usage, information management (including storage and disposition), operation of systems, and security will be updated semiannually	Modems
		Electronic messaging system
		Local area network and software support

---



failure. In addition, the network should be capable of interfacing with equipment supplied by more than one vendor. We explore this issue at the end of this report.

You want a network that is easy to install, maintain, and occasionally reconfigure. Flexibility is important. You want to be able to connect your equipment to the network easily, with no impact on hardware or software.

Finally, there are software issues, such as the need for servers, security, and network management. These can be supplied in an integrated fashion with the LAN or purchased separately. The former solution may mean that you have fewer vendors to deal with and simpler training. The latter solution may give you more flexibility in your choices.

### Specific LAN Requirements

Let us consider specifically what you will require of a network. Much of the information was discussed in the beginning of the report, and a knowledgeable consultant or vendor can work with that sort of information. This section will give you a better idea of what to ask for, a checklist of what has to be provided.

We can group our concerns into five areas:

- Services
- Traffic
- Reliability
- Growth
- Installation, maintenance, and training

#### Services

The functions performed by the network that are most visible to users and management are referred to as *network services*.<sup>2</sup> A primary consideration in developing network services is the physical network environment. Is the network all in one building or spread over several? If the latter, must the buildings be linked together? Where is the equipment located? What space is available for network components? What false ceilings, conduits, and buried cable runs exist for wiring? Are there any special environmental problems?

Next, what type of equipment will be supported? This consideration leads to the question of interfaces. Some devices, such as terminals, may require a standardized communications interface, such as RS-232C. Others, such as personal computers and workstations, may interface to the network through a communications board mounted in the chassis of the computer. Only resources that need to be shared should be on the network. For example, if you have a special plotter that must be driven by the software in one and only one computer, hook the plotter to that computer—don't make them communicate over the network.

There is also the software interface to consider. For example, what protocols are provided to allow terminals to communicate with a variety of computers?

Next, consider the type of information to be transmitted over the network. This is the whole point of having a network. For data, your primary concern is compatibility. Given the types of equipment you will have and the types of data to be transmitted, the network and the software in the attached devices must provide the communications

software needed for full compatibility. Issues of communications architecture and server software need to be addressed.

Security and privacy are a concern. When multiple user groups have access to the network, the network must provide means for isolating information and restricting access to it. Similarly, unauthorized users should be kept off the network altogether.

Another concern is remote communications. You may want to link two installations by a satellite data channel and routers or bridges. Or you may want to augment your in-house processing capability with outside services, available over a packet-switching network such as Telenet or Tymenet. You may also want to provide dial-in ports on a modem server so that personnel who are traveling can enter the network from anywhere in the country.

Finally, one service is absolutely essential to the successful operation of a network: network management. Some level of monitoring and control is needed.

#### Traffic

The network must have the capability to meet the expected traffic load. This was explored previously in this report. Throughput and response-time requirements must be specified. Especially important are peak-load requirements.

How does this translate to a specific LAN capacity? A good guideline is to estimate the total load and then purchase a LAN with at least 10 times that capacity. That figure may seem excessive, but it is not. First, you are more likely to underestimate than overestimate your capacity needs. Second, demand will grow.

#### Reliability

The network must be available to its users a very high percentage of the time: there must be a long period between component or network failure, and a minimal period before repairs are completed. A useful measure of reliability is *mean time between failure (MTBF)*. For the transmission medium, which includes the cable, amplifiers, taps, and so on, an MTBF of 175,000 hours is reasonable. For intelligent network devices, such as bridges or network interface units (NIUs), an MTBF of 30,000 hours is a good goal.

Another aspect of reliability is the *error rate* of transmitted information. This is usually expressed as the rate of bit errors. There are two types of error rates to be concerned about: undetected and detected. An undetected error is discovered the hard way: The numbers didn't balance, or someone didn't get a paycheck. Detected errors are those discovered by the communications logic and automatically corrected, usually through a retransmission. Undetected errors are obviously unacceptable, and the communications software used across a LAN should be able to prevent these, or at least reduce them to an acceptable minimum. Detected errors are not as bad, but they do cause unnecessary network overhead.

Good reliability values are a detected-error rate of 1 bit in  $10^9$  and an undetected-error rate of 1 bit in  $10^{12}$ . Detected-error rates are easily monitored. Undetected-error rates must be estimated from off-line tests using artificial traffic.

**Table 5. Productivity Citings for Representative Equipment Types**

Equipment	Input Phase		Citings
	Benefits	Drawbacks	
Dictation	Input is four times faster than longhand	Clarity of thought and expression is required	6.25-12% time savings/day—Herbert M. Kaplan, <i>Words</i> , International Word Processing Association, June-July 1980, pp. 40-43
	Transcription is twice as fast as reading longhand or shorthand	Pre-organization of material by dictator is necessary	
	Any secretary can transcribe	Many originators resist dictation	
	Priority work can be handled		
	24-hour input is allowed		
Electronic typewriter with OCR font	Correction time is reduced	No text editing is possible	See optical character recognition citing
	Input by OCR reader into production equipment is possible	No storage capacity is available	
Optical character recognition	Every typewriter with an OCR font can be a low-level input device for text editing and data manipulation	Generally only a limited number of fonts can be read	600% increase in throughput, Compuscan sales literature, AW-5B-018.0
	Re-keyboarding of input text and data is eliminated	Scanning errors are possible	
	Production equipment is freed for word or data processing	Specific input formats may be required	
	Work distribution is enhanced	Accuracy depends upon ribbon, strike, paper, and other variables	
Personal (professional) terminal	Data entry, access, and retrieval time can be reduced	Training is required	50% productivity rise, Emerick G. Zouks, <i>Business Week</i> , April 7, 1980, pp. 81-82
	Paper and supplies can be saved through electronic capture of keystrokes	CPU downtime can affect the use of the terminal	
	Can be expanded, reconfigured, or networked as applications expand and new requirements evolve	Response degradation may occur during peak periods	

**Growth**

No matter how good your analysis of current requirements, you cannot expect to satisfy all your local area network needs once and for all. For example, you may anticipate that your word processing staff will double in the next three years. You should be able to attach these additional stations to the network easily, with no disruption of network operations. Also, the network must have the capability to handle the increased traffic generated by the new stations.

New types of applications or devices might be added to the network (for example, a facsimile machine). This will require new protocols. If such growth is to be possible, the network hardware and software must be capable of easily accommodating new protocols. This is most easily done with an OSI-based architecture because virtually all work on new protocols is in that framework.

**Installation, Maintenance, and Training**

The preceding criteria had to do with the capability and capacity of the network. This subsection speaks to an equally important consideration: the service provided by the vendor. First, the vendor must plan, with your cooperation, the layout of the network. This includes the physical

placement of all cable (including taps, splitters, and amplifiers) and wiring and the location of all network components. The vendor must assure that the layout meets all fire and building codes.

Following installation, maintenance of the network is an ongoing responsibility, divided between you and the vendor. How that responsibility is divided varies; the next few paragraphs present an example.

The customer has the following responsibilities for the network hardware:

- Maintain a permanent survey of the network through the use of logging files and traffic statistics.
- When a problem occurs, perform diagnostics to attempt to localize the fault, using vendor-supplied test equipment and diagnostic software.
- When the fault is localized, reconfigure the network, if possible, to provide continued service.
- When the vendor has corrected a failure, assist the vendor in checking the network.

Customer personnel will require training from the vendor to perform these tasks.

**Table 5. Productivity Citings for Representative Equipment Types (Continued)**

Equipment	Production Phase		Citings
	Benefits	Drawbacks	
Blind automatic word processor	Correction time is reduced  Light change text editing is handled	Text editing functions are limited  Large amounts of text cannot be manipulated  Storage may be limited	Average 48-69% productivity improvement if used for original and revision typing; average 69-98% if used for revision only; NARS standards
Standalone display text editors	Extensive text editing is easily handled  Input and formatting is facilitated by a CRT  Saves paper and supplies through electronic capture of keystrokes	Some are unprogrammable  Large databases may not be manipulated  Storage may be limited	Average 75-133% productivity improvement if used for original and revision typing; average 104-181% if used for revision only; NARS standards
Shared logic word processor	Extensive text editing is easily handled  Input and formatting is facilitated by a CRT  Saves paper and supplies through electronic capture of keystrokes  Different tasks may be performed at the same time  Some are programmable by the user	CPU downtime can be a problem  Specially trained personnel may be necessary to administer the system  Response degradation may occur during peak periods  System backup may be limited	Average 84-89% productivity improvement if used for original and revision typing; average 115-122% if used for revision only; NARS standards
Minicomputer	Can be expanded, reconfigured, or networked as applications expand and new requirements evolve  Performs many different tasks at the same time  Can support simultaneous users  Is programmable  Paper and supplies can be saved through electronic capture of keystrokes	Early obsolescence on large investment is a risk  System backup may be limited  CPU downtime can be a problem  Response degradation may occur during peak periods	Same statistics as for shared logic word processors  Data processing figure totally dependent on each application
Data processing system	Handles many applications  Is programmable  Can be expanded, reconfigured, or networked as applications expand and new requirements evolve  Performs many different tasks at the same time  Can support a large number of users  Handles large amounts of text or data	CPU downtime can be a problem  Specially trained personnel may be necessary to administer the system  May be susceptible to response degradation during peak periods  System backup may be limited  Techniques and procedures usually are alien to the office environment	Same statistics as for shared logic word processors  Data processing figure totally dependent on each application

The vendors maintenance responsibilities follow:

- Cross-check customer diagnostics, if necessary.
- Replace failed components. (To minimize lost service time, faulty components are replaced rather than repaired.)
- After correcting a failure, check the network.

If the network includes substantial software, which is likely, software maintenance becomes quite complex. This should be left entirely to the vendor.

Training is closely related to maintenance. As mentioned, some customer personnel need training to participate in maintenance activities. One or more individuals will need to be trained in network management functions and perhaps security functions. End users, in general,

**Table 5. Productivity Citings for Representative Equipment Types (Continued)**

Equipment	Output Phase		Citings
	Benefits	Drawbacks	
Word processing impact printer	High quality print is produced	Changing printwheels can be time-consuming	15 to 55 characters per second burst speed, <i>Datapro Reports on Word Processing</i> , April 1980  148-533% faster than electric typewriter capability
	Carbon copies can be created	Inserting paper may be required	
	Automatic single sheet feeder or continuous form paper may be used		
Word processing nonimpact printer	High quality print is produced	Usually prints no carbon copies	77 to 92 characters per second burst speed, <i>Datapro Reports on Word Processing</i> , April 1980  770-918% faster than electric typewriter capability
	Automatically feeds paper		
	Typstyles are changed electronically within the printer		
Data processing printer	Prints at high speeds	Print quality is usually unacceptable for word processing output	40-120 characters per second (matrix), <i>Auerbach Computer Technology Reports #31</i> , 1978, p. 13  1184-3554% faster than electric typewriter capability  150-2000 lines per minute (line), <i>Auerbach Computer Technology Reports #31</i> , 1978, p. 13  4813-6417% faster than electric typewriter capability
	Carbon copies can be created		
	Usually incorporates an automatic paper feed		
Photocomposers	Word processors may serve as a means of keyboarding for preparing photocomposer input	May be less expensive to procure this service from outside vendors	20-80 lines per minute, <i>Datapro Reports on Office Systems</i> , September 1979  641-2566% faster than electric typewriter capability
	User can save between 30-40% in the final required number of pages	Compatibility with other systems may be a problem	
	OCRs may be used to generate photocomposer tapes		
	Document preparation time may be decreased		
Micrographics	Recording on microfilm consumes as little as 2% of the space occupied by the same records on paper	Archive quality of film images is questionable	25% (Commander Lloyd C. Burger) to 62% (Reuben Donnelly) access/retrieval time savings, <i>Modern Office Procedures</i> , May 1977, p. 60
	Only seconds are involved in retrieving one of a million records filed within reach of a seated operator	Quality control and inspection procedures may be maintained during the filming, processing, and storage activities	
	Duplicate microfilm files kept off premises to protect against loss of vital information	Complexities exist when indexing for automated retrieval	
	Magnetic tape data is made readable on microfilm in a fraction of the time required for printing out on paper	High costs may be associated with conversion of existing paper files to microfilm images	
	Microfilm records retention cost is considerably lower than paper records systems		

should require little or no training. At most, the application user should have to learn a logon procedure. Beyond that, the network should be transparent to the user.

### Type of Vendor

One final point to consider is the type of vendor. There is a broad spectrum here, but we will group vendors into two categories to clarify the issue: the network vendor and the data processing vendor.

**Table 5. Productivity Citings for Representative Equipment Types (Continued)**

Equipment	Benefits	Distribution Phase	
		Drawbacks	Citings
Intelligent copiers	Document storage can be reduced	Copier may be used as a printing press with increased per page costs	75-600 characters/second transmission speed, IBM 6670 literature (G54 1006)
	Speed of document communications can be enhanced	Potential exists for excess copying	When 600 cps compared to 2 days for mail to arrive, 3,927 times faster for page with 50 lines (65 characters/line)
	Scope of document communications can be enhanced	Potential exists for nonbusiness copying	
	Acts as a convenience copier		
Facsimile (FAX)	Electronically sends text, graphic data, photographs, drawings, or charts with little difficulty	Line costs are relatively high	30 seconds, 6 minutes to transmit 8½ x 11 page, <i>Datapro Reports on Office Systems #2</i> , June 1980
	Documents transfer much faster than mail, messenger, and so on	Compatibility with receiving device must exist	When 6 minutes compared to 2 days for mail to arrive, 480 times faster
		Sizes of sending and receiving documents are limited	
		Certain devices require station-to-station coordination	
Executive telephone	Reduces staff time in using the telephone	Difficult to cost justify	No citing available
	Provides arithmetic capabilities		
Word processing/data processing system with communications feature	Increased speed of document or data communications	Possible compatibility problems with mainframe or minicomputer	40% reduction in dissemination time, <i>Report on Electronic Mail</i> , 4th Quarter 1978, Yankee Group, p. 13

*These sample productivity improvement citings have been derived from representative sources available as of March 1, 1989. New technological developments are rapidly changing the productivity to be derived from the use of these representative equipment types. Therefore, the information needs to be continuously researched and updated.*

*Several items are listed here that we have not discussed before. A blind automatic word processor is a unit without a display, with text usually stored on magnetic cards or tape. An executive telephone is a combination telephone and calculator. Features may include call pickup, automatic dialing, one-line display, calculating, clocking, and appointment calendaring.*

*After selecting some candidate equipment, you need to judge the benefit of any item compared to its cost. Then you can trim the list to those things you are prepared to buy.*

*Although this is a preliminary list, it gives you and any potential vendor a clear idea of your requirements.*

**Table 6. Capacity Checklist**

For each type of information exchange	Initiator device and location Responder device and location
For file transfers	Amount of data for transfer Time allotted for transfer
For transactions	Amount of data requested Response time
For total system	Average loading Peak loading Performance goals
Future growth	Number of devices File size Added functions

The network vendor's primary business is to sell local area networks. Typically, network vendors are independent companies or subsidiaries of firms other than computer companies. Network vendors make their money from the sale of the network.

The data processing vendor, on the other hand, is in business to sell data processing equipment. A LAN is a means of meeting a customer's need for linking equipment purchased from the vendor.

Which type of vendor is preferable? If you already have or are planning to purchase most of your data processing equipment from one vendor and that vendor offers a LAN, it makes sense to get the LAN from the same vendor because it simplifies maintenance responsibilities. Otherwise, you might be better off with a LAN vendor because LANs are the vendors specialty.

**Table 7. Workload Generated from Each Source Type**

Type of Source	Peak Data Rate (Kbps)	Duty Cycle (%)
Heat, vent, air-conditioning, alarm, and security	0.1	100
Line printer	19.2	50-90
File server or block transfer	20,000	0.1
File server or file transfer	100	10-30
Mail server	100	30-50
Information server or calendar	9.6	1-5
Information server or decision support	56	20-40
Word processor	9.6	1.5
Data entry terminal	9.6	0.1-1.0
Data enquiry terminal	64	10-30
Program development	9.6	5-20
Laser printer	256	20-50
Facsimile	9.6	5-20
Voice, immediate	64	20-40
Voice, store and forward	32	30-50
Video, noncompressed	30,000	50-90
Video, freeze frame	64	50-90
Video, compressed	400	20-40
Graphics, noncompressed	256	1-10
Graphics, compressed	64	10-30
Optical character reader	2.4	50-90
Gateway	1,000	0.1-1.0
Host, 0.5 MIPS	128	20-40
Host, 5 MIPS	1,000	20-30

**Table 8. Vendor Selection Criteria**

Total cost	
Meets requirements	
Expandable incrementally in cost	
Capable of interfacing with equipment supplied by multiple vendors	
Ease of	Installation Maintenance Reconfiguration Interconnection
Software	Servers Security Network management

**References**

This report is a revision of material that appeared in F. Derfler and W. Stallings, *A Manager's Guide to Local Networks*, Prentice-Hall, 1983.

<sup>1</sup>The tables and many of the ideas in this section are adapted from an excellent publication by the National Bureau of Standards (now the National Institute of Standards and Technology), *Guidance on Requirements Analysis for Office Automation Systems*, NBS Special Publication 500-147, March 1987.

<sup>2</sup>Many of the ideas in this section are adapted from another excellent publication by the National Bureau of Standards (now the National Institute of Standards and Technology), *The Selection of Local Area Computer Networks*, NBS Special Publication 500-96, November 1982. ■

# Choosing a NOS: Banyan, Microsoft, and Novell

## In this report:

Microsoft LAN Manager .....	3
NOS Comparison Table .....	4
Novell NetWare .....	8

## Datapro Summary

Choosing a network operating system (NOS) for your organization's local area network (LAN) is one of the most important decisions you may ever make. The NOS is the platform upon which a company's information system is built. NOS selection has direct and widespread impact on the usefulness and potential of an organization's information processing capabilities. Selecting a NOS involves careful evaluation of a company's specific requirements, and thorough investigation of available NOSs. While there is a wide variety of LAN operating systems on the market, the dominant product for enterprise networking is NetWare from Novell. Banyan VINES and Microsoft LAN Manager are the primary challengers to NetWare's market dominance.

## Banyan VINES

### Overview

Virtual Networking System, or VINES, is developed and marketed by Banyan Systems, a private firm located in Westboro, MA. Although Banyan is an underdog in the network operating system (NOS) race with the smallest installed base and market share, it offers some of the most valuable and coveted NOS features on the market.

### Background

- Year established: 1983
- Gross annual sales: \$97 million
- Number of employees: 600
- President and CEO: David C. Mahoney

### Sources of Revenue

- Software: 80%
- Hardware: 20%

—By Katherine Wollerman and  
Richard Scruggs  
Business Systems Group, Inc. (BSG)

### Statement of Direction

"To simplify the use and management of distributed networks."

Banyan targets VINES to the *Fortune* 500 for large WAN installations. The StreetTalk Global Naming Scheme is well suited for such large distributed WANs. David Mahoney, Banyan's president, is focusing on pushing LANs closer to the world of mainstream large-system data processing.

### Product History

VINES 1.0 shipped in 1984 and ran on the Banyan Network Server (BNS); the more powerful Corporate Network Server (CNS) and smaller Desk Top Server (DTS) followed. Following Novell's lead, Banyan has decreased its reliance on hardware to concentrate on software; the CNS is the only server still offered by Banyan.

VINES/286 (June 1986) and VINES/386 (May 1988) were released to run on Intel 80286 and 80386 PCs, respectively. VINES/486 was introduced in July 1990. In April 1991, Banyan unveiled VINES 4.10. The current version, VINES 4.11, was released in September 1991.

### Strengths

- StreeTalk global naming scheme (GNS) is the naming scheme upon which other vendor's GNSs are modeled.
- Management of multiple file servers is easy.
- Can support up to ten printers on one file server.
- Excellent performance with multiuser database management products.
- Solid telecommunications options. (However, best options are available on CNS servers, not VINES 4.11).
- Good built-in electronic mail system with ties to StreetTalk.
- Easy to install with automated ability to upgrade or downgrade workstation software.
- Unlimited number of concurrent users on one server.
- Unlimited number of open files.
- Automatic administrator alerts and warnings regarding system performance.
- Built-in support for connecting to minicomputer and mainframe hosts.
- Symmetric multiprocessor (SMP) technology allows an operating system to utilize the full power of its hardware, significantly increasing speed and performance.

### Limitations

- Small market share.
- Third-party application and support software is limited.
- Does not support duplicate concurrent hard disks and/or controllers, which would allow the system to continue operating normally when one disk/controller fails.
- Limited hard disk fault tolerance.
- Maximum number of hard disks supported per file server is two (660MB each); greater disk storage available on CNS systems.
- Does not support Named Pipes programming interface for workstation applications.
- UNIX foundation imposes some performance penalties.
- Workstation RAM requirements to access LAN are large at 112KB.
- Does not support full Apple Filing Protocol (AFP) connectivity to Apple products.
- Relatively slow performance.

### Analysis

Banyan VINES is a UNIX-based operating system primarily known for its multiserver networking features. It is renowned for its capability to support a large number of nodes and geographically separated servers. With VINES, information system managers can replace their wide area terminal systems with PCs that are both locally networked and connected across a wide area.

### Installation and Management

VINES is easier to install than LAN Manager or NetWare. The setup process is almost completely self-contained, requiring little use of instruction manuals. The ease with which a NOS is installed is often a good indication of its

ease of maintenance; this holds true with VINES. A network administrator can install and maintain a VINES network of 25 servers and 1,000 users with little difficulty.

VINES tailors its management features after those found in traditional UNIX systems, thereby offering a no-frills path to installation. There is no mouse use and no fancy windows or pull-down menus. Commands are entered at the command prompt or selected from basic menus. User interface with VINES follows the same command-line format.

VINES documentation is substantial, well organized, and easy to use, with one exception. The documentation does not contain a central location for information on error messages, a quality that would be helpful to systems administrators.

Most of the popular third-party applications, such as Wordperfect, Lotus, and Oracle, are VINES-compatible. Most DOS applications will "unofficially" (without Banyan certification) work with VINES. Banyan is continually adding to its certified list of applications. A catalog of VINES-compatible third-party applications is available.

### StreeTalk

Managing large networks is simplified with StreeTalk. Banyan's outstanding global naming service (GNS). StreeTalk, a distributed database, translates logical names into physical internet addresses. Changes to the database are automatically replicated across the network, making it easier for system administrators to move resources from server to server to relieve network load. A nice feature recently added to Banyan's GNS package is StreeTalk Directory Assistance. This feature allows users to substitute actual names when they do not know the network address. Familiarity with this GNS is necessary to manage a VINES network because StreeTalk is the structure around which all network resources and user accounts are organized.

StreeTalk is one of VINES' major strengths, setting the standard in GNSs. Novell and Microsoft have not yet introduced a GNS as advanced as Banyan's. Analysts and users have commented that neither Novell's NetWare Naming Service (NNS) nor LAN Manager's Domain Naming can compare to the ease and comprehensiveness of StreeTalk. GNSs will continue to increase in importance as wide area networks proliferate.

### WAN Capabilities

VINES includes many features suited for large networks. Because VINES runs on top of UNIX (System 5 release 3), it can run multiple tasks for multiple users. VINES supports symmetric multiprocessing (SMP), an innovative feature which allows the NOS to utilize the full power of its hardware, significantly increasing speed and performance. VINES supports an unlimited number of open files and concurrent users on one server.

Banyan has received widespread acclaim as a technical leader, and many VINES network managers feel that they are on the cutting edge of local area networking. This is because VINES is equipped with many of the features that users of other NOSs are still waiting for. Support for mainframe and minicomputer host connections, for instance, is a built-in feature of VINES. VINES also simplifies the design of networks with thousands of users, connected by StreeTalk global naming and directory services, and linked by integrated gateway, bridge, and router software.



### Enhancements

VINES 4.10 included some important improvements over previous versions. Banyan doubled the size of the VINES Toolkit, adding Dynamic Link Libraries (DLLs) for both Windows and OS/2. These DLLs are a collection of shared program functions that can be simultaneously referenced by the NOS and by multiple applications. They provide more efficient memory usage and reduction in application size, and make further versions of VINES applications easy to upgrade. The expanded toolkit also included two new applications programming interfaces (APIs). The Mail Client API permits developers to generate new electronic mail applications that communicate with the VINES mail transport system. The VINES Network and System Management API allows developers to create network management clients which are integrated with VINES' existing network management capabilities.

VINES 4.11 includes an enhanced version of VINES SMP that adds the AT&T StarServer E to the list of SMP servers supported by VINES. Other additions include enhanced support for Compaq Intelligent Array disk systems and internal tape drives; support for diskless client workstations; support for more LAN adapters; administrator-forced logout; remote booting and enhanced StreeTalk Directory Assistance; closer integration of StreeTalk with application development tools; and OS/2 NETBIOS support for application sharing between DOS and OS/2 users. New wide area communications features include X.29 Dial-in (allowing multiple PCs to dial into a VINES 4.11 server over a single X.25 connection) and the ICAplus Intelligent Communications Adapter.

### Limitations

The breadth of VINES support for both users and resellers leaves something to be desired. Banyan is beginning to rectify this by implementing various programs. For example, ten new channel partner programs have been announced which are designed to furnish worldwide Banyan resellers with better sales support tools, more responsive technical service, quicker access to marketing and product information, and enhanced communications resources. This program will also help improve end-user support.

A small market share (approximately 7.5%) is often seen as Banyan's biggest drawback. There is a perception that the NOS war is being fought solely between Microsoft and Novell, and some IS managers are reluctant to work with Banyan. In light of recent events, especially 3Com's withdrawal from the NOS market, managers are concerned that smaller firms like Banyan may not survive in this competitive market.

Banyan recognizes these weaknesses and is taking calculated steps to win a more substantial market share. A February 1991 alliance with AT&T supplied Banyan with considerably more clout in the value-added reseller and distribution channel. The alliance allows resellers already authorized to sell either Banyan VINES or AT&T StarGroup products to sell both. A recent alliance with Compaq should also give Banyan more credibility in the NOS marketplace.

### Summary

VINES is a solid choice for users with large or wide area networking needs. The many built-in features, such as mainframe and minicomputer connectivity and StreeTalk global naming service, make VINES a good choice for large networks. Banyan is also competitive in its pricing, offering VINES at \$5,995 for an unlimited number of users.

## Microsoft LAN Manager

### Overview

Microsoft is a public firm based in Redmond, WA. It is the largest and most successful microcomputer-based software company in the industry, with 1990 revenues of over \$1 billion.

Microsoft designed the LAN Manager NOS as an OS/2 application. It consists of a mouse-oriented, point-and-shoot interface. The current version LAN Manager 2.0, is relatively inexpensive—a five-user license costs \$995, and a license for an unlimited number of users costs \$5,495. Its key features include managing multiple servers and fast file copying. Drawbacks stem from its relative immaturity, which may cause snags during installation.

### Background

- Year established: 1975
- Gross annual sales: \$1.1 billion
- Number of employees: 7,000
- President: Michael R. Hallman
- CEO: William H. Gates III

### Sources of Revenue

- Software: 87%
- Hardware: 13%

### Statement of Direction

"To be a leading force in the international growth and development of the computer industry, by developing software tools that are at the heart of those machines."

Microsoft's networking strategy, focused on the future of a company's networking needs, is "based on the idea of connections" to various products and support.

### Product History

Microsoft codeveloped OS/2 LAN Manager with 3Com, but never intended to sell it directly; LAN Manager was to be an OEM product. When the two companies agreed to develop the new NOS in 1987, Microsoft hoped to match the success of its MS-Net operating system, which it never sold directly but which formed the basis for IBM PC LAN Program, 3Com 3+, and many others. Initially, over 40 vendors were enlisted to sell LAN Manager under their own name, including 3Com, IBM, AT&T, Ungermann-Bass, Digital Equipment, Unisys, Hewlett-Packard, and NCR.

As an OEM product, LAN Manager was unable to win a significant market share from Novell. In February 1991, Microsoft announced that it would sell its own version of LAN Manager directly. Since that announcement, the most important OEM vendor of LAN Manager, 3Com, has withdrawn from the NOS market entirely.

### Strengths

- Customer base expected to grow steadily over next three years, especially in *Fortune* 1,000 market.
- Global naming scheme uses domain server concept.
- Direct queries to a mainframe.
- OS/2 client peer-service feature.
- Sophisticated client/server applications.

**NOS Comparison Table**

Features	Banyan VINES 4.11	Microsoft LAN Manager 2.0	Novell NetWare 2.2	Novell NetWare 3.11
<b>PC Operating Systems and Environments</b>				
DOS	Yes	Yes	Yes	Yes
OS/2	Yes	Yes	Yes	Yes
Windows 286	Yes	Yes	Yes	Yes
Windows 386	Yes	Yes	Yes	Yes
HP New Wave	No	Yes	Yes	Yes
PC-MOS/386	Yes	No	Yes	Yes
Presentation Manager	Yes	Yes	Yes	Yes
Distributed/Parallel Processing	Yes (Named Pipes) (2)	Yes (Named Pipes) (1)	Yes (Named Pipes and RPCs) (3)	Yes (Named Pipes and NLMs) (4)
<b>APIs</b>				
NETBIOS	Yes	Yes	Yes	Yes
Named Pipes	Yes	Yes (5)	Yes	Yes
SNA Gateway APIs	Yes	Yes	Yes	Yes
APPC/LU6.2	Yes	Yes	Yes	Yes
<b>System Fault Tolerance</b>				
Database: Verify Multiple Record/File Updates	Yes	Yes	Yes	Yes
Database: Record Roll Back/Roll Forward	Yes	Yes	Yes	Yes
Hard Disk: Dynamic Bad Block Mapping/Circumvention	Yes	Yes	Yes	Yes
Hard Disk: Duplicate Hard Disks	No	Yes	Yes	Yes
Hard Disk: Duplicate Hard Disks/Controllers	No	Yes	Yes	Yes
Hard Disk: Correct Sector Failure w/Duplicate Hard Disks	No	Yes	Yes (6)	Yes (6)
UPS Monitoring	Yes (7)	Yes	Yes	Yes
Duplicate Concurrent File Servers	No	Yes	No	Yes (3.1) (8)
<b>SQL Database</b>				
SQL Support	Yes	Yes	Yes	Yes
<b>Advanced Storage Features and Performance</b>				
Maximum Number of Concurrent Users	Unlimited (9)	992	100	250
Maximum Number of Open Files	Unlimited (9)	8,000	1,000	100,000
Maximum Number of Concurrent DOS 3.1 Locks	Unlimited (9)	64,000	1,000-2,000	1,000-2,000
Maximum Number of Hard Disks	2 (660MB each)	8	32	1,024
Maximum Number of Volumes	Unlimited (9)	24	32	32
Maximum File Size	Unlimited (9)	2GB	255MB	4GB
Maximum Number of Directory Entries per Volume	Unlimited (9)	Does not apply	32,000	2,097,152
Maximum Volume Size	Unlimited (9)	2GB	255MB	32TB
Maximum Disk Space per File Server	1.2GB	96GB	2GB	32TB
Split Volume Across Multiple Disks	No	No	No	Yes (32 disks) (10)
Split Files Across Multiple Disks	No	Yes	No	Yes (32 disks) (11)
Multiple Hard Disk Seeks w/One Controller	Yes	Yes	Yes	Yes

**NOS Comparison Table (Continued)**

Features	Banyan VINES 4.11	Microsoft LAN Manager 2.0	Novell NetWare 2.2	Novell NetWare 3.11
<b>Advanced Storage Features and Performance (Continued)</b>				
Multiple Hard Disk Seeks Across Multiple Controller	Yes	Yes	Yes	Yes
Elevator Seeking	No (12)	Yes	Yes	Yes
Minimum Amount of Server RAM	4MB	5MB	2MB	2.5MB
Maximum Amount of Server RAM	16MB	16MB	12MB	4MB
FoxBase+/LAN Application Performance	Excellent	Good	Excellent	Outstanding
<b>Miscellaneous Advanced Features</b>				
Automatically Reconfigure Buffers and Caches	Yes	Yes	Yes	Yes
Manually Reconfigure Buffers and Caches	Yes	Yes	Yes	Yes
Add New Hard Disks Dynamically	No	Yes	No	Yes
Add or Remove Volumes	Yes	Yes	No	Yes
Add or Remove Print Services	Yes	Yes	Yes	Yes
Add or Remove Communications Services in Server	Yes	Yes	No	Yes
Add or Remove Network Adapters or Protocols	Yes (13)	Yes	No	Yes
Server Software	UNIX Processes (15)	OS/2 Processes (14)	VAPs (16)	NLMs (17)
Load Server Applications Dynamically	Yes	Yes	No	Yes
Recover RAM Used by Unloaded Application	Yes	Yes	Yes	Yes
Invoke Loading of Another Application	Yes	Yes	Yes	Yes
<b>RAM Usage by Workstation Software</b>				
DOS Workstation LAN Drivers	80KB-110KB	2KB	54KB	58KB
DOS Workstation Token-Ring Drivers	None	9KB	20KB	20KB
DOS Workstation NETBIOS Drivers	20KB	25KB	26KB	26KB
Load DOS Drivers into High Memory	Yes	Yes	Yes (3rd party)	Yes
<b>Security and Server Management</b>				
Access Control Lists/User Groups	Yes	Yes	Yes	Yes
File-Level Security	Yes	Yes	No	Yes
Supervisory Rights for Workgroup Managers	Yes	Yes	Yes	Yes
Encrypted Passwords on LAN Cabling	Yes	Yes	Yes	Yes
Encrypted Data on LAN Cabling	Yes	No	Yes	Yes
Encrypted Tape or Hard Disk Backups	Yes	Yes	Yes	Yes
Intruder Lockout	Yes	Yes	Yes	Yes
Intruder Alert and Auditing	Yes	Yes	Yes	Yes
Force Password Change Every X Days	Yes	Yes	Yes	Yes
Audit of Logons	Yes	Yes	Yes	Yes
Limit User Disk Space Usage	Yes	Yes	Yes	Yes
Automatically Disconnect Inactive Workstation	Yes	Yes	Yes	Yes
Automatically Reconnect Workstation After Inactivity Logout	Yes	Yes	Yes	Yes
Forced Disconnect by Supervisor	Yes	Yes	Yes	Yes

**NOS Comparison Table (Continued)**

Features	Banyan VINES 4.11	Microsoft LAN Manager 2.0	Novell NetWare 2.2	Novell NetWare 3.11
<b>Administrator Alerts and Warnings</b>				
Disk Full	Yes	Yes	Yes	Yes
Excessive Errors	Yes	Yes	Yes	Yes
Intruder Alert	Yes	Yes	Yes	Yes
Printer Problems	Yes	Yes	Yes	Yes
Print Job Complete	Yes	Yes	Yes	Yes
<b>Global Naming Scheme (GNS)</b>				
GNS Type	Distributed	Domain	None	Domain (opt.)
XNS Clearinghouse Support	No	No	No	Yes
Sun Yellow Pages Support	No	No	No	Yes
CCITT X.500 Support	Planned	No	No	Yes
<b>Protocol Support</b>				
TCP/IP	Yes	Yes	Yes	Yes
AFP	No	Yes	Yes	Yes
SMB	No	Yes	No	Yes
NFS	No	No	No	Yes
OSI	Yes	Yes	No	Yes
XNS	Yes (XNS Base)	Yes	No	Yes
DLC	No	Yes	No	Yes
<b>Bridging and Internetworking</b>				
Maximum LANs per Server	4	4	4	16 (18)
Maximum LANs in External Bridge	4	4	4	4
IBM Token-Ring Bridge Support	Yes	Yes	Yes	Yes
Internetwork Servers via Asynchronous	Yes	Yes	Yes	Yes
Internetwork Servers via X.25	Yes	Yes	Yes	Yes
Internetwork Servers via T1	Yes	Yes	Yes	Yes
<b>Communications and Connectivity</b>				
Asynchronous	Yes	Yes	Yes	Yes
Single PC Remote Access	Yes	Yes	Yes	Yes
Multuser Remote Access	Yes	No	Yes	Yes
X.25	Yes	Yes	Yes	Yes
HDLC	Yes	No	No	Yes
T1	Yes	Yes	Yes	Yes
IBM SNA	Yes	Yes	Yes	Yes
IBM SNA via Token-Ring	Yes	Yes	Yes	Yes
IBM SDLC	Yes	Yes	Yes	Yes
IBM BSC	Yes	Yes	Yes (3rd party)	Yes (3rd party)
IBM System/3X	Yes	Yes	Yes	Yes
IBM AS/400	Yes	Yes	Yes	Yes
Digital Equipment	Yes	Yes (3rd party)	Yes (VMS) (19)	Yes (VMS) (19)
Apple Macintosh	Yes (3rd party)	Planned	Yes	Yes
AFP-Compliant Apple Macintosh	Yes (3rd party)	Planned	Yes	Yes
Sun	Yes	Yes	No	Yes

**NOS Comparison Table (Continued)**

Features	Banyan VINES 4.11	Microsoft LAN Manager 2.0	Novell NetWare 2.2	Novell NetWare 3.11
<b>Miscellaneous Add-on Products</b>				
Database	Excellent (Oracle)	Good	Excellent	Does not apply
Network Diagnostics	Good (NetMan) (20)	Good (NetMan, NetView)	Good	Planned (NetView)
Miscellaneous Utilities	Fair	Good	Excellent	Does not apply
<b>Printing</b>				
Maximum Number of Printers on File Server	10	8	5	16
Non-File Server Print Servers (21)	10	3rd Party	3rd Party	3rd Party
Availability of 3rd-Party Products	Good	Good	Excellent	Does not apply (extensive APIs)

- (1) Standard OS/2 applications utilizing interprocess communications allow users to run multiple applications concurrently on a PC and extend those applications over the network to access other network resources.
- (2) OS/2 services (e.g., SQL servers) run as a process within UNIX on the file server. VINES inherently runs many distributed or parallel processes such as StreeTalk updates, least cost routing of information, and mail messaging.
- (3) In addition to Named Pipes, NetWare Remote Procedure Calls (RPC) generate ISO-compatible network code that allows processes running on different machines with different operating systems to communicate at the application level.
- (4) OS/2 processes (e.g., SQL servers) can be run in dedicated computers on the LAN and can be called via the NetWare Requester and OS/2's Named Pipes. NetWare 3.11 NLMs load in the file server and can be loaded or unloaded on demand. An NLM is available to allow OS/2 processes to run within a NetWare 3.11 file server.
- (5) 150 additional APIs are included above and beyond the standard set of 250 which are part of OS/2.
- (6) When a read failure is encountered, the duplicate hard disk is used to recover from the error. The bad area is marked as unusable and its data is written to a new sector.
- (7) Upon notification of an extended failure, all files are closed, disk heads parked, and server shut down. When power is restored, the server reverts to the same state as before the shutdown (including software processes) and is automatically reconnected to workstations.
- (8) If a NetWare 3.11 file server fails for any reason (e.g., power supply, mother board failure, etc.), its backup immediately takes over. Users experience no loss of network function.
- (9) Due to UNIX technical details and the design of VINES and StreeTalk, these items have no practical limit on VINES LANs or WANs. Principal limiting factors are disk space and available WAN communications services.
- (10) Splitting a logical volume across many physical hard disks transparently provides very large storage areas for database applications.
- <sup>11</sup>Splitting files across multiple hard disk drives can improve database system performance. While one hard disk receives the first portion of a file, other hard disks and controllers queue up the remaining portions.
- (12) Commonly used system software such as StreeTalk and UNIX are automatically placed in the most efficient place on a hard disk.
- (13) Most VINES resources can be loaded and accessed on the fly; an example is communications services. Another example is the loading of token-ring drivers for simultaneous use of VINES, TCP/IP, or IBM source routing protocols on a token-ring adapter.
- (14) Multiple OS/2 processes can be run on the server concurrently (e.g., SQL server applications).
- (15) Banyan servers have a suite of software that runs as UNIX processes and supports server functions. Examples include communications (e.g., 3270 gateway), Oracle, StreeTalk, and Banyan's Vanguard security. OS/2 processes such as SQL server run as processes under UNIX in the file server.
- (16) Value Added Processes (VAPs) are written by Novell or third parties to provide added functionality to a NetWare file server. They load in the file server at boot time and cannot be unloaded. An example is the NetWare for Macintosh VAP.
- (17) NetWare 3.11's NetWare Loadable Modules (NLMs) load in the file server. NLM software can be written by Novell or third parties and provide added functionality to NetWare 3.11 file servers. NLMs can be loaded and unloaded on demand.
- (18) Most server hardware cannot support 16 LANs per server. The actual number of LANs possible is a function of the total number of available slots for network interface cards.
- (19) In addition to standard asynchronous Digital VT100 terminal emulation, Novell provides NetWare for VMS. NetWare for VMS software runs on a Digital VAX and provides NetWare 2.2 file server functions on PCs connected to the VAX. Improved file transfer and terminal emulation functions are provided.
- (20) Banyan provides NetMan as a diagnostic tool. NetMan is a comprehensive diagnostic product that reviews local and remote servers and resources.
- (21) Maximum number is a function of the third-party software used. The typical number is 6 to 16.

- Only NOS with auto-reconnect feature that re-establishes severed communications with workstations when the file server comes back on-line.
- Mouse-based point-and-click operation.
- Strong administrative features.
- Inexpensive five-user package available.

#### Limitations

- Expensive client support beyond ten days (\$2,495 for ten incidents within one year).
- Client software and servers consume a great deal of RAM.
- No direct server-to-server bridging.
- Problems supporting large number of users.
- Problems arise from LAN Manager's relative immaturity.
- No Macintosh support.

#### Analysis

Despite its early lack of acceptance, LAN Manager is considered a strong contender in the NOS race. Based on the OS/2 operating system, LAN Manager provides good performance for small- to medium-sized networks, a broad set of features, and aggressive pricing. LAN Manager's faults arise mainly from its relative immaturity.

#### Management Features

LAN Manager 2.0's basic features are solid. Security features include password encryption, file protection, and delegation of some administrative functions to different users. Network maintenance can be performed through a single menu-driven program. Domain Naming, Microsoft's global naming service, allows administrators to group multiple servers into a domain. The group of servers can then be managed as a single entity. All servers in a domain share user IDs, groups, and resources.

LAN Manager 2.0 is equipped with helpful administrative features unavailable in other network operating systems. In the event of server failure, an auto-reconnect feature re-establishes broken communication ties with workstations as soon as the server comes back on-line. Another feature allows a server's resource configuration to be saved and restored in the event of a system crash.

Installing LAN Manager 2.0 is easy but time consuming. Installation is menu-driven, and help screens are available at the click of a mouse. However, the installation process contains a major drawback—hardware requirements for LAN Manager are higher than for other NOSs because it requires OS/2. OS/2 also makes installation more involved than necessary.

LAN Manager 2.0 documentation is helpful and does a good job of explaining potentially complex procedures, such as installing and managing the server software. One issue that could be addressed further is performance tuning.

LAN Manager 2.0 is tightly integrated with both OS/2 and Windows, a definite plus for users of these operating systems. Support for Macintosh clients, however, is not yet available. Another limiting factor is that two servers on separate network segments cannot share files or transfer information without a third-party bridge.

Through various network testing, LAN Manager has been found to be best for small- to medium-sized networks. When the capability to handle various data loads

was tested, LAN Manager kept the pace with the competition when data loads were light. When the data load was increased to simulate the load of 16 users on 300 workstations, LAN Manager's performance suffered.

#### Limitations

The most prominent drawbacks of LAN Manager 2.0 seem to stem from its relative immaturity. LAN Manager runs into problems when handling large loads while running with two processors in the server. When using Xcopy, only the top-level directory is copied from the server, instead of the entire directory tree. Problems like these are small, but bothersome.

Another flaw in LAN Manager is its need for RAM. For a 386 or a 486 server, the most often used, 6MB of RAM is required, and Microsoft recommends 9MB. Between 67KB and 178KB is used by the DOS client software. Client/server applications need even more RAM.

#### Summary

LAN Manager 2.0 is a solid NOS, well suited to small- and medium-sized networks. It offers strong network administration and fault-tolerance features, as well as attractive pricing. The consistent, mouse-oriented interfaces make life easier for the users and system administrators.

LAN Manager 2.0 is also aggressively priced. A five-user package is priced at \$595; add-on packages are available in ten-user increments for the same price. An unlimited-user license costs \$5,495.

---

## Novell NetWare

#### Overview

Novell is a public firm located in Provo, UT. Novell offers a line of NetWare products, including NetWare 3.11 and NetWare 2.2. The NetWare products are the most mature of the NOSs discussed in this report, and lead the pack in overall performance. Novell dominates the NOS market—various sources place NetWare's market share at anywhere from 60% to 70%. Despite these outstanding strengths, NetWare also has some drawbacks, including a relatively high price and fewer built-in network management products than its competitors.

#### Background

- Year established: 1983
- Gross annual sales: \$497.5 million
- Number of employees: 2,100
- President and CEO: Raymond J. Noorda

#### Sources of Revenue

- Software: 80%
- Hardware: 20%

#### Statement of Direction

Novell's stated goal is to be the "standard of standards." NetWare permits the use of microcomputer, minicomputer, or mainframe operating systems, dissimilar workstations, multiple protocols and topologies, a variety of communications links, and virtually any application. NetWare joins these into a powerful NOS. NetWare serves as the integrating component, and facilitates connectivity among products of various vendors.

### Product History

NetWare 2.2 is the ninth generation of the NetWare 286 product line; NetWare 3.11 is the eighth generation of the NetWare 386 product line. There have been versions of NetWare for the Zilog Z80, Motorola 68000, and Intel 8086, 80286, and 80386. The NetWare product line has matured from a simple disk server system to a workgroup file server system, and is now a network computing system.

### Strengths

- Acknowledged best performer by many independent testers; known for consistently superior performance, mature technology, and hardware independence.
- Extensive fault-tolerance features.
- Support for OS/2, Named Pipes, and SQL products.
- Supports duplicate concurrent file servers, allowing one server to fail while the other keeps the system operating.
- Excellent overall system security features, LAN management functions, and usage statistics.
- Varied connectivity options.
- Only Apple Filing Protocol (AFP)-compliant support for Macintosh.
- Clear market leader—approximately 60% market share; largest worldwide installed base.
- Wide availability of add-on products from third-party vendors.
- NetWare 3.11 supports up to 1,024 hard disks on server, with maximum storage capacity of 32TB.
- NetWare 3.11's NetWare Naming Service (NNS) supports other vendors' GNSs.
- New releases are backward-compatible with earlier NetWare versions. (This kind of support is a corporate goal for Novell.)
- NetWare is a true network operating system—not an application that runs under another vendor's operating system.
- Strong set of Application Programming Interfaces (APIs).
- Complete implementation of TCP/IP built into core.
- Supports the widest variety of hardware in industry.

### Limitations

- Slight time lag in producing workstation shells for new versions of DOS and OS/2.
- NetWare 2.2 supports maximum of only 100 concurrent users; NetWare 3.11 supports only 250 concurrent users.
- NetWare 2.2 has limited storage features—maximum number of open files per server is 1,000; maximum volume size of 255MB (greatly expanded in NetWare 3.11).
- NetWare 2.2 does not support NetWare Naming Service.
- OS/2 processes not supported in server.
- Relatively expensive.
- Cannot perform direct queries to IBM mainframe.
- NetWare Naming Service for NetWare 3.11 must be purchased as option for \$1,995.

### Analysis

Novell bills NetWare 2.2 as a product “designed to meet the needs of the workgroup computing segment of the network computing arena.” Made for smaller groups, NetWare 2.2 is available in 5-, 10-, 50-, and 100-user configurations. NetWare 2.2 replaces all existing NetWare 2.1x products (ELS, Advanced, and SFT NetWare). It provides superior fault tolerance, high performance, Macintosh support, capability for a VAX to function as a file server, strong communications support, VAPs (value-added processes, a file server-based application), OS/2 support, and SQL server support. One feature it does not include is a global naming scheme, which is less important in small installations. An appealing quality of NetWare 2.2 is its path upgrade. Users who start out small do not have to decide which product to buy; they can grow their LANs simply by buying bigger licenses.

NetWare 3.11 is a high-end NOS designed specifically for large businesses to meet their complex, enterprise-wide computing needs. NetWare 3.11 connects DOS, OS/2, Windows, Apple Macintosh, or UNIX NFS workstations with each other and with host systems. In addition to the standard features provided by NetWare 2.2, NetWare 3.11 provides two to four times the performance of NetWare 2.2, is very transaction and heavy-usage oriented, and is positioned to take advantage of very large (32TB) and fast storage systems as they become available. Additionally, NetWare Naming Service (NNS) supports other major vendors' GNSs.

Novell's NetWare is not based on any one particular platform, as VINES is based on UNIX and LAN Manager on OS/2; NetWare was built from the ground up to be a network operating system. This architecture ensures that all Novell products work together, connecting many different vendors' hardware and software.

### Installation and Management

The installation of a NetWare operating system, although fairly intuitive and trouble-free, is the most involved of the NOSs covered in this report. However, adding a user to a NetWare network is quicker, requiring only that an administrator copy a few files onto the workstation's disk.

Novell has established an extensive program in which one can become a Certified NetWare Engineer (CNE). CNEs are trained in-depth on the installation and maintenance of a NetWare operating system, and they must pass a series of tests to become certified. Companies that send their employees to this program will have an in-house certified expert on Novell NetWare operating systems, ensuring that support is always available. Novell also offers NetWare on-line support (through CompuServe) for NetWare products. These various forms of support ensure that a NetWare operating system will not be without the support it needs.

### Limitations

Two prominent features of NetWare 3.11 are the NetWare Naming Service and the Remote Management Facility. Both features facilitate managing large, multiserver internetworks. An option priced at \$1,995, NetWare Naming Service allows servers to be grouped into domains, and users assigned to these domains. The users receive access rights to various servers. With this feature, a user logs in only once—instead of logging in to each server—to access network resources. Although Novell's global naming service is good, it does not quite match up to the competitors' GNSs. Its domain support is not as integrated with the

base product as LAN Manager's, and it is not as comprehensive or as elegant as VINES StreeTalk.

NetWare's biggest drawback is its price. NetWare 3.11's least expensive package, supporting 20 users, is expensive, at \$3,495. A 100-user package can be purchased for \$6,995, and a 250-user package for \$12,495. NetWare 2.2 is offered in 5-, 10-, 50-, and 100-user packages, ranging in price from \$895 to \$5,495.

---

This report was developed exclusively for Datapro by Katherine Wollerman and Richard Scruggs of Business Systems Group, Inc. (BSG). Katherine Wollerman is a consultant with BSG SI Consulting's applications development group. Richard Scruggs is BSG's director of business development. BSG, based in Houston, TX, is a national systems integration company specializing in business solutions based on client/server, network computing technology. BSG can be reached at (713) 965-9000.

### Summary

Novell possesses two strong advantages in the NOS market—product maturity and market share. NetWare products repeatedly install and run smoothly, while other operating systems frequently experience minor but annoying problems. With the largest market share and worldwide installed base, support is not an issue. Novell is a solid, safe choice in NOSs.

Novell's NOSs are hardware-independent, allowing network supervisors to integrate different, often incompatible types of networking hardware within a single network. NetWare supports the largest variety of hardware of any NOS on the market. For example, NetWare is the only operating system to support the Macintosh. NetWare for Macintosh 3.0 allows Macintoshes to access NetWare 3.11 servers directly (this feature is not available with NetWare 2.2). ■



# Selecting a 486

## In this report:

The Soul of a New Machine .....	2
Memories are Made of This .....	3
Of Bits and Buses .....	3
Finding Your Dream Machine .....	3
Should You or Shouldn't You? .....	4
The Once and Future Chip.....	5

## This report will help you to:

- Evaluate the capabilities of 486 microprocessors.
- Anticipate the future evolution of 486 systems.
- Determine whether a 486 is the system that will best meet the needs of your organization.

## Introduction

What's faster than a speeding bullet? The i486, the latest and greatest offering from the Intel family of micro-processing chips. The i486 is *fifty times faster* than the 8088 chip. Put another way, it can scan the entire Encyclopedia Britannica in two seconds.

Great Jeopardy fodder, but what if you don't *need* to scan the entire Encyclopedia Britannica in two seconds? Should you stop reading this report? No, you should not. Don't worry, I'm not here to make a hardware fashion victim out of you.

Instead, I'm going to show you what the 486 offers the business user and help you decide whether you even need that much power. In some cases, you may be better off with a

386DX, a 386SX, or even—surprise!—a 286.

The 486 marketplace is an interesting one. "The market is split into four segments right now," says Bruce Stephen, a research analyst with International Data Corporation, a market-research firm in Framingham, Mass. "Number one, as a LAN server running *NetWare* or some other network operating system. Two, as a Unix hub for a multiuser Unix PC. Three, the person who's got a very intensive mix of applications, running a number of large applications with a real need for multitasking. Four, on the technical side, the 486 is appealing to people doing CAD-CAM and other numeric-intensive types of applications."

You may have heard 486s are expensive, and you're right, for the most part. Depending on how they're configured, these muscle boxes can run anywhere from \$10,000 to more than \$30,000 putting them out of

---

This Datapro report is a reprint of "Trailblazing with the 486" by Jackie Fox, pp. 22-27, from *PC TODAY*, Issue 2, Volume 5, February 1991. Copyright © 1991 by Peed Corporation. Reprinted with permission.

reach of many desktop users. But surprise again—some 486s list at less than \$4,000. Even at higher prices, a 486 can be surprisingly cost-effective, depending on your needs. But before we get to the price-performance issue, let's look under the hood of these souped-up machines and see what makes them so special.

### The Soul of a New Machine

Intel's i486 chip is at the heart of these sleek PCs. The i486 is more fully integrated than earlier chips, although it retains full backward compatibility. "The critical things about the 486 are its integration in performance tied to its compatibility," says Tom McDonald, Intel's product marketing manager for 386 and 486 chips. "We've integrated 386 capabilities onto one 1.2-million transistor chip." Those capabilities include the following.

#### A 387-Compatible Math Coprocessor

The coprocessor is typically used for number crunching in CAD and spreadsheet programs. While the 387 math chip is a separate, optional chip on the 386DX, it's built into the CPU of every 486 chip. Since the coprocessor is directly on the CPU, it's faster. "AutoCAD will run more than 30 percent better on a 486," says Joe Kua, director of research and development and acting president of Arche Technologies, Inc.

#### An Internal 8K Cache and Cache Controller

The cache speeds up memory access by using faster, more expensive static RAM (SRAM), in contrast to the dynamic RAM (DRAM) used by main memory. (Because DRAM is volatile, it has to constantly be refreshed, or recharged, so it doesn't lose its contents. SRAM is faster because it doesn't require refreshes.) The cache keeps a copy of the most frequently used data so the CPU can look in the cache first, thus speeding up the search process.

How much difference are we talking about? "On a 25MHz 486 you're talking about 25-30 nanoseconds for the CPU's clock cycle," says Jeff Cheng, product engineer for Eltech Research, Inc. "DRAM just can't keep up, because today's technology for DRAM is limited to the 60-80 nanosecond range. SRAM can go 10 to 15 nanoseconds." In other words, the CPU has to wait for DRAM, but it doesn't have to wait for SRAM.

Although nanoseconds may not seem like much, they have a way of adding up. This performance gain is particularly important at the professional level, according to Eric Leppanen, product marketing manager for AST Research. "If having a 486 vs. the alternative enables a person to calculate their spreadsheets ten minutes a day earlier than it otherwise would, you calculate their time out for a year and that gets real interesting," Leppanen says. "You save a lot of money over time."

You may wonder how often the CPU ends up accessing main memory anyway because the cache didn't have what it needed. Not often, according to McDonald. "The hit rate with internal caches is very, very high," he says. "It's typically well over 90 percent and depending on what you're writing it can be as high as 99 percent."

Sending data within a chip is obviously faster than sending it back and forth between separate chips, but that's not all Intel has done to speed things up. "Although the whole chip is completely redesigned, there seems to be this naive perception that we just put these chips together and that's all we did," says McDonald. "A 25MHz 486 is typically two to three times faster than a 25MHz 386, depending on the application. That's obviously because we optimized the instructions."

A 486-25 is also faster than an equivalent 386 running at 33MHz, with the 486 getting 11 million instructions per second (MIPS) and the 386 clocking in at eight MIPS, according to Cheng.

#### Burst Mode

The i486 chip also has a nifty feature called burst mode, the PC equivalent of Han Solo kicking the Millennium Falcon into hyperspace. Both 386s and 486s process data 32 bits at a time internally, but they handle that data quite differently. Let's say you want to process 128 bits of data (four 32-bit chunks). On a 386, that normally takes eight clock cycles. The first four clock cycles are spent by the CPU specifying where to find each of those four 32-bit chunks. The other four clock cycles are spent by memory returning each chunk back to the CPU.

A 486 system using burst mode is different. "In a 486, the first clock cycle is used to calculate four locations at once," Cheng says, "but it still takes four clock cycles to get the data. You're talking about five clock cycles vs. eight, so you're actually reducing access time by three-eighths of the time."

---

## Memories are Made of This

You don't just buy a 486 chip, however; you buy a 486 system, and not all 486 systems are created equal. For example, not all 486s are designed to take advantage of the 486 chip's built-in burst mode.

"Companies who have just implemented the exact 386 design wouldn't get the benefit of the burst bus," says Jim Rilly, Intel's senior applications engineer. "Someone looking for absolute optimum performance would look for that feature. Even if you don't have it you get much more performance than on a 386 system, but you get the most performance if you take advantage of burst mode."

Secondary or external caches are another system variable. Some system manufacturers use an additional RAM cache to make memory access even more efficient.

When do you need an external cache, and when is the 8K internal cache good enough? "For DOS applications and unsophisticated software, the 8K cache is good enough, but for Unix and a lot of other things, the 8K cache is a drag, it's not enough performance," says Kua.

Some manufacturers choose not to include an external cache. Instead, they tweak standard memory for more performance, with memory interleaving the most popular option. "Interleaving takes place in main memory. It divides the memory into different memory banks; one memory bank does refreshing while the other memory bank is available for the CPU to access, so you're actually reducing the workload for the DRAM. For a product that does not have RAM caching, they're definitely using interleaving," Cheng says.

Whether they choose interleaving or a secondary cache, the reason manufacturers fine-tune the memory architecture is to provide the fewest wait states possible. You may not care about interleaving vs. caching, but you should care about wait states.

Wait states vary by machine, but they refer to the amount of time the CPU spends waiting for something to happen. If the CPU expects to get data back within two clock cycles, and it does, then you have zero wait states. If the CPU expects to get data back in two clock cycles, and it takes four clock cycles instead, then the CPU has one wait state. If it takes six clock cycles, the CPU has two wait states.

"Comparing zero and one wait state, with one wait state you're talking about the CPU being idle 50 percent of the time [during memory operations]," Cheng says. "If the CPU is spending half its time waiting for memory to finish the job, then you're wasting the CPU's resources, and that's the most expensive resource of the computer.

"There are three ways to get close to zero wait-state performance," Cheng says. "The first is to use expensive SRAM. It's true zero-wait state, but the cost is just too much. The second is RAM caching, and that's the most popular way because it's the most effective one and it gives you the most effective performance. The third way is interleaving main memory. That's the cheapest way but the efficiency compared to RAM caching isn't as good."

---

## Of Bits and Buses

Systems vary in the input-output bus as well. You can buy systems with either 32-bit Micro Channel or EISA (Extended Industry Standard Architecture) buses. You may be surprised to learn that some systems feature 16-bit ISA (Industry Standard Architecture) buses. Each of them has its strong points.

"We decided on ISA because it's cheaper and faster to bring out," Cheng says. "With EISA the peripherals that support it are limited, and the cost to make an EISA machine is higher than an ISA machine. We chose ISA because it's stable and people can take advantage of the 486 CPU."

If you decide you want to go full-steam ahead with a 32-bit bus, EISA offers the advantage of backward compatibility with 8- and 16-bit cards; Micro Channel does not.

"Micro Channel's lack of backward compatibility is a disadvantage," admits Scott Schafer, NCR Corporation's director of workstation marketing for the United States, "but we've found that people don't bring their cards forward anyway because when you use an 8-bit card in a 32-bit machine, the whole EISA machine has to slow itself down, and people buying a high-performance product really don't want to do that."

---

## Finding Your Dream Machine

Given all the variations in 486-based systems, how do you make a good choice? It depends on your

needs. "What to look for in a 486 system all depends on what applications you're running, and that's the first basis for whether you need a 486 at all," says David Middleton, manager of PC product marketing for NEC Technologies, Inc.

One good place to start is determining whether the machine will be used as a standalone machine or as a file server. Many manufacturers are taking a two-pronged approach, offering both high-end servers and more reasonably priced desktop models. IBM, NEC, AST Research, Advanced Logic Research, NCR, and Compaq offer both desktop models and machines designed to be used as file servers.

"File server vs. desktop are two distinct markets," says an IBM spokeswoman. "The requirements for each of those markets are different. For instance, in a file server, one wants greater storage capacities as well as greater expansion capabilities. Those are two key ingredients."

"Basically you need to consider your software first," says Cheng. "By determining what kind of software you'll run it's easier to shop for a good 486. Let's say you use a file server application. You can eliminate the graphics issue because it's handling communications, but if you're doing CAD, then graphics need to be considered."

Scott Schafer agrees. "People who do imaging and AutoCAD would be very interested in the implementation of the graphics subsystem," he says.

The disk input-output system is another critical area. "Again, it all depends on your requirements," says Middleton. "If you're running CAD, then you need lots of data storage but that data storage doesn't have to be very high performance. What that means is once you have the diagrams and software loaded from disk, you're running a memory-intensive environment. Then it's not necessary to have a very high-speed disk subsystem."

When do you need a high-performance disk system? "If you're running a multiuser machine servicing 10 to 15 users, or if your machine is a file server on a 100-node network, that disk subsystem needs to be very high performance," says Middleton. "It depends on what you need, but the disk input-output generally becomes a significant bottleneck in overall system performance."

"For a disk-intensive program such as a database, the disk drive is very critical to the performance of the machine," agrees Cheng. "If you

don't have fast enough devices the whole machine will slow down because the CPU always has to wait for the data before it can process it."

If you need maximum performance from your disk subsystem, you'll want to consider a caching disk controller, which operates on the same principle as the memory caches we've been discussing. A caching disk controller contains its own RAM cache, which it can search much more quickly than it can search the hard disk.

You don't have to be in the market for a file server to benefit from a solid disk I/O subsystem, however. "Today's big programs all require a lot of read and write and a lot of storage, so the disk factor is the major factor people need to consider when they buy a 486," says Cheng.

---

### Should You or Shouldn't You?

Before you decide which 486 to buy, however, decide whether you need a 486 at all. "If you pay that much money to buy a 486, it's better to have applications that require that kind of high-speed processing," says Cheng. "CAD and networking are the major ones, and Unix also requires that kind of speed because it's a multiuser environment."

You don't have to be an engineer or a LAN guru to benefit from a 486, however. "If you're into any financial analysis, if you do large spreadsheets or things of that sort, you're typically configuring a 386 with a 387 math coprocessor," says Leppanen. "If you cost that out vs. a 486 where you get the numeric coprocessor effectively for free, you'll find that the cost difference between the 386 and 486 begins to get real interesting. In many cases, you'll be better off just going with a 486 up front."

I promised I wouldn't make you a hardware fashion victim, so here's the other side of the story. "I'm not saying this technology is for everyone," says Arche Technologies' Michael Bruzzone. "For what most people do, a 386-33 should be sufficient performance."

"The average user should consider a 486 if they need a network server or a powerful engineering system," says Thereza Lam, president of Eltech Research, Inc. "For general purposes, the business or home user should need a 386."

Even the 386's horsepower may be overkill. Remember the 286? "If you're still running

character-based DOS, a 286 is probably just fine," says John Dunkle, a vice-president at Workgroup Technologies, Inc., a market-research firm based in Hampton, New Hampshire. "A 286 or a 386SX purchased today is still a safe buy. Don't get sucked into needing the latest and greatest because that's what everybody else has. Let the application dictate the platform."

"Software developers are very pragmatic and say what's the standard today, what has the biggest installed base I can write to," Dunkle says. "Well, gee, guess what it is today, it's the 286. Most PCs shipped are still the 286—43 percent, or roughly 4.5 million. That's because the average selling price of the 286 on the street is less than \$1,000.

"What users see in performance isn't the chip architecture, and it isn't the operating system," Dunkle says. "It's 'On the application I'm trying to do, is the performance adequate to meet my requirements?' If the answer is no then they go to the next level of technology. It's that easy."

"If you want to get into *Windows* very heavily, then the platform you need to commit to is the 386DX, and if you want to go beyond that into imaging and perhaps OS/2, then the i486 begins to make sense," Dunkle says. "But those are application hybrids rather than mainstream corporate computing, which is computing for the rest of us. Don't let the platform dictate how you justify the system."

Bruce Stephen also counsels patience. "You should know that prices are coming down. If you absolutely have to buy in now it should be very beneficial to what you do; it should raise your productivity by a very large amount. Like any other

technology, if you can afford to be patient and wait a little bit, and wait for the volumes to increase, that will drive prices downward."

Prices are already dropping. In October, AST Research announced the debut of the *Bravo 486/25*, which carries a suggested retail list price of \$3,995. And Positive Corporation added another twist when it introduced the *Positive 486/25*. The Positive 486/25 includes 2MB of RAM, a 106MB hard drive, and six expansion slots for the list price of \$2,995, if you purchase it through a retail warehouse such as Price Club or Whole Club. (The system costs more if you order it direct from Positive Corporation.)

---

### The Once and Future Chip

Even though prices are dropping, you'd still like your PC to be a long-term investment. Making a smart purchasing decision is tough, because the hardware development cycle is becoming shorter. "The 8088 sold for four years as mainstream," Dunkle said. "The 286 sold for three years. The 386 has been selling for two years, and now we've got the i486 competing at the same time."

It would help if you could just replace the microprocessor instead of buying a whole new system, and some companies now offer that option. IBM, AST Research, and Advanced Logic Research offer models that place the CPU on a removable board.

"Basically what the future brings is that you'll be able to upgrade a system by adding a couple more CPUs, the way you upgrade a system today by adding memory," says McDonald. ■



# Selecting Mass Storage Devices

## In this report:

Backup Storage .....	2
New Storage Trends.....	3
Native or Compressed? .....	4

Buying Considerations .....	4
Massive Mass Storage .....	6
Narrowed Choices and Trade-Offs .....	8

## Datapro Summary

Larger, more complex software applications make it necessary for 386- and i486-based personal computers to use mass storage systems with greater capacities. This report offers an overview of the types of mass storage technologies available for high-capacity systems. It focuses on hard disk drives, rewritable optical disk drives, and tape as the primary storage technologies. The report also outlines purchasing guidelines for those considering buying a mass storage device.

Powerful 386- and i486-based PCs are placing new demands on data-storage systems. With the proliferation of gigabyte-plus database applications, the need arises for greater storage capacities and increased operating efficiencies. This is evident in applications such as high-performance CAE and desktop publishing workstations, as well as data-packed LAN file servers.

Many PC applications need 1 gigabyte of storage capacity or more, and mass-storage technology must keep pace with this need. A variety of technologies are available for this new multigigabyte environment, and conflicting considerations are involved in choosing among them.

This Datapro report is a reprint of "Giga-Storage," by Richard A. Peters, pp. 201-213, from *Byte Magazine*, Volume 16, Number 5, May 1991. Copyright © 1991 by McGraw-Hill, Inc. Reprinted with permission.

## Primary-Storage Choices

A wide range of mass-storage technologies for high-capacity systems exists on the market today. They compete with and often complement each other (see Table 1). Any examination of what to buy begins with knowing if you intend to use the device as the primary- or backup-storage medium.

The hard disk drive is almost omnipresent with personal computer systems today. It is the dominant primary data-storage product and will remain so until the next century.

Hard disk drives supplanted tape drives more than 10 years ago as the preeminent primary-storage medium because the technology could randomly access data. Random access makes it possible to retrieve data anywhere on the disk in milliseconds. Many suppliers provide 1-gigabyte-plus hard disk drives

**Table 1. Data-Storage Technologies**

Technology	Benefits	Negative Issues
<b>Hard disk</b>	Random access High capacity	Lack of removability
<b>Rewritable optical</b>	Random access High capacity	High drive cost High media cost Lack of standards
<b>Quarter-inch cartridge</b>	Low cost Compatible among multiple sources High data transfer rate	Low capacity, but increases are evolving quickly
<b>4-mm DAT helical scan</b>	High capacity Low media cost Fast search capacity	Lack of compatibility among multiple sources High drive cost Low data transfer rate
<b>8-mm helical scan</b>	High capacity Low media cost	Single-source supplier Low data transfer rate

with varying costs, capacities, and software-derived performance features.

Hard disk drives are fixed devices with finite storage capacities. The finite capacity makes it important to know how much storage you really need now and in the future. Since the medium is not removable, the data processing system needs to provide for backup of the database. Backup data storage systems must be rewritable and use removable media.

The rewritable optical disk drive is becoming a popular option for primary storage. Rewritable optical drives, currently featuring up to 600-megabyte capacities per removable disk, record data on extremely thin platters that can be removed like floppy disks and moved easily from system to system. The optical disk provides a two-sided medium that you must manually turn over to access the other half of available storage.

Demand for optical storage technology began with the advent of rewritable optical devices, which you can reliably reuse as often as you wish. In fact, testing by manufacturers has demonstrated that you can perform 1000 read/write cycles per day for 40 years with no degradation of data, disk, or drive mechanism. There are three main OEMs of rewritable optical disk drives—Ricoh (San Jose, CA), Sony Corp. of America (San Jose, CA), and Hitachi America (Tarrytown, NY).

Ideal as primary storage devices in image-oriented applications in vertical markets, optical drives also have notable drawbacks. The key drawback is their data-access times. The random-access times for optical drives are 5 to 7½ times higher than those for hard disk drives (50 to 150 ms compared with 10 to 20 ms). A primary factor is the weight of the intricate optical drive head. Even though the respective weights can be measured in grams, these heads tend to be 20 to 50 times heavier than hard disk drive heads.

The upfront cost for rewritable optical disk technology is also very high. The average price for a 300-MB rewritable optical disk drive is about \$5,500, compared with below \$3,000 for a 300-MB hard disk drive.

Finally, while hard disk drives can give you more than 1 gigabyte of continuous storage, current rewritable opticals give you a maximum of 300 MB per side—you must then manually turn them over to obtain additional storage. This amount of storage may be too small for LANs with more than 10 users.

Rewritable optical disk drives have some utility in backup and archiving applications, but high costs for the drive and medium make the technology cost-effective only when justified by other applications.

Another type of optical technology should also be noted. You can only write to WORM (write once, read many times) optical disks once, but you can read them as often as you need to—much like a phonograph record (if anybody still remembers what they are). WORM drives are useful in specialized backup and archival applications, such as the storage of legal and financial documents, because the stored information cannot be altered.

## Backup Storage

To protect against catastrophic data loss, you need to keep a backup copy of the data stored on your primary devices. Periodic system backups performed to copy data from the primary device to an offline medium become increasingly time consuming and impractical with the growing size of a database.

With today's multigigabyte PC applications, having a backup mass storage unit is a necessity.



You use the backup device to make copies of frequently used data, transfer older data from the primary unit, and retrieve archived data with maximum efficiency. While tape drives are no longer used as primary-storage devices, they remain the most attractive medium for inexpensive, removable, and high-capacity backup storage.

The quarter-inch cartridge is the leading tape drive technology. It was invented specifically for data processing applications. QICs now account for more than two-thirds of all tape drive shipments, and they will continue to be the dominant backup technology well into the mid-1990s. The QIC manufacturing community began introducing products with storage capacities above 1 gigabyte late last year, and 6-gigabyte products are projected by early 1993.

In comparison with competing technologies, QIC drives are low in cost, with retail prices under \$2,000 for a 1.3-gigabyte system; 1.3-gigabyte cartridges retail for about \$35. Data transfer rates for current products reach up to 600,000 bps, allowing backup of 1 gigabyte in under 30 minutes. Any file can be accessed in 40 seconds or less in 1-gigabyte-plus tape devices.

QICs are also enhanced by the clear performance standards that manufacturers in this field have developed. Through the work of a cooperative organization known as Quarter-Inch Cartridge Drives Standards, manufacturers now supply products that are interchangeable between drive and medium regardless of origin. Some key suppliers of these products include Tandberg Data (Westlake Village, CA), Wangtek (Simi Valley, CA), Sankyo Seiki America (Torrance, CA), and Archive (Costa Mesa, CA).

Two new types of high-capacity tape drives, 8mm helical-scan tape and 4mm digital audiotape, have made their presence felt in the last few years using a technology known as helical scan. Helical scan describes how the tape travels over the read/write heads, which are mounted on a spinning drum aligned diagonally with the recording track. With the drum spinning rapidly and the tape passing over the drum slowly, the head writes data in a diagonal pattern corresponding to the pitch of the head, with a high tape-to-head velocity. Helical-scan drives function the same as videocassette and commercial DAT recorders, and intricate error

checking and redundancy must be implemented to produce an acceptable error rate for data processing applications.

The 8mm helical-scan tape drives are provided solely by Exabyte (Boulder, CO). These drives provide the highest capacity-to-volume storage ratio of any mass-storage device currently in use (326 MB per cubic inch) for storage capacities of from 2½ to 5 gigabytes. The higher-capacity drives sell for about \$3,900 at OEM quantities. Medium costs are in the \$10 range.

DAT systems use a data-recording standard devised by Hewlett-Packard and Sony, called Digital Data Storage. DDS is currently licensed by 17 different companies.

DAT storage devices use the same tapes that DAT recorders use for entertainment purposes. The needs of the audio world dictate that DAT provide long recording times (2 hours) with relatively slow transfer rates (180,000 bps). The audio recording industry has lobbied successfully to protect its intellectual property rights against unauthorized copying by DAT devices. These restrictions have reduced the availability of consumer DAT drives.

The capacity of the DAT cassette is about 2 gigabytes. Medium costs could be as low as \$7 each, but some experts say a special cassette may be more desirable. The cost of such a medium may be \$15 each. Drive costs at the OEM level are under \$1000.

DAT offers a fast-search capability that allows access to a file in an average time of about 15 seconds. Comparable time for a QIC is about 30 seconds.

---

## New Storage Trends

Recent technological breakthroughs in rewritable optical disk drives and high-capacity tape drives have at least one common thread: They all provide unlimited storage because they use high-capacity media that you can quickly insert and withdraw from the drive.

This trend toward removability is further supported by jukeboxes, or autochangers. These devices have recently become available for all three major storage technologies, although optical devices seem to be leading the way. A jukebox typically contains two drives and a mechanical arm

## Native or Compressed?

The compress-or-not-compress issue has existed for years, but the advent of data-compression chips, performing what is termed *loss/less* compression, has made it popular again.

Data compression maintains data integrity and compresses the data in real time—transparently. It removes redundancy from a closed set of information symbols (i.e., a data block) without any loss of information. In simple terms, it records more data in a smaller

space. *Compression ratio* is the length of compressed output relative to the length of uncompressed input. The key is that data stored in a compressed form must be reliably retrievable.

Benefits can be derived from implementing data compression at the peripheral level. You can theoretically increase the peripheral's capacity by the factor of the compression ratio and can increase the peripheral's internal continuous-transfer rate because it's

recording less data. For example, a 2-to-1 compression factor could double the peripheral's sustained data rate.

However, Mike Casey from InfoCorp (Cupertino, CA) pointed out, "If data compression is implemented at the system level, no benefit will be obtained by incorporated data compression in an intelligent disk or tape drive." Also, if data interchange is a requirement, all systems involved not only need compression but must use the same hardware/firmware implementation of it, or they will find their data unreadable.

Remember, not all data lends itself to compression. First, the benefits will vary because the

amount of data redundancy varies widely among types of data; this variation changes the compression ratio. Second, you will benefit from compression only if the host system can support the higher continuous-transfer rates. Third, the benefits will be negated if the compression process cannot keep up with system needs. Fourth, compressed data may actually expand when compressed again. With compression, there are no hard and fast rules.

Compression is not the complete answer to the need for increased capacity and data rate. The decision to adopt data compression must be carefully thought out because of its associated

used to select and load one of many disks stored within it. A newly announced optical jukebox contains 32 optical disks for a total storage capacity of 17.9 gigabytes.

Another development unfolding, particularly for very large storage applications, is the mixed-media mass-storage systems. These systems use a combination of hard disks, optical disks, and tape-to-store files. Where the data is stored depends on how frequently it is used—a particular file will automatically migrate from online hard disk storage to slower optical and tape systems as the frequency of its use decreases. Such mixed-media systems allow you to take advantage of all technologies, using each to its maximum potential.

### Buying Considerations

There are several criteria you should consider when you plan to purchase a mass storage medium (see Table 2). The issues relate to the specifications of the products and to the particulars of your individual applications.

### Table 2. Buying Consideration Checklist

Specification-Related Criteria	Application-Related Criteria
Random/sequential access	Security:
Capacity	Shock
Speed:	Environment
Transfer rate	Data:
Access time	Shelf life
Compatibility	Transportability
Fault tolerance	Quality
Data integrity	Mixes and matches
	Growth potential
	Costs:
	Drive
	Media

Several specifications are fundamental to choosing appropriate primary and backup storage, beginning with how a mass-storage device obtains data—randomly or sequentially. A random-access device stores data randomly and retrieves it using an identifying address. A sequential access device stores data in a prescribed ascending or descending sequence and retrieves it by searching for it from the beginning to the end of the file.

risks. If you are interested in unattended backup, it's wise to rely on native capacity instead of compressed capacity. With native capacity, you know that the data will fit on a particular medium and that it will take a certain length of time to back up.

Many standards and implementations of data compression exist, such as Hewlett-Packard's, STAC's, IBM's, and a number of other proprietary algorithms. You could almost say it's *algorithm du jour*. Given the rapid advances in computer throughput, it is extremely important that the compression algorithm be able to handle higher data rates of future systems. Some cannot.

In addition, different companies, using the same compression algorithm, may offer unique implementations of it. Furthermore, multiple compression algorithms on top of multiple formats (e.g., digital audiotape's DDS and Data/DAT) create more confusion. You can only roll the dice and hope your selected product, format, algorithm, and vendor are around in the coming years.

Whole sections of the computer industry, mainly at the high end, are attracted to 8-mm tape's large data-storage capabilities and low cost, but they've held back because of transfer rate. In these cases, adding data compression will satisfy

their needs until 8-mm products with further native-performance increases are introduced.

When looking to the future, technologies that can be extended without compression are ultimately much better than those that can only be extended with compression. You can always improve the native performance by adding compression to it. Exabyte (Boulder, CO) currently has plans to extend the 8-mm tape's transfer rate to 1 megabyte per second and the cartridge capacity up to 10 gigabytes using current native technology.

Be aware that some backup peripherals are

being advertised with large capacities and high transfer rates—improvements made by using data compression. Because so many variables affect data compression, you may not be able to achieve these large advertised performance capabilities as promised.

Data compression is an added benefit, but it's not the total answer. Native capacity and transfer rate are also important.

—Grant Wilcox

Random access is typical of hard disk and optical disk drives, and sequential access is typical of tape drives. Random access is a more direct and faster method of accessing data, but you must weigh that against the increased cost, the size of your database, and the intended use of the device (as a primary- or secondary-storage unit). Sequential access is slower and less convenient, but it is much less expensive.

How much storage capacity is enough to meet your needs in a cost-effective manner? Having too much is no better than having too little because you will be saddled with expensive, underutilized storage equipment.

To estimate your data-capacity requirements, consider the following example. Assume you have 20 users on a LAN. Allocate 40 MB for system and common application files and 20 MB per user. Assume one 200-MB common database. Under this formula, the drive size required is 640 MB. However, the growth factor is also important, and you should factor in roughly twice a user's current needs for growth. The recommended drive size, then, exceeds 1 gigabyte.

Speed is another key performance consideration. Speed specifications include the product's access time, measuring how fast it can locate data, and its transfer rate, measuring how fast it can move data from one place to another. Sometimes it's difficult to know how fast is fast enough. You must weigh the importance of speed against your other needs. A few milliseconds' worth of speed could come with an unacceptable price tag.

It is also worth noting that the data-rate capability of a storage device is often not the limiting factor on a system's ability to move data to and from the device. Average data transfer rates above 300,000 bps for peripheral devices are unusual on today's PC systems, so a tape drive's data transfer rate of only 200,000 bps is often the optimum solution.

Compatibility of the mass-storage device with the LAN operating systems and topologies in which it must operate is also important. Widely used LAN operating systems include Novell's NetWare 286/386 and Microsoft's MS-NET for IBM PC-compatible LAN environments, and

## Massive Mass Storage

### Resource Guide

For *real* mass storage, be prepared to drop a couple of hundred thousand dimes in the jukebox.

Jukeboxes (also called auto-changers) provide unprecedented on-line storage for PC, workstation, and LAN users. They range in capacity from a few tens of gigabytes to over a terabyte, and they are available in CD-ROM, WORM (write once, read many times), rewritable optical, and tape-based configurations. Listed below are some manufacturers of jukebox mass-storage systems.

Advanced Graphics Applications, Inc.  
90 Fifth Ave.  
New York, NY 10011  
(212) 337-4200

Alphatronix, Inc.  
2300 Englert Dr., Suite C  
Research Triangle Park,  
NC 27709  
(919) 544-0001  
fax: (919) 544-4079

Aquidneck Systems International, Inc.  
650 Ten Rod Rd.  
North Kingstown, RI  
02852  
(401) 295-2691  
fax: (401) 295-1851

AT&T  
100 Southgate Pkwy.  
Morristown, NJ 07960  
(800) 247-1212  
(201) 898-8000

Bell & Howell Document Management Products Co.  
6800 McCormick Rd.  
Chicago, IL 60645  
(708) 675-7600  
fax: (708) 675-9271

Control Data Corp.  
P.O. Box 0  
Minneapolis, MN 55440  
(612) 853-8100  
fax: (612) 853-5300

Cygnnet Systems, Inc.  
2560 Junctions Ave.  
San Jose, CA 95134  
(408) 954-1800  
fax: (408) 954-9391

Delta Microsystems, Inc.  
5039 Preston Ave.  
Livermore, CA 94550  
(415) 449-6881  
fax: (415) 449-6885

Digital Equipment Corp.  
146 Main St.  
Maynard, MA 01754  
(508) 493-5111  
fax: (508) 493-8780

Dilog Corp.  
1555 South Sinclair St.  
Anaheim, CA 92806  
(714) 937-5700  
fax: (714) 978-2420

Epoch Systems  
8 Technology Dr.  
Westborough, MA 01581  
(508) 836-4300  
fax: (508) 836-3802

Exabyte Corp.  
1685 38th St.  
Boulder, CO 80301  
(303) 442-4333  
fax: (303) 442-4269

FileNet Corp.  
3565 Harbor Blvd.  
Costa Mesa, CA 92626  
(714) 966-3400  
fax: (714) 966-3490

AppleTalk/AppleShare and A/UX from Apple and TOPS from Sun Microsystems for the Macintosh world.

Topology refers to the physical layout of the components of a LAN—a bus topology connects all devices in one line; a ring topology connects each workstation to two other workstations in a circle. The topology that you employ can affect how the storage device will perform. With ring topologies, for example, the more PCs that are present, the slower the LAN, which can slow down data access and transfer times.

You will also want to examine fault tolerance features: Do you want or need duplicate systems for continuous storage operation in the event that one should fail? What data integrity features does the product provide to prevent the accidental erasing or contamination of your data?

### Further Factors

There are a number of other issues that you need to take into account before choosing a 1-gigabyte-plus mass-storage device.

Data security is the first of these issues. What happens if the drive or the medium is dropped? The ability of the device to withstand shock is very important. Is the device rugged and impervious to the environment? Some mass-storage units are sensitive to magnetic fields and x-rays. You also need to find out the unit's operational range in terms of temperature and humidity.

You should consider the projected longevity (shelf life) of the storage medium and drive, and how long you intend to store the data on the device. The demand for WORM optical drives continues in spite of the arrival of rewritable optical ones because WORM technology provides permanent storage that can't be overwritten.

The transportability of the medium and drive can be a key factor. Data-storage applications used

Hewlett-Packard Co., Inc.  
Disk Storage Systems  
Division  
11413 Chindin Blvd.  
Boise, ID 83714  
(208) 323-3290  
fax: (208) 323-3991

Hitachi America, Ltd.  
Computer Division  
Peripherals & Systems  
Marketing, MS:500  
Hitachi Plaza  
2000 Sierra Point Pkwy.  
Brisbane, CA 94005  
(800) 283-4080, ext. 877  
(415) 589-8300

Laser Magnetic Storage  
International Co.  
4425 Arrows West Dr.  
Colorado Springs, CO  
80907  
(800) 777-5674  
(719) 593-7900

Literal Corp.  
2180 Executive Cir.  
Colorado Springs, CO  
80906  
(719) 579-0460  
fax: (719) 579-0450

Memorex Telex Corp.  
6422 East 41st St.  
Tulsa, OK 74135  
(800) 950-3465  
(918) 627-1111  
fax: (918) 628-2768

Micro Design  
International, Inc.  
6985 University Blvd.  
Winter Park, FL 32792  
(800) 228-0891  
(407) 677-8333  
fax: (407) 677-8365

Micro Technology, Inc.  
5065 East Hunter St.  
Anaheim, CA 92807  
(714) 970-0300  
fax: (714) 970-5743

Optimem  
297 North Bernard Ave.  
Mountain View, CA 94043  
(415) 961-1800  
fax: (415) 961-8913

Pinnacle Micro, Inc.  
15265 Alton Pkwy.  
Irvine, CA 92718  
(800) 553-7070  
(714) 727-3300  
fax: (714) 727-1913

Pioneer Communications  
600 East Crescent Ave.  
Upper Saddle River, NJ  
07458  
(201) 327-6400  
fax: (201) 327-9379

Reflection Systems, Inc.  
P.O. Box 611608  
San Jose, CA 95161  
(408) 432-0943  
fax: (408) 432-0843

Ricoh Corp.  
Sales Office  
3567 Parkway Lane,  
Suite 150  
Norcross, GA 30092  
(404) 446-3533  
fax: (404) 447-4102

Sony Corp. of America  
Peripheral Systems Co.  
Sony Dr.  
Park Ridge, NJ 07656  
(201) 930-1000  
fax: (201) 573-8608

Storage Dimensions, Inc.  
2145 Hamilton Ave.  
San Jose, CA 95125  
(408) 879-0300  
fax: (408) 879-3397

Storage Technology  
Corp.  
2270 South 88th St.  
Louisville, CO 80028  
(303) 673-5151  
fax: (303) 673-5019

Summus Computer  
Systems  
17171 Park Row,  
Suite 300  
Houston, TX 77084  
(713) 492-6611  
fax: (713) 492-0092

Trimarchi, Inc.  
P.O. Box 560  
State College, PA 16804  
(814) 353-9120

Wang Laboratories, Inc.  
One Industrial Ave.  
Lowell, MA 01851  
(508) 459-5000

in banks, insurance companies, government agencies, corporate records departments, and multiple-site companies all require easily transportable storage equipment, sometimes above all other considerations.

Product quality is fundamental. This is basically a question of looking at the steps the manufacturer has taken to ensure reliable performance. These steps include user troubleshooting and fault-isolation diagnostic routines, the warranty of the mass-storage unit, the availability of information from the manufacturer or supplier, and postsale support.

Another critical issue is how well the device mixes and matches with other mass-storage devices that may already be in place in the computer system. Are the devices compatible with each other, and do their collective data capacities equal what you need?

The growth potential of the respective technologies is also worth looking at in terms of selecting a solution that you will be able to live and grow with in the long term. QIC's relatively low data densities and large available tape area make upward migration to larger capacities an ongoing evolutionary process.

The very high data densities of DAT permit using a small cassette with a small available recording area, making upward migration more difficult. The 8mm helical scan also has higher data density than QIC and a larger available record area than DAT.

Current QIC densities are about 0.6 MB per square inch, while 4mm DAT is about 3.8 MB per square inch on a medium area of less than 25% of QIC. QIC has more room to grow. Table 3 summarizes the current comparative situation for backup-tape technologies.

**Table 3. Backup-Tape Technologies**

Technology	Capacity	Transfer rate (bytes/sec.)	Form factor	Avg. access (sec.)	Backup time for 300 MB
QIC	250MB	90,000	5¼" HH	40	60 min.
	525MB	200,000	5¼" HH	30	25 min.
	1.00GB	300,000	5¼" HH	30	20 min.
	1.35GB	600,000	5¼" HH	36	10 min.
4-mm DAT helical scan	2.00GB	180,000	3½"	20	30 min.
8-mm helical scan	2.50GB	246,000	5¼" FH	484	21 min.
	5.00GB	500,000	5¼" FH	90	12 min.

Hard disk technology appears to achieve higher data capacities almost monthly, while rewritable optical disk capacities are moving at a slower pace.

Last but not least, there are the costs of the storage device and medium. The cost of a mass-storage subsystem varies greatly depending on the technology and manufacturer. However, cost increases with capacity. As important as cost is, it comes into perspective when you think about the possibility of losing your data due to a choice made on the basis of cost alone.

### Narrowed Choices and Trade-Offs

When you draw up your list of requirements How much capacity do you need? How much speed? What kinds of compatibility?—the choices of mass-storage systems narrow. There will be inevitable trade-offs involving price, functionality, and performance. A lower-cost unit will not display lightning-quick throughput speed; you may not be able to easily move a very high capacity system; or you may not be able to afford the most reliable system.

Optimizing your PC data-storage capabilities is a multistep process that involves many different considerations taking particular care to plan for growth. Nothing is cost-effective if it does not provide what you need. ■

# Contracting Fundamentals

## In this report:

The Negotiation Process .....	2
The Contract .....	2
Contract Administration .....	5
Summary .....	5

## Datapro Summary

Understanding the basics of contract negotiation and administration is imperative for project managers, since contracts are commonly a crucial point of liability, as well as a source of protection for the project manager's organization. Although it is not necessary to be proficient in the law field, it is helpful to be familiar with contractual terms such as warranties, indemnification, liquidation of damages, and others.

Although project managers are not normally trained in law, their day-to-day job activities require them to make decisions and take actions that can have significant legal impact for themselves and their companies.<sup>1</sup> Today some companies have established contract procedures for the negotiation as well as the administration of contracts. They routinely inform and train the managers of projects that may involve contracts about the details of these contracting policies. Fortunately, this is becoming a more popular practice. Frequently, however, project managers are handed a completed contract and are expected to proceed with little or no guidance until problems arise. The results can be extremely costly, time consuming, and embarrassing to the project manager and the company. Forward-thinking companies are finding it profitable to involve their project managers in the negotiation phase of developing a project's contract, particularly in those cases where performance to the letter of the contract constitutes the total project effort.

This Datapro report is a reprint of Chapter 15, "A Project Manager's Guide to Contracting," pp. 242-253, from *Project Management: Strategic Design and Implementation* by David I. Cleland. Copyright © 1990 by TAB Professional and Reference Books. Reprinted with permission.

Even in situations where the project manager conducts a project for his or her own employer, he or she is likely to be involved in one or more contracting efforts. In that case, however, the project manager would be more likely to be selecting contractors to accomplish activities, groups of activities, or work packages—that is, to perform some sub-set or portion of the project. Thus the manager might contract out the design of pipe hangers in a utility power plant construction project, or the development of safety and evacuation procedures for a nuclear power project.

In either of these cases the project manager must assume the role of contract manager with the primary purpose of assuring that contractors meet their obligations on time and on schedule, and perform the work to the documented specifications. Thus the project manager is likely to be involved in the contracting process in one way or another. While not a lawyer, the project manager must certainly recognize the legal aspects of his or her activities and be sufficiently aware of their implications to demand and obtain legal assistance where necessary.

How, then, can project managers best protect themselves and their company in these activities? First, they should recognize that the company's liability can arise in three principal ways: through warranties made either during negotiations or through

inclusion in the actual contract; through indemnification clauses or limitation of liability clauses in the contract document; or through the project manager's conduct in fulfilling the terms of the contract, or actions they might take which could be termed negligent.

This report reviews several methods project managers can use to protect themselves and their company during the negotiation of contracts and the conduct of projects. To perform effectively as contract managers, project managers must understand the practical aspects of three key issues: the negotiation process itself; the types of clauses which can be negotiated during the formation of a contract to limit liability and share risks, such as warranties, and the ways to fulfill the terms of the contract to avoid actions which could be termed negligent. Beyond that, project managers must understand the limits of their own knowledge and be prepared to obtain legal help when the situation demands more detailed knowledge of the law.

### The Negotiation Process

Negotiation is a process by which parties with differing interests reach agreement through a process of communication and compromise. According to Roger Fisher and William Ury in their book *Getting to Yes, Negotiating Agreement Without Giving In*,<sup>2</sup> there is a straight-forward, no-nonsense strategy for firmly pursuing your own interests while still getting along with those whose interests conflict with yours. This proven method of negotiation consists of four principles.

First, it is important to separate the people from the problems, realizing that the "other side" is also represented by people who get angry, frustrated, fearful, hostile and offended when they fail to interpret what has been said the way it was intended. Conversely, the "other side" frequently involves people who do not mean what you understood them to say. This misunderstanding may reinforce existing prejudices and fuel an endless cycle of actions and reactions until a solution becomes impossible. The process of negotiation is doomed when we fail to deal with the other negotiator as a person prone to human reactions.

Fisher and Ury's advice in this area is to first separate the substance of the negotiation from the relationship between the parties involved and to deal directly with the people problems which may exist. They suggest putting yourself in the other persons' position to see the situation from their point of view, and to recognize and understand their emotions as well as your own. Successful negotiators will refrain from deducing the intentions of the other parties from the perspective of their own fears. They will not blame other parties for their own problems but will discuss both sets of perceptions with the other parties, thus involving them in the process of communication. The successful negotiator will give credit for the good advice and ideas of others, and will frame the final proposal so that it is consistent with their values and seems a fair outcome for both sides.

The second point in successful negotiation is to focus on the common interests in the process rather than the opposing positions of the parties. The basic conflict in a negotiation normally lies in the goals of the parties being represented, and more frequently than not these goals can be reasonably well reconciled such that both parties win. The difficulty in accomplishing this is that negotiators frequently become over-committed to specific positions and defend those positions without examining other ways of achieving the overriding goals or objectives.

The difference lies between strategic and tactical thinking. The tactical negotiator will defend the detailed position. The strategic negotiator will look for other positions that may aid both parties in achieving their goals. Behind the conflicting surface interests of the two parties can often be found common areas which can be identified and used to maximize the goal accomplishment of both parties. Acknowledge the interests of the other party as a legitimate part of the problem, be concrete but flexible in the ideas offered as solutions, and be firm in dealing with the problem and yet open and supportive to the human being on the other side. This approach improves the relationship between the parties and increases the likelihood of reaching an acceptable agreement, as both parties attempt to attack the problem and not each other.

The third point Fisher and Ury make concerning successful negotiation is to generate a number of options and possible solutions that advance shared interests and creatively reconcile opposing interests prior to beginning the negotiation. Being creative under the pressure of the negotiation process and in the presence of the other party is difficult at best. At the negotiating table several options can be presented and the preferences of the other party can be solicited. The most preferable option can then be adjusted until a solution acceptable to both parties can be reached.

The fourth point is to insist that results be based on some objective criteria. Fisher and Ury point out that no matter how well you understand the interests or how ingeniously you invent ways of reconciling those interests, conflict will continue to exist. The best way to resolve those differences is to base the solution on an objective standard other than the will of the two parties involved. That standard could be the market value, replacement cost, industry practice, allocation of risk based on investment, or a long-term cost/benefit analysis.

These four points combined are termed "principled negotiation," a concept which involves developing an agreement based on the merits of the contract, as opposed to which party wins or loses. Successful contract negotiation requires practice and skill in the application of these tactics, but the benefits to be gained from the mutual respect and satisfaction that tends to accompany such an approach can provide inestimable value to both the company and the project manager. This kind of relationship frequently leads to add-on contracts and a continuing, profitable relationship between satisfied clients and contractors.

### The Contract

During the negotiation of the contract as a whole, separate clauses must be understood and incorporated into the final document to legally limit the liability and to allocate the manager's and the company's risk. The manager can assist in avoiding liability by understanding and using warranties, indemnification, liquidation of damages, and limitation of liability clauses.

### Warranties

The concept of warranty is based on the seller's assurance to the buyer that the goods will meet certain standards. A warranty imposes a duty on the seller who can be held liable by the buyer if this duty is breached. The buyer either can sue to recover damages or can rescind or cancel the agreement.



There are two types of warranties that are made when the seller enters an agreement with the buyer. *Express warranties* are promises actually spoken or written in the agreement, while *implied warranties* are promises the Uniform Commercial Code automatically includes in a transaction, whether they are written in the contract or not. For example, the implied warranty of *merchantability* says that goods or products must be reasonably fit for the ordinary purpose(s) for which they are used. This means the quality must be comparable to the quality of goods which would pass without objection in the trade, and that the product will perform safely according to the general standards of the industry. The second warranty implied in a transaction is that of *fitness for a particular purpose*. This warranty arises when any seller knows the particular purpose for which the buyer will use the goods, and where the buyer is relying on the seller's expertise in selecting suitable goods for that purpose.

In contracting with both express and implied warranties in mind, professionals can protect themselves by being very precise. When express warranties overlap the implied warranties, the express warranties will control, except for the warranty of fitness for a particular purpose. A good contract will contain a warranty giving a definite start and end for the period of time during which the promises are to be upheld. It will state a specific remedy of repair or replacement and describe the costs which will be paid by each party if specific circumstances or problems occur. And most importantly, it will specifically state the exact coverage of the warranty, specifying whether it deals with the design, the manufacturing, or the operation of the product, or possibly some combination of the three. Remember it can be as important to state what a warranty does not cover as it is to state what a warranty does cover.

Lawsuits against professionals and contract managers are becoming alarmingly more commonplace. The traditional standard applied to an individual's performance was based on negligence, i.e., performing with a reasonable standard of care. Thus, in order for negligence to be proven, it would need to be demonstrated that the manager failed to perform a duty (as a reasonable manager would) defined in the contract and this failure would have to be established in court. Even though the end result of the services failed to meet expectations, the manager might not have performed the service negligently. Thus, there might be no basis for a claim.

Until now, professionals in the majority of jurisdictions have not been held to the higher standards of strict liability usually reserved for those involved in the production of goods. However, in some states, recent cases have made inroads into the strict liability area on the theory that design professionals, for example, deal in an exact science, that exact results can and should be expected, and that a detailed professional report, design or plan is no different than the production of goods. As a result, these products should be governed by the same standards as those referred to in the production of goods.<sup>3</sup> In *Tamarac Development Co. v. Delamater, Freund & Assoc.*, 234 Kan. 618, 675 P.2d 361, 365, an architectural/engineering firm was held liable under a breach of implied warranty for failure to supervise and check the accuracy of the grading contractor's work for his conformance with the specific elevations. The original suit was brought on negligence charges, but was barred because of the statute of limitations. The plaintiff then brought suit and was successful in a breach of implied warranty action, which has the longer limitations period.

The contract negotiator should be aware that inserting warranty language into contracts for performance of services can have the effect of raising the standard of care. If services are warranted, they are guaranteed over and above the common law duty to perform in a non-negligent manner. Thus, if clauses are worded to require performance "to the highest standard of care," the contractor is in effect agreeing to increase his or her liability exposure.

### Indemnification

Indemnification is the act of making reimbursement to a person for a loss already incurred by that person. There are two general types of indemnification: common law and contractual. The common law concept of indemnification can best be understood by reference to a factual situation: A contractor is under contract to complete a project for an owner on the owner's property. Some person, perhaps a worker or even an innocent bystander or trespasser, is injured on the work site. Generally the owner has a high duty to any persons on his or her property and so the injured person will typically sue the owner to recover his or her expenses and possibly punitive damages. Under common law indemnification, the owner is entitled to recover any amount paid to the injured party from the contractor provided that the owner did not contribute to the injury. If the owner contributed to the injury, he or she has no right to recover monies from the contractor.

Because of the harsh results to the owner for any contribution to the injury, two developments have taken place. First, distinctions have been drawn between active and passive negligence on the part of the owner. *Active negligence* was defined as the owner's actions which directly caused an injury. *Passive negligence*, on the other hand, was identified as the owner's failure to act to prevent an injury. By virtue of these distinctions, the owner generally has been allowed to recover from the contractor if he or she was passively negligent, but not if he or she was actively negligent. In the second development, comparative negligence concepts adopted in many states allocate financial responsibility on a percentage basis of contribution to the injury. Because of these two concepts, owners wishing to allocate responsibility away from themselves frequently seek various contractual provisions in which contractors performing work for them agree to indemnify and hold them harmless against all claims, losses, and damages arising out of the work performed by the contractor. Similarly, contractors seek to obtain agreements from owners to protect themselves against claims by third parties.

Most states have statutes which limit to some extent contracts which seek to relieve a contractor of responsibility for the consequences of its own professional negligence. On the other hand, most states will enforce indemnity agreements by which the contractor agrees to indemnify the owner. Traditional indemnity requires the party seeking reimbursement through indemnity clauses to actually incur a loss before seeking reimbursement. In order to require the indemnifying party to defend against a claim from the outset, most contracts use the "hold harmless" language—"the owner shall be held harmless in all cases of injury arising from the operations of the contractor."

Indemnity provisions vary considerably from contract to contract as to the extent of the liability transferred. Some provisions may be so severe in their results that their legality hinges on whether or not they are contrary to public policy. These provisions can be categorized into three main classifications.

1. The broad form, the most severe of the three, obligates the indemnitor to indemnify and hold harmless the indemnitee against all loss arising out of the performance of the contract even if the indemnitee is solely responsible for the occurrence of the loss.
2. The intermediate form holds the indemnitor responsible for all claims or suits arising out of the contract except those arising out of the sole negligence of the indemnitee.
3. In the limited form, one party agrees to indemnify the other only for the claims arising out of the indemnitor's negligence.

In situations where both parties are jointly negligent, both are generally held liable. Under this type of clause, the parties are basically reconfirming their liabilities under common law.

Protection can be built into the indemnity provision for the contractor by adding a cap to the amount indemnified, limiting it to the same amount provided in the limitation of liability provisions (see "Liquidation of Damages" below). Additional protection for the contractor can be added by specifically wording the provision to indemnify third-party damages only. Without this language, indemnity obligations can be interpreted to hold the contractor liable for damages suffered by the owner himself. In negotiating these clauses, use reason and fair play with the other party. Consider the level of involvement in the work and allocate the level of risk accordingly. In negotiating third party claims, the contractor should not be expected to be responsible for those claims resulting from actions over which he or she has no control. It is important to remember that the indemnity provision survives the completion of the contract. This is especially important in cases of claims against a company by an employee who contracted an occupational disease and sues the company, which in turn sues the indemnitor or contractor. While insurance will usually cover this type of claim, the high dollar amount of deductibles required for this type of coverage may be prohibitive for many contractors.

### Liquidation of Damages

Frequently, parties to a contract wish to transfer liability for delays in completion of the work, including missing deadlines for the delivery of equipment and materials, or for failing to complete the project or any part thereof. If the contractor assumes liability for late completion or delivery, damages will be assessed in accordance with the terms of the contract. The damages specified usually are in terms of a specific amount of money to be paid for each day the work is late. Liability under these clauses can be limited in different ways depending on the circumstances surrounding the project.

The most common method, which should be included as a minimum, is to place a cap on the amount of damages to be paid, keeping in mind that excessive amounts may be found to be punitive and thus deleted from the contract by a court of law. In general, the amount of damages to be paid should be justified in some way by the amount of loss incurred by the owner as a result of the late delivery or completion.

A second way to limit liability is to include a *force majeure* provision. This common contract clause is used to protect the parties in the event that part of the contract cannot be performed because of an event outside the control of either party, or which could not be avoided by using

due care. A broad force majeure provision should also be included when providing for subcontractor delays as well. If total liability for subcontractor delays cannot be included in the force majeure provision, the contractor's liability for these delays should be dealt with elsewhere in the contract. This is frequently accomplished by structuring the contract so that damages are passed on to the subcontractors in amounts proportionate to the subcontractor's contribution to the overall task.

Third, if the contract calls for performance testing of the system prior to acceptance, the contractor should ideally have as much input as possible in determining the testing criteria. The contractor should also want to be involved in supervising the administration of the test within established deadlines for test completion and client acceptance. When the conditions of the test do not conform to the normal use of the product, performance testing becomes a means of sharing the risk between the client and the contractor. Under such nonstandard conditions, it is imperative that the testing conditions be carefully defined and in consonance with the contractor's understanding concerning the appropriate use of the product being provided.

A fourth method for controlling liability involves the careful definition of scheduled deadlines or milestones for each phase of the workscope. The contractor should carefully review all scheduled deadlines to assure they are reasonable given his capacity to perform. Ideally, the contractor will have proposed the major deadlines and milestones as part of the negotiation process. Such schedules should include a reasonable safety margin for dealing with potential problems in delivery.

A well-negotiated liquidated damages clause will adequately protect the contractor if liability is defined with reasonable limits. This type of clause also can be structured to provide an incentive bonus for the contractor who, instead of completing work after a scheduled deadline, completes the project prior to the designated date. Again, the bonus must be reasonably determined with a cap on the amount to be received in lieu of other compensation, and again it would be useful to justify the amount of the bonus and possibly the cap by the amount of savings and benefits that would accrue to the owner as a result of the early delivery or completion.

### Limitation of Liability

Limitation of liability clauses in contracts usually are the result of company-determined policies for all contract situations in which the company may be involved. Company policy usually requires the signature of an executive if liability is open-ended or if a limitation is over a set amount. Care must be taken, however, that the clause is tailored to fit the circumstances of the individual contract. Reliance on standard clauses lifted from the policy manual frequently leads to limitations which can be considered unreasonable in a court of law. If the limitation is found to be unreasonable, the company may be left open to a broad-scale liability, rather than to the limited scope initially intended.

When negotiating these limitation of liability clauses, the other party to the contract usually will be more accepting of the clause if it is introduced at the outset of the negotiation rather than added as company boilerplate terminology during the conclusion of negotiations. From the outset of negotiations, the project manager must be knowledgeable about the reasons behind the company's policy and the terminology used in the limitation of liability

clauses. The client's concerns with the clauses must be addressed so that they will be able to actively participate in the tailoring process which will inevitably ensue.<sup>4</sup> The project manager must understand both the policies and the terminology so he or she can simultaneously address the needs of the client and protect the interests of the employer.

The most successful strategy is to negotiate a commercially reasonable cap on the amount of liability based on the circumstances at the time the negotiation takes place. These circumstances include the fees charged by the company for the contract, the scope of general insurance coverage obtainable (considering the often high deductible), the nature of the work, and the resources of the client. The acceptance of a reasonable amount of liability also will reduce the likelihood that the clause later will be found to be either an exculpatory provision (that is, wording which unreasonably relieves the contractor of an obligation for the safety or quality of the product), or considered to be a penalty rather than a commercially acceptable allocation of risk. Either of these findings by the court would render the clause unenforceable. By the same token, a clause that shifts the entire responsibility to the contractor may not be enforceable. The wise negotiator will therefore carefully define the limitation of liability clause rather than hope that the courts will find a standard clause unenforceable.

The breadth or brevity of the clause being used should also be closely scrutinized, since any ambiguities will be construed against the drafter of the clause. The clause which is too short or which has been reduced to "plain English" may increase the risk of ambiguity. If the limitation of liability clause is too general, it may be construed to cover only one type of legal obligation but not others. It is important to limit liability on all possible legal theories under which a claim may be made. Project managers should understand that liability lies in active and passive negligence principles as well as in the more commonly recognized breach of contract, tort, and indemnification actions.<sup>5</sup>

## Contract Administration

*Contract administration* is the responsibility for supervising the work to be done under the terms of the contract, preparing and processing the changes that inevitably will be made, providing interpretation (with legal assistance if required) of contract language, and approving invoices as the work is being performed. Project managers need to be aware that in performing any of these responsibilities, conduct which could be deemed less than reasonable under the circumstances could give rise to charges of negligence against the manager, the company, or its employees by any other party involved in the contract. Such behavior also could lead to allegations of breach of contract. Normal company policy calls for executive-level personnel to have overall responsibility for establishing company contracting policy and ensuring that it is maintained consistently throughout all levels of administration. The negotiation, administration and control of all client contracts must be in accordance with this established company policy. To this end, it is frequently the project manager who must establish or implement the appropriate control mechanisms to ensure that the company's most current contracting and compensation policies are being executed.

Several brief but important principles of contract administration should be understood by the project manager.

Prior to issuing a contract, it is the project manager's responsibility to ensure that all required signatures, comments, and approvals have been secured or documented. No work should be performed before the final contract is issued or pending the contract, a formal letter of authorization to begin work has been received. This letter, of course, must specify precisely what work is to be accomplished pending the contract, as well as the obligation the company is willing to accept for completion of this work. The letter also should specify that the terms and conditions of the standard company contract expressly govern any work performed.

Another contract administration responsibility is maintenance of the contract files. As simple as this may sound to the uninitiated, the process of documenting the actions of the company and the contractor can become very complex. However, careful documentation also can be critical to the company if legal problems occur later. "Maintaining the files" includes carefully referencing all communications concerning the project with the proper accounting, project, and/or job number. It also includes keeping copies of the contract itself and all contract-related documentation, especially all changes, addenda, or supplements to the original contract, and assuring they are all properly cross-referenced. The project manager should make it a habit to record daily events in a permanent and orderly manner. This information usually is recorded on company forms which request information about problems that have occurred, the dates on which management became aware of the problem, and the steps taken to generate a solution. Additional documentation in a contract file generally includes the request for proposal, memoranda of any verbal communication concerning the contract, all internal review and approval documents, subcontractor status report sheets, verification of employee hours and overtime, quality assurance documentation, materials price and delivery verification, and records of timely and accurate billing and invoicing. Additionally, the project manager must ensure that the cumulative services and billing for those company services do not exceed the approved scope and budget for the client.

Of particular note is the problem of changes that occur during the conduct of the contract. It is common practice in industry for contractors to "buy in" to a job: i.e., to bid well below what they expect their costs to be for the contract knowing that they can "get well" by pricing later changes at their own discretion when the company is "locked in" to their services. Assuring that a clearly understood and well-documented change control process is defined and in place can be absolutely critical to the successful completion of the contract. The clear documentation of all changes, the price changes involved, and the authorization to proceed must be included in the contract files, as this is a prime area for litigation.

## Summary

Understanding techniques of allocating risk and implementing them while negotiating and administering contracts is a major factor in being a successful project manager. Contracts are frequently a critical basis of liability exposure as well as a major source of protection for the project manager's employer. Every project manager who conducts or administers contract work should be able to identify the principal bases of liability exposure in the contracts they are likely to encounter. They should understand how to construct clauses that reasonably limit the liability

exposure of the company they represent, and when to get legal help because the complexity of the issue has exceeded their knowledge or level of responsibility. Once these sources of liability have been recognized, a basic knowledge of successful negotiation strategies will enable them to develop contracts in conformance with the policies of the company they represent.

Finally, in assuming the role of a contract administrator or contract manager, project managers who have an understanding of practical contract management issues will be able to recognize the importance of carefully documenting the performance of the activities which they supervise. Most important, they will understand the value of formal legal assistance in creating contracts which will avoid costly and lengthy legal settlements, which are at best disruptive to themselves and their employers.

---

## References

<sup>1</sup>This report was prepared by Dr. MaryAnne F. Nixon, an assistant professor of business law in the School of Business, Western

Carolina University, Cullowhee, North Carolina. Dr. Nixon is an attorney who consults and conducts seminars on the legal aspects of project management. She is a feature editor of the Legal Issues column for the *Project Management Journal*.

<sup>2</sup>Roger Fisher and William Ury, *Getting to Yes—Negotiating Agreement Without Giving In* (Boston, Mass.: Houghton Mifflin Company, 1981).

<sup>3</sup>In *Vincenzo v. Trus Wall Systems, Inc.*, (a case before the Connecticut Superior Court), the court held "Trus Wall did 'manufacture' the design it sold to Universal Builders in the sense that it created the design and put it into a form from which the actual product took shape. . . The definition of 'product' should not be so narrow as to exclude a design."

<sup>4</sup>Roger S. Mertz, "How Can I Limit My Monetary Liability by Contract Clauses?" *Avoiding Liability in Architecture: Design and Construction* (New York: John Wiley & Sons, 1986), 279.

<sup>5</sup>*Ibid.*, 281. ■

# Evaluating Potential Consultants

## In this report:

The Evaluation Process .....	2
Evaluating a Consultant's Skills.....	2
Measurable Objectives.....	3
Experience and References .....	3
Client Fears .....	4
Consultant Fears.....	5
The Consultant's Marketing .....	5

## Datapro Summary

Evaluating the qualifications and suitability of potential consultants can be a challenging and time-consuming task. By properly assessing consultants and providing clear objectives, you can have a productive and rewarding working relationship with the right consultant.

The degree of precision with which you evaluate the suitability and capability of the potential consultant is likely to be an important, even determining, factor in your ultimate satisfaction with the results achieved from the consultation. Over the years, I have heard numerous complaints from clients regarding the quality of services provided by consultants. In most cases such dissatisfaction can be directly related to factors that could have been determined in the course of careful preconsultation evaluations and interviews.

The reader might well ask, then, why, if danger signs were apparent and even obvious in the first place, were they not discovered? There are, undoubtedly, many reasons, but the most significant is the imprecision or sheer laziness with which the client approaches the task of evaluating his or her own needs and the suitability of a potential consultant. If the client fails to fully understand the precise outcomes the consultant must produce, the result may be dissatisfaction and a counterproductive expenditure of funds. It also provides clear

evidence that the client does not fully understand his or her own problem or comprehend the full range of needs to which a response was desired.

In reality, many consultations could be avoided (or reduced in scope), saving considerable dollars, if clients were disciplined enough to think their needs and problems through with sufficient precision to identify the ways in which a consultant's services are really needed. Further, such precise analysis will often serve to identify a more specific or narrow set of skills required from a consultant. This can be advantageous in that the assessment of a narrower set of skills may result in the client's ability to obtain a professional at lower cost or to avoid the expense of planning and elaborate needs analysis expenditures often associated with consulting. One consultant with whom I am familiar often says, "I wish my large corporate clients were more like my smaller entrepreneurial clients." This is a very telling statement. Like many consultants, this one finds that working with a precise, no-nonsense, frugal client is often more productive than working with casual, often undisciplined, free-spending bureaucrats. In short, more corporate clients would be well-served behaving as if the money they spent were their own, not someone else's.

Successful use of consultants often requires as much work on the part of the client as it does on the part of the consultant, at least in terms of planning and strategy. Consultant and client must understand the need, the objectives, the limitations, and

This Datapro report is a reprint of Chapter Six, "Evaluating Potential Consultants," pp. 43-59, from *How to Select and Manage Consultants*, by Howard L. Shenson. Copyright © 1990 by Howard L. Shenson. Reprinted with the permission of Lexington Books, an imprint of Macmillan, Inc..

the nature of the results to be achieved. This can only occur when both parties have a shared understanding of all the parameters related to the consultation. The more the client is able to provide precise, even measurable, objectives to the consultant, the greater the likelihood that the consultation will produce the desired outcomes in a cost-efficient fashion. And contrary to conventional wisdom, though certainly exceptions do exist, most good consultants do not seek to build continuing client dependency on their services. Like good clients, good consultants are busy, over-committed, and not in need of more ways to spend their time. With proper evaluation on the part of the client, limited experience with a consultant can be productive and rewarding.

### The Evaluation Process

How, then, do you evaluate a potential consultant? How do you find out whether someone is the right consultant for you? There are many ways. I think that the best way is to use what I call a "here and now" basis of evaluation—meet your potential consultant in person.

You must schedule an interview with the prospective consultant. When you call to make an appointment, briefly describe your problem or reason for seeking consulting services and make clear that you are seriously seeking a suitable consultant for your needs and would like to set up a personal interview. In a sense, evaluating a potential consultant is no different from evaluating a prospective employee. You need to sit down face-to-face, ask the hard questions, listen for the right answers, and reject those people who do not provide them. I wish I could tell you there is an easier way of evaluating consultants, but there simply is not.

The operating style of consultants is an important factor in evaluation, too. In addition to their technical know-how, ideas, concepts, and approach to the problem, consider how they operate. There are consultants who are too creative, too entrepreneurial, too shoot-from-the-hip to be comfortable with a stodgy bureaucratic organization. There also are consultants who are too bureaucratic themselves, too conservative to suit a creative, innovative organization. A mesh between the personality or management style of the consultant and the personality and management style of the client is essential. If they lack compatibility of style and approach, they are likely to have difficulty working together to produce the necessary results.

### The First Meeting Between You and the Consultant

The first face-to-face meeting between client and consultant can be a productive, rewarding, and instructive session. To be effective, however, the client must come to the meeting with full understanding that this is a meeting between equals and peers and not a meeting between superior and subordinate. Also, the client must be aware that the consultant's time is of equal value to the client's and that the consultant (read "a worthwhile consultant") has as much right to reject the client's business as the client has to reject the consultant's services.

With this environment of mutual respect, the first meeting allows the parties to get to know one another and to determine whether they will find it of mutual benefit to work together. While it is obvious that you will interview the consultant, a good consultant will also interview you. You should be a little suspicious if he or she does not. Any consultant who does not find out what you want, what you

have already done relative to your needs and problems, and what business arrangements are to your liking is probably too hungry and too willing to be a good consultant.

### The First Four or Five Minutes

The first four or five minutes of the first meeting between the client and consultant may be crucial, but be cautious. You may well reject a qualified person within the first four or five minutes simply because he or she does not give a good interview. That will be your loss because someone who might have been very qualified leaves without ample consideration on your part. Clients need to remember that many consultants are not as adept at salesmanship, even subtle salesmanship, as others. In general, consultants prefer doing their work to marketing their skills. In a day and age when the client encounters very polished presentation tactics, it is important to look beneath the surface for substance.

It is unlikely however, that you are going to make the mistake on the other side of the coin. That is, if you invite a consultant to meet and talk to you, and he or she sounds suave and sophisticated, organized, and confident in the first few minutes, you will keep the consultant talking. You will perhaps talk to him or her two or three more times, but if you are at all good at interviewing and observing people, you will figure out that this person is not qualified or not the right person for your job. You will have wasted your time, but you probably will not have made a mistake.

So, while the first four or five minutes are crucial, I think they tend to be so only in the sense of possibly causing you to dismiss someone who is very qualified because you do not allow him or her adequate time to demonstrate skills and competency.

### Evaluating a Consultant's Skills

Next, you will find questions that could be asked of consultants during the initial interview—questions, that, I have found very quickly help to discover if the consultant is qualified. You will undoubtedly want to modify these questions to some extent, but nonetheless, the answers you will get will be useful. And you will discover that retaining a consultant is not very different from retaining an employee in terms of the questions you ask and the criteria you assess. You must always rely on your ability to evaluate people and decide whether they are honest, come up with ideas that make sense, and make proposals relevant to your problem and needs.

### Seven Quick-Check Questions to Evaluate a Consultant's Skills

1. What do you regard as our principal need or problem?
2. What can you offer us that other consultants we have interviewed/your competitors have not been able to provide?
3. How will we measure or evaluate your success in meeting our needs/solving our problems?
4. Are you willing to work on a performance basis—that is, to be compensated on the basis of the results you produce?
5. What related experience have you had in working with organizations similar to ours or with other organizations in this industry or field?
6. What related experience have you had in working with needs and problems similar to ours?

7. Who may we contact as a reference about the services you provide?

#### Additional Questions to Ask Yourself and/or the Consultant

- How can I profit?
- Why can I profit?
- Where can I profit?
- Who says I will profit?
- What will I profit?
- When can I profit?
- Do I need this service?
- Do I really want this service?
- Can I really afford this service?
- Will I make use of the outcomes?
- Am I being given a good deal?
- Should I check out the competition?
- Could I get this service for less?
- Is this consultant honest, reliable, and right for me?
- Is the consultant knowledgeable?
- When do I need to make my decision?

#### Outcomes and Expectations

Most certainly in this first meeting there will be a discussion of the outcomes and expectations of the consultation. What is this consultant expected to produce? You are going to get a better performance and better results if you can tell the consultant, as specifically as possible, what a consultant must accomplish or produce to satisfy you.

If you do not tell the consultant this, he or she should ask. One of the signs of a weak consultant, in my mind, is when the consultant fails to ask what the client wants to accomplish, if the client has failed to communicate these goals precisely. If all the consultant wants to do is sign the contract, get under way, and figure out later what has to be done, you are likely to encounter problems.

However, often a client decides on the nature of his or her problem or need before ever calling a consultant; once this decision is made, a client may tend to become committed to it. Then, if the consultant discovers that the problem is different from the client's perception of the problem (and such is frequently the case), there can be some conflict in the relationship as the consultant attempts to change the client's thinking or behavior about the particular problem. This ego conflict may also work the other way. Once the consultant has done a formal needs analysis—a diagnostic—for the client, the consultant may become very committed to it. Consultants should be able to bury their egos for the benefit of the assignment or the client, but they are human (though some people would dispute this) like everyone else. So the relationship may become strained if the client disagrees with the consultant's needs analysis. The first meeting will be greatly enhanced if both client and consultant are not overly committed to their previous thinking. Moreover, a good consultant should do the needs analysis in what I would call a "real time" fashion—as needs are discovered, as the situation is understood, a good consultant should confirm these with the client so that the final report of the needs analysis does not come as an absolute surprise to the client.

#### Measurable Objectives

Whenever possible, the client and the consultant should work together to establish tangible, measurable, observable criteria—things like dollar profit, sales quotas, number of units produced, and number of people who pass the test at the conclusion of a training program as a result of the consultation—to serve as the basis of evaluation. However, frequently there are no tangible factors appropriate for evaluation, in which case, the second best thing to do is to arrive at some proxy measurement.

For example, if we are going to bring in a consultant to train teachers so they will be able to give their students better vocational job placement and better work experience later on, how will we know when we have been successful? We may have to wait five to seven years to find out if, in fact, the students do get better vocational placement. Obviously, we cannot wait five to seven years to determine whether the consultant should be paid; payment has to come out of this year's budget. So we may develop some kind of a proxy measurement. We may just find a test instrument and if the teachers can pass the test, then we will believe they have the right attitudes, the right skills, and the right knowledge to be able to bring about improved performance on the part of the students.

That is a proxy measure. When a proxy measure is inappropriate, we can, of course, resort to our old standby—gut feelings and instinct. "I am happy. I am satisfied." What you and the consultant should discuss very precisely before the consultation is what you, the client, will have to see to assure you and satisfy you that the consultant has done the job well. Now perhaps that is something very specific: "We must have a 14 percent increase in profits or you fail," or maybe it is, "I just have to feel better." Although it is difficult to say in the abstract, strive to make your goals as tangible and specific as possible; if that is not possible, find a proxy if you can.

#### Business Arrangements

It is important for both parties to discuss the nature of the financial arrangements and the contractual terms and conditions. Consultants collect their fee in a variety of ways—by the day, by the hour, fixed-price, and so on. Each method of payment results in the client taking more or less risk. A solid understanding of each party's risk-taking propensities and the ways they might work together should be an integral part of the first meeting.

#### Experience and References

Once you are satisfied that the consultant's here-and-now answers to your questions are appropriate, that the consultant's operating style is suited to your organization and your particular needs—indeed, that this may be the right consultant for you—you will certainly want to inquire into his or her experience and may well want to check some references. You may also ask for and check the consultant's references before the first meeting. Perhaps after checking references, you will already know this consultant is not right for you, in which case you need never have a face-to-face meeting. However, it is probably no wiser to dismiss a potential consultant whom you have never met on the basis of references alone (except in the very rare case that all the references are negative) than it is to retain a consultant after a first meeting without ever checking his or her references. It is at your discretion whether you ask

for references before the interview or during the interview, but do check into them. You will find the results well worth your time and effort.

There are different kinds of references consultants can give you. There are supplied references, which are what I have when I reach into my briefcase and take out a typed sheet of names, addresses, and telephone numbers in response to a client's request for references. I say, "Here they are. Please feel free to contact any of these people." This is the least useful form of reference to a client because obviously the consultant is not going to list anyone who would say anything negative or critical. This does provide a way of checking out the consultant, but not a great way unless you happen to know personally and professionally one or more of the people on the reference list, and you think you will get a totally straight answer and honest opinion.

Another form of reference is what I would call a peer reference: talk to other consultants who may know the consultant whom you are considering. This is not at all an uncommon practice. Consultants tend to know one another to some extent, or to have heard of one another, or at least to know third parties who know both of them. And consultants are fairly good judges of other consultants; moreover, unlike doctors, they are not nearly as closemouthed about the problems and the limitations of their peers. So you may well find it within your interest to talk to other consultants.

Perhaps the best source of reference is former clients of the consultant. If the consultant is willing to tell you who the former clients are, or if you know who the former clients are, this is an excellent place to check references. Many consultants, however, and properly so, do not reveal the names of clients. They regard their clients and the work they have undertaken for them as confidential. You should pay careful attention to what a prospective consultant says about past clients. If the consultant lacks discretion and becomes your consultant, he or she may be indiscreet regarding you and your organization.

There are three other evaluation procedures that may be quite useful. The first is to research the consultant. Examine what he or she has done, for whom, and where. Talk to people who know the consultant; that is, create your own list of references. Second, evaluate the consultant's contributions to his or her field. What has the consultant written and/or researched? Where does he or she speak? With whom in the field does the consultant keep company? Heads of trade and professional associations and scholars in the field may be helpful resources. Third, take a good look at who is running the consultant's practice. How much in control of his or her destiny is the consultant? A consultant who is too complacent is likely to be unsuccessful and not very busy. When I retain a consultant, I want someone who is in control, intelligent, innovative, assertive, and who knows what he or she wants and at which points he or she plans to make a stand.

## Client Fears

In a survey of 610 clients, I discovered that there are specific fears clients have about using consultants. These are presented next in priority order. The first meeting is an excellent place for you to ask questions to determine whether you have these fears about a particular consultant.

### Consultant Incompetence

The most prevalent fear is that the consultant is incompetent; he or she looks and sounds good, but really is useless.

Have the consultant give you sufficient information about his or her knowledge, skills, and experience for you to rule out the fear of consultant incompetence.

### Continuing Dependency

Clients fear they will get hooked on consulting and will continuously need help, never being free of the consultant. You need to press the consultant for ways in which the consultant will make this a turnkey activity. Find out how he or she will turn things over to you and train and inform you and your staff so that you are not continuously dependent on the consultant's services.

### Lack of Managerial Control

Clients also fear the consultant will be so powerful and so overwhelming and so much in control that the client organization will lose control of the situation. Management is not and should not be the consultant's role. It is not the consultant's responsibility to make decisions; that is management's responsibility. The consultant's role is to advise management about decisions to be made and to provide documentation, evidence, support, and recommendations, without taking away decision-making authority from the client. So make sure you and your consultant understand these essential distinctions.

### Excessive Fee

If you fear the consultant's fee is too high, have him or her justify for you exactly what will be done. The consultant should be willing to break down the cost and show why the services cost what they do, until you are comfortable with the cost analysis. You may discover that the fee seemed excessive because you have said certain things about the project to the consultant or the consultant has made certain assumptions about the project that are not in fact accurate.

You may also find that you and the consultant have different expectations about the quality of the result to be produced. Consultants, whenever possible, like to deliver Rolls Royces to their clients. But some clients prefer Pintos. If that fact is not communicated, the consultant will seemingly be charging too much money for the services the client is expecting, and the client will be unhappy.

If the client is the Department of Defense, and it is trying to figure out where to put antimissile missiles, then I would expect it to want to have about 99.9% accuracy in its decision. It is going to pay a great deal of money to get 99.9% accuracy. Getting 90% certainty about a decision carries a substantially lower cost than buying that extra 9.9%, which has a proportionately much greater cost. If, however, the client is the City Parks and Recreation Department, and it is trying to figure out how many people used the park this summer, then perhaps it is willing to live with 80% certainty and considerably less cost; it is not as vital that it buy 99.9% accuracy.

You must make clear to your consultant what level of response, what level of result you are looking for. If you do not communicate this information, the consultant should determine this information from you, but sometimes consultants make assumptions. Do not allow this to happen. To alleviate excessive fees, specify to the consultant what level of quality and expense you deem desirable and necessary, and find out from the consultant what agreement or what discrepancy exists between your needs and the consultant's needs.



### Time Availability

Many clients fear the consultant will have insufficient time to complete the client's project on time. Therefore, have your consultant justify how much time the consultation will take, when the project will take place, where the major milestones will be, and when the project can be expected to be completed. Agree to what will happen, if anything, if it is necessary to adjust the time schedule.

### Evidence of Failure

Sometimes clients are afraid that the need for a consultant will suggest to others in the industry, or in the client's organization, that such a need is a failure on the part of the client. This is rarely accurate or true. It should be the good consultant's job to ensure that the client will be protected from any embarrassment, that the consultant will work behind the scenes in a manner supportive to the client's self-interest, and, further, that the consultant will place his or her ego backstage to make the client look good.

### Disclosure of Proprietary/Sensitive Data

Some clients fear that in using consultants they may be disclosing sensitive or proprietary data that could, in the wrong hands, damage their interests in some way. For this reason, a consultant's discretion is very important. Any sign of indiscretion is perhaps a sign that you cannot release sensitive material to the consultant. You may even wish to have your consultant sign a disclosure statement that says the consultant will not release confidential data. Although the disclosure statement itself will not prevent the consultant from being indiscreet, it will make this concern more prominent.

### Improper Diagnosis/Needs Analysis

Low on the list of fears is the concern that either the client or the consultant may have improperly diagnosed the need for the consulting services, resulting in wasted time, effort, and money. Thus, it is important that both parties confirm the needs analysis/diagnosis, regardless of which party originally undertook the analysis. Whenever possible, needs analysis and problem definition should be done by client and consultant, working together.

### Lack of Impartiality

And finally, the last of clients' fears about making use of consultants is that the consultant will lack impartiality. The consultant has, as a requirement and responsibility of being a professional consultant, an obligation to serve only the client's best interests. The consultant should not be beholden to any other interests whatsoever or have any side deals, commissions, kickbacks, referrals, or the like, which could cause the consultant to lose objectivity—the true value to the client.

Therefore, examine your consultants. Scrutinize them to make sure they are not tied in or influenced in any way that might cause them to make decisions other than in your best interests.

### Client Fears Alleviated

While you may not completely eliminate your fears about retaining a consultant (and a certain measure of caution is not unwise), by evaluating your potential consultant, communicating your needs and concerns, and questioning and listening to the consultant's answers, you should be able to alleviate most of your fears. Of course, if after carefully evaluating the prospective consultant you still have considerable wariness about hiring him or her, then it is probably time to move on to another.

## Consultant Fears

The most common consultant fears are that the client will cheat the consultant and not pay, that the consultant will get bogged down in the client's organization because of an inability to define the problem and agree to the scope of work. Consultants also fear the client won't provide adequate support to the consultant, or that the people with whom the consultant has to work in the organization will be counterproductive to what the consultant is trying to achieve. As a client, it is your responsibility to make sure that your consultant need have none of these fears. Even if the consultant does not communicate any fears to you (and chances are, the consultant will not), you should be prepared to define objectives, to provide support and a supportive staff, and to pay fully for satisfactory services rendered.

## The Consultant's Marketing

The consultant expects it will take, on the average, about a day of his or her marketing time to sell five days of time. Thus, if your consulting project involves fifty days of the consultant's time, it may well be that in the process of proposal writing, interviews, and phone calls, the consultant could spend up to ten days' time in the marketing stage. That is a considerable amount of meetings and a considerable amount of getting to know one another. In ten days' time you get to know someone fairly well, so if in ten days you have not figured out whether this person is competent, you are probably never going to figure it out, not even after the consulting work is done.

The longer the consultation, the smaller the percentage of time usually required in marketing. The reverse is also true: many consultants will tell you that it is just as difficult to sell a twenty-five-day assignment as it is to sell a ten-day assignment. One of the factors that adds to the consultant's marketing time, of course, is a client's lack of precision and direction. A proposal I completed while writing this report illustrates the point. A Fortune 100 company flew five of its executives to Los Angeles for a meeting with me at an airline club at Los Angeles International Airport. Despite my constant efforts to pin them down about objectives and desires, it became apparent to me that they had no inclination to be specific. Either they had not thought through their objectives and needs with sufficient precision to be specific, or they were testing my ability to be expansive and creative. We left the meeting with my commitment to send them a proposal to outline "how my services would be of value."

I know of no more difficult proposal to write. The information was too vague. For two weeks I did no writing but thought (mostly subconsciously) about what my proposal would say. In one day, it flowed out. I proposed twelve "mini-projects" that would serve their needs. The response took less than a week (rather spectacular turnaround for a large company), and the outcome was favorable: four of the specific projects were agreed to with certainty, with four others highly likely in the second year. As many consultants have learned, they often have to stimulate the client's thinking with a potpourri of interesting ideas to focus the client more specifically on desires, needs, and objectives.

Should you expect to pay for the consultant's time during the marketing phase? You will, whether you expect to or not. You will pay for it indirectly in the form of overhead after the contract is awarded, or you will pay directly.

Only a small percentage of consultants, however, charge directly for their marketing time. If it is not a direct charge, the consultant will treat marketing as an overhead expense and will charge all future clients indirectly. So yes, the clients will pay for it. Someone has to pay for marketing.

It is not at all uncommon for one federal agency I know of to cause an expenditure of \$150,000 worth of consulting time in proposals and marketing to award a \$30,000 contract. The agency contractors go to fifty people asking for proposals. Each spend approximately \$3,000 to respond to the government's R.F.P. That makes \$150,000 that has been spent to get a \$30,000 contract that is awarded to one person. This is one of the reasons consultant's fees tend to be somewhat high. Shopping is much more common in the public sector than in the private sector, where few clients shop. A statistic that amazes me is that in 83% of all consultations, the client never gets a second bid. In 68% of consultations, the client never even talks to a second consultant, meaning 68% of the work is awarded without the client talking to anybody else or getting any alternative bids. This amazes me as much as a similarly high number of people who will not shop two dealers when buying a car. Everyone thinks that people shop for cars. They do not. They go to the first dealer and buy a car. I would suggest that you do not go to the first consultant and buy consulting services without shopping around, and by this stage in the consulting process, ideally you will have searched and evaluated sufficiently to have found the perfect consultant for your desires and needs.

### Summary

Taking the time and effort to evaluate the qualifications, capability, and suitability of potential consultants can be a daunting task, yet the results of careful evaluation—finding the right consultant for you—can be immeasurably valuable. As you have learned from this report, to discern whether a prospective consultant meets your needs, desires, and style, you need to:

- Conduct a personal, face-to-face interview.
- Ask essential questions to establish the consultant's qualifications:
  - What does the consultant regard as your principal need or problem?

–What can he or she offer you that competitors cannot provide?

–How will evaluation of success take place?

–Will the consultant work on a performance basis?

–What experience does the consultant have working with organizations similar to yours, or other organizations in the industry or field?

–What experience does he or she have in working with needs and problems similar to yours?

–Who are the consultant's references?

- Discuss with the consultant the desired outcomes and expectations of the consultation.
- Establish with the consultant measurable objectives for the consultation.
- Discuss financial arrangements and contractual terms and conditions.
- Assess your own fears—fear of consultant incompetence, continuing dependency—lack of managerial control, excessive fee, time availability, evidence of failure, disclosure of proprietary information, improper diagnosis/needs analysis, lack of impartiality—with the consultant, determining if they are increased or alleviated.
- Alleviate the consultant's potential fears.
- Check references.
- Shop around. Do not hire the first consultant who seems to meet your needs without evaluating the competition.

Consider all these factors when evaluating your potential consultant. Most of all, listen to your gut feelings and instinct: if a consultant seems to have all the right answers, impeccable references, and excellent experience but you somehow feel he or she is not right for you, you are probably correct; if a consultant seems a little unsure, unpolished, and inexperienced, but you are convinced of the quality of his or her abilities and feel you would work well with this person, then you are also probably right. ■

# Buying Third-Party Network Support

## In this report:

Software Support.....	2
Watching the Clock .....	2
Response Time .....	3
Buying Network Support: A Shopping List.....	3
Contracts.....	4
Document the Network.....	4

## Datapro Summary

Organizations spend too much money on planning and implementing LANs not to acquire comprehensive support and service. When you are in the market for a service provider, let flexibility in packaging be the guide. The customer should note the service provider's pricing, software support, maintenance and repairs, spare parts, and contractual policies.

Maintaining a network often requires complementing your own expertise with outside help. Buying network support services resembles buying a suit. You can buy a suit off-the-rack and have it fitted, or you can have a suit custom made. A service provider can nip and tuck an existing support package, or it can craft a support program to perfectly fit your needs. Here are tips on how to choose a service provider and what your support package should contain.

## Buying Support

Factors, such as network size and complexity as well as business applications, will determine how much and what kind of support you should buy. Some companies want cradle-to-grave coverage on every network component; some choose to pay as they go for maintenance and repairs. When you're buying service, look for a company that offers you flexibility.

The most important test is the service provider's primary business. Look for a company whose business is networks, not a PC shop that wants to expand its revenue base. "You just can't put a network at risk

with a third-party organization that's not dedicated to networking," says George Christian, vice president of service marketing for BancTec (Dallas), which has offices in 151 cities. "To fix the hardware is generally straightforward, but to bring the network back up and recover the data and what was lost during a crisis is something else."

When buying support, "I caution network administrators to look at this from a business perspective, not a computing perspective," says Leo Spiegel, executive vice president of marketing and business development for LAN Systems (New York), a systems integrator with eight U.S. locations, including San Diego. For example, a law firm should weigh the potential for lost revenue and missed deadlines if its network goes down and its attorneys can't access files against the cost of support, not how much the LAN cost to install.

The decision to buy hardware support contracts or pay as you go should be based on risk-benefit analyses. You need to assess the risk of being self-insuring, says Herb Klein, general manager of the network management services center for Network Management Inc. (NMI, Fairfax, VA), a national systems integrator.

Klein cautions companies to consider what the failure of the system would cost them, not what the system cost. "In many cases, people buy too little support for very, very critical business items," he says.

This Datapro report is a reprint of "Suitable Service" by Elizabeth Dougherty, pp. 45-52, from *LAN Magazine*, Volume 6, Number 3, March 1991. Copyright © 1991 by Miller Freeman Publications, Inc. Reprinted with permission.

A company's need for outside help with its LAN services, such as design and ongoing maintenance, depends on where a network is in its life cycle. A company initially may need more help, but this need for maintenance support typically levels out as time goes by and the network stabilizes. "Often you'll start out with a lot of support services, and as you gain expertise, you'll bring more and more services in house," says Nina Burns, principal with Network Marketing Solutions (Menlo Park, CA), a market research firm.

When you're buying a LAN support contract, you need to decide what hardware and software components to cover. With a network you'll want to look for a service provider that offers expertise in LAN software support. You need a company that, for example, can fix your file server, reinstall the network operating system, restore your data, and get the network—not just the server—up. Make sure a contract for a file server provides this type of software service.

Besides the file server, you should consider buying service contracts for critical components such as bridges, routers, hubs, backup systems, UPSs, gateways, and communications servers. If an essential component fails, the network may be brought to a standstill. Also keep in mind the most likely points of trouble. Cabling problems, for example, account for approximately 95% of LAN problems, says Mark Cuban, president of MicroSolutions (Dallas), a large systems integrator.

For less expensive or less critical components, such as workstations and printers, you might be better served by having spares and swapping those in while others are serviced.

## Software Support

Whether you buy software support depends on your management philosophy and level of in-house expertise. "Companies want a mix between what they provide themselves and what they outsource," says Burns. "With software, they want to build expertise in-house." Burns says the cons of relying on outside software support include the financial outlay, the loss of developing the expertise in-house, and the lack of control.

## Software Issues

With software, you should think about whether you want help with repairs and upgrades or help in learning and using the applications. For example, a contract might cover support under the normal operation of software, which assumes a certain level of knowledge. You might not get support "that takes care of a lack of expertise," NMI's Klein says. A contract also should state what customer-installed software is covered.

A company should consider hiring someone to operate a help desk to provide onsite or telephone assistance with PCs and applications, suggests Klein. "Many internal resources go toward solving PC problems," he says. "A company can cut costs by buying help desk services."

Upgrades are another software issue. Decide what your company policy will be on upgrades, and verify that your service provider and service contract are in sync. If you're not going to buy upgrades realize that vendors concentrate their support and training efforts on the latest versions.

Think, too, about preventative maintenance. Does your contract cover periodic checkups? Often diagnostics can

uncover problems before they bring the network down. Include in the service contract a schedule for preventative maintenance.

A service provider may be in a good position to suggest ways to fine-tune your network. MicroSolutions performs network audits that search for problems and better ways to do things. The cost is \$50 per workstation \$500 per server, with a minimum charge of \$2,500.

Determine how services are priced and what your payments specifically cover. Does the provider offer fixed annual rates, hourly rate, special bulk discounts, or a combination? How flexible are the packages? In determining which price structure is best for you, consider the type and frequency of service that you require.

A fixed-rate sum functions much like an insurance policy. You know what your expense is going to be, and you can budget for it. If your equipment needs repairs, the cost is covered. If nothing needs to be fixed, you have peace of mind that if it had, you would have been covered.

Annual fees typically are based on a percentage of the purchase price of equipment or on the type of equipment. Often you'll be charged per node or file server.

LAN Systems prices its contracts as a percentage of the retail price of the hardware. MicroSolutions typically charges \$3,500 to \$5,000 per server per year, which includes 10 hours of onsite service per month, upgrade installations, and unlimited telephone support.

BancTec prices annual contracts on the amount of hardware and number of users. "The big problem with 'time and materials' work is readiness," says Christian. For example, if you pay \$5,000 for a service contract on your file server and nothing happens to it, you might feel like you didn't get anything for your money. "That's not really true," he says. That fee "insures that a trained technician and parts are available. If [something happens and] you don't have parts, that extends your downtime. Over the long run, you pay a premium."

Hourly rates can be paid on an as-needed basis or sold at a prepaid bulk rate, which may include the service provider placing a higher priority on responding to your service requests.

## Watching the Clock

Be careful in estimating how many hours of support you will need. "Hourly fees tend to run up quickly," points out Burns.

Winson Olson & Co., a network systems integrator based in Santa Ana, CA, offers its clients the option of a fixed-rate service contract for hardware or support contracts in groups of bulk hours that apply to on-site or telephone support. The company recommends the hourly option to its clients. "It makes good business sense," says Dale Winson, executive vice president. "Clients pay for the services that they receive, and we get paid for the services that we provide."

Winson Olson prices annual service contracts that cover parts, labor, and maintenance at approximately 15% to 20% of the purchase price of hardware. "People tend to buy these more traditional service contracts," Winson says. "I don't think it's really a good deal for the customer for what we charge for them. Someone always wins and loses with services contracts."

Evernet Systems (Los Angeles), a network systems integration company with 15 U.S. offices, offers fixed price, repair-or-replace hardware contracts based on a variable percentage of the equipment's retail price. A retainer,

which includes discounts off hourly support rates and training costs, covers all other support. The company estimates how many hours of support that a client will need and has the client pay for a year's worth in advance. If the full allotment isn't used, the remaining hours carry over to the next year.

Strategic Network Designs (Clark, NJ), a systems integrator that employs three full-time certified Novell engineers and specializes in LAN-to-WAN connectivity, changed from a per-server annual fee to a fee that includes a set block of support time. "Some people were paying more than they needed for support," Rovner says. "We also had people who abused it. They wouldn't open a manual. So it's an incentive [for a client] to open a book and document internally. We recommend a block of time—the larger block the greater the discounting. We've found that that's fairest to all. If you're self-sufficient you don't need a large block."

This pricing structure helps Strategic Network Designs plan its staff resources and scheduling to meet its time commitments. Software support is offered on a hourly basis. Hardware repair has fixed rates, structured similarly to those used in car repair shops.

Intellogic Trace offers pricing based on "support requests." Each service request ends when a problem is resolved, which often takes many telephone calls, says Intellogic Trace's LAN program manager Cliff Mountain. A service contract will allow customer a certain number of requests, typically 24 to 28 per site per year. When you purchase combined hardware and software support, there are no additional charges for parts. When you purchase only software support, the company charges time and materials for hardware repairs.

Also look at how parts and utilities for repairs are priced. If you need a third-party utility to diagnose or fix a problem, the contract should spell out whether it will be priced at a retail or discount level.

You also should specify how a service provider will document its services, for example, through monthly activity reports.

Keep an eye on the bottom line, but don't lose sight of what you need to keep your network up. As LAN Systems' Spiegel cautions, "It's worth paying a premium to someone who knows your system. Get competitive bids, but two percentage points isn't going to matter if the person can't do his job."

## Response Time

Response time is one of the most critical components of a service contract. The biggest problem is defining what it means.

"Vendors offer a range with different meanings," says NMI's Klein. "Does it mean time to answer telephone? Time to get back with a patch or replacement part? Time to go out to a facility? In all cases it is probably not reasonable that someone can guarantee the time it will take to repair [what's broken]." A client needs to distinguish "the difference between responding to and applying all reasonable efforts and actually solving the problem," he says.

Choosing a response time is "a risk-benefit decision for companies based on how long they can afford to live with a problem," Klein says. "If they want for a four-hour on-site response time outside of business hours, they will pay a premium for that. They have to decide whether the nature

## Buying Network Support: A Shopping List

1. Do you want to buy fixed-rate contracts or pay as you go?
2. Is the service provider flexible with issues such as pricing?
3. Are networks the service provider's primary business?
4. How do you want to divide support tasks between in-house staff and outside service providers?
5. What hardware and software components do you want to cover?
6. Does the service contract cover the file server and the network operating system?
7. Do you want help with software repairs and upgrades or with learning and using the applications?
8. Does your contract cover preventative maintenance?
9. How are services priced?
10. What response time do you need, and how does the service provider define response time?
11. Do you need service during business hours or around-the-clock?
12. Does the service provider use remote access software or monitoring tools?
13. What spare parts does the service provider keep in stock, and how fast can it get parts?
14. Will the service provider make a replacement unit available to you during repairs?
15. Does the contract include a document of understanding, a scope of work document, an escalation policy, and a "no excuses" clause?
16. How does the service provider document your network and any changes to it?
17. What do current clients and vendors say about the provider?
18. How long has the service provider been in business, and what is its financial history?
19. How many technicians work for the service provider, and are they certified by the maker of your network operating system?
20. What is the ratio of technicians to sales personnel?
21. What is the ratio of service contracts to technical network engineers; is it 10-to-1 or better?
22. What is the staff turnover rate?
23. What are the service provider's biases?
24. Do you feel comfortable with the service provider's staff?
25. Can the service provider support sites in several cities?

of their business warrants that premium by looking at what possible failures of systems can mean to the operation of their business."

Common response times are half-day, eight-hour, four-hour, and two-hour. Typically companies contract for a two-hour response time for mission-critical applications and equipment, such as file servers.

A service contract should specify what the response time will be for returning telephone calls and for having a technician on-site. Telephone response times, typically a half hour or an hour, can be key. "About 75% of problems can be solved over the telephone," says Intellogic Trace's Mountain.

You also will want to specify whether response times are during business hours or around-the-clock. Find out

who will respond. A technician, not an administrative assistant, needs to call back. Do technicians carry beepers 24 hours a day?

Does the service provider use remote access software, such as Triton Technology's (Iselin, NJ) Co/Session, which allows the service provider to dial into your network and remotely provide support or trouble-shoot. This saves the time and expense of an on-site service call.

During on-site service calls, Intelogic Trace's (San Antonio) certified Novell technicians carry a portable PC that allows them to connect to Intelogic Trace's network and perform functions such as accessing databases and reloading NetWare.

Netlan (New York) offers LAN monitoring packages that include use of tools such as Frye Computer Systems' (Boston) Frye Utilities for Networks, which lets you establish performance thresholds and set alarms. The firm's support contracts include other toolbox programs for network management, a set number of service incidents per year, and unlimited telephone support.

### Spare Parts

Another critical component is the service supplier's spare parts inventory and supply stream. You need to know what it keeps in stock and how fast it can get parts.

What is the supplier's relationship with distributors and manufacturers? "Understand that relationship and understand where they are in the chain—the closer to the manufacturer, the better the chance of getting high quality responsive service," says Klein. "With Federal Express and just-in-time inventory management, you're often better served by a vendor who has arranged to obtain parts rather than the vendor who stocks them."

"Spare parts is the difference between service and no service," says BancTec's Christian. "Anyone can get a technician there. When you pay a charge for more rapid response time, the assumption is that when the technician gets there to fix the machine that he'll have the part. Are the parts on-site or in that city? If the parts are not, they're not getting two-hour response time."

You might consider keeping spare parts, which could be owned by your company or the service provider, at your installation. "Depending on your risk profile, this could be money well spent," says Spiegel.

Look at whether the service provider can provide a replacement unit during repairs and address the network nuances of swapping hardware. It can be tricky to use a substitute file server. Christian calls it "unrealistic. You're asking to break something while you're putting it together. Swapping a file server isn't the answer. You have to have the spare parts on hand."

Think about whether the price of keeping costly replacement units on-site is worth it for your organization. Plan for and keep off-site backups. If the nature of your business warrants it, you could use a service provider's network while yours is being repaired.

### Contracts

A contract should include a document of understanding that spells out the service provider's responsibilities and the client's responsibilities.

This should address issues such as who is allowed to contact the service organization, the hours of support, what the charges cover, and what applications are covered.

A scope of work document should describe agreed-upon work, and an escalation policy should describe procedures for solving disputes.

A contract always should define the final point of responsibility. "You don't want to be in a finger-pointing situation," says Cuban. "You want to make sure who has the ultimate responsibility is defined in a 'no excuses' clause." A systems integrator, by definition, works with products from different companies. The contract should spell out what the outside organizations' and the system integrator's roles will be in solving problems.

Ask vendors a lot of what-if questions about their services. Spend time thinking of different scenarios where you would need to use the support services. Ask "If this happened, is it covered?" By fashioning questions, you'll improve your own understanding of your needs. Then get the service provider's responses in writing as part of the contract.

### Document the Network

Documentation of a network's configuration is essential for maintenance and future growth. "The only things for certain with LANs is that they grow and change," says Cuban. "How does your support person document those changes?"

MicroSolutions, for example, wrote a program to record network configurations. It charges \$75 per workstation and \$500 per server for documentation. "The key is having a change log so that you can keep track of what's happening," Cuban says. Ask to see sample documentation of other clients' installations.

Also make sure that outside contractors provide you with documentation of what they do, including what's been done to network, how it's configured and how defaults are set.

Strategic Network Designs keeps documentation of its clients' installations both at its office and on-site.

If you maintain good internal records, you might not need to pay for a service provider to duplicate your efforts. "It's always good to have documentation, but someone has to be willing to pay for that," says Dennis Passovoy, Evernet's vice president of technical services.

### Reference Checks

You always will want to check references from current clients, especially ones that have similar attributes, such as size of company and type of LAN. If a vendor cites a large company as a client, the level of service that it commands might not be the same as what a smaller company receives.

Check with other vendors, such as Novell, directly. Make sure the service provider has technicians who are authorized to service your equipment and have experience with your type of LAN installation. Ask to see written proof that the service provider is authorized to support both your hardware and your software products.

Study the track record and financial history of the provider.

The number of technicians—and their LAN expertise—is key. Take a look at their resumes. Make note of how many years they've been out in the field fixing LANs. Are they certified by Banyan or Novell? "If most of the engineers have less than two years' experience, you're going to be an experiment for them," says MicroSolutions's Cuban.

What is the ratio of technicians to sales personnel? "We keep about two engineers for every sales person," says Winson. We'd almost like it to be higher than that."

What is the ratio of service contracts to technical network engineers? Cuban recommends a 10-to-1 cutoff point, where you would want to have a minimum of one technician for every 10 contracts.

What is the staff turnover rate? If the turnover rate is high, the staff will learn at your expense.

Make sure the service provider is big enough for your needs. "You want depth," NMI's Klein says. "There are small companies that provide fine service, but a company is accepting some risk that the staff members who support them may not be available. In larger companies, there's greater depth and greater resources available to solve problems."

The service provider should have experience with more products than you currently use, so that it can support you as your network expands.

Consider the company's biases. Are you working with a truly independent service provider who's primary concern is finding the right mix of products and services to solve your problems? A service provider owned by a manufacturer might be more interested in selling its equipment than supporting a multivendor environment. A national chain, too, might be more interested in selling you its line of products than the best products for your needs.

When shopping for a service provider, remember that keeping a LAN up requires teamwork. Get to know the people in the organizations, and pick a company that has a staff whom you feel comfortable working with.

---

## Site Support

Supporting sites in more than one location has its own requirements.

"Most of the [service] companies are small and don't have multiple offices," Cuban says. "They'll subcontract support. You must be careful to have the reseller fully define what a subcontractor will do, or contract [directly] with the subcontractor."

When buying support for sites in different cities, NMI's Klein says a company should consider whether it handles purchasing and other activities centrally or independently at the branch offices. "If the company norm is to do these things centrally, clearly there is an opportunity to get some

saving in a volume purchase, even with service," Klein says. "The question the buyer should ask is: 'If I'm in city A, how are you going to take care of me if this happens?'"

The ideal scenario is contracting with one company that has service capability in all of your branch cities. "Shoot for company with multisite locations, but if it doesn't have a site in some cities, make it that company's responsibility to cover the other cities through subcontracts or other mechanisms," says Evernet's Passovoy. "Execution gets very complicated if you have different providers in each city. In an out-of-the-way area, you may have to be prepared to pay an extra fee to support that site and keep the benefits of having a single contractor."

"Let the local offices buy it from regional maintenance providers, or go with a national organization coordinated from corporate headquarters," BancTec's Christian says. "They can be equally as successful."

---

## Economic Woes

Economic woes are a boon and a burden to network support's growth.

There has been "dramatic growth in selling network support contracts to large companies, which haven't been as able to hire people or transfer people into network support," says Cuban. "Head count is a big issue. They'd much rather buy a contract than hire."

"More companies are trying to do the same thing with less resources, Intelogic Trace's Mountain says. "MIS directors' budgets are flat or growing 4% to 5%, while LANs are growing 40% to 50% per year." Its expensive to train and retain an in-house staff.

Plan for staff turnover by negotiating arrangements for an outside network administrator and training for a new administrator in advance so that you're not trying to make these arrangements under duress, suggests MicroSolutions's Cuban.

"MIS departments are being forced to reduce budgets," Spiegel says. "Maintenance is one of the first places. The tendency in marketplace is to pay as you go, which is OK, if you have a relationship with a vendor. It makes a lot of sense to cover as much as possible under contract."

But when you're buying support, try not to be pound-wise and penny foolish. After fitting your LAN to your business, don't scrimp on its upkeep. Or you might pay the piper as well as the tailor. ■





# Installing a LAN

## In this report:

Designing the Physical Plant .....	2
Software Selection.....	3
Hardware Selection.....	4
Protocol Selection.....	5
Cable Installation.....	6
Interface Hardware .....	6
Installing the Server.....	7
Facility Support .....	8

## Datapro Summary

Installing a local area network (LAN) is not an easy task. With proper planning and some understanding of the technology, managers can minimize difficulty and disruption. This report is an easy-to-read guide to the steps involved in the physical installation of a LAN. It includes discussions of requirements analysis, plant design, hardware and software selection, and file server installation.

## Deciding What You Need to Do

Installation of the physical components of a LAN is the single most time consuming, expensive, and difficult part of making a local area network part of your business. Once you've decided that you need a LAN, you need to decide exactly what services you want the network to handle. Some of these will be obvious to you, since you used them as a way of justifying the LAN in the first place. Others must exist to make the LAN work.

After you've decided what you want the LAN to do, you have to decide where you'll put the LAN, what LAN software will work best for you, what kind of equipment you need to make the LAN work, and where all of this material is going to go. This is a complex process, and it's best accomplished in concert with others who'll be using the LAN, as well as those required to support it.

The first step is to decide what you want the LAN to do. This is called *requirements analysis*, and it means that you need to make a list of requirements. This list will

serve as a guide when you select the hardware and software that will form your LAN, or it will give you something to work with when you hire the company that will install the LAN for you.

## Requirements Analysis

What do you want the LAN to do? What kind of work do you do? How big will the LAN be? Will the LAN connect to any other networks, computers, or outside services? Where will the LAN equipment go? Where do the prospective users work? The answers to these questions are the requirements your LAN must meet. You develop requirements by asking questions about the LAN and listing the answers.

### What Must the LAN Do?

If your primary objective is electronic mail, then the LAN must support it. Likewise, you must list file sharing, print sharing, or group scheduling as functions you want the LAN to accomplish. You should also list specific software you want the LAN to support, since not all application software works equally well with all LANs.

### What Kind of Work Do You Do?

If you're planning the LAN for an engineering shop that uses computer-aided design heavily, your requirements are different from an office that uses a LAN to support word processing on shared files. Either of these has different needs from an office that

This Datapro report is a reprint of "Physical Installation," Chapter 5, pp. 71-93, from *Executive Guide to Local Area Networks* by Wayne Rash, Jr. and Peter R. Stephenson. Copyright © 1990 by COMPUTE! Publications, Inc. Reprinted with permission.

performs a great deal of data entry and retrieval from a central database. You must specify the type or types of work that must be accomplished, and you must have some idea how much of this work is actually being done at any one time.

### How Many Users Are There?

A LAN that must support two thousand users is different than one that must support twenty users. Where possible, you should be able to specify how many users are doing what kind of work. Networks have upper limits on the number of users they can support. Sometimes the limits are specific (Novell Advanced NetWare 286 can support 100 users, for example) while other limits are practical. (3Com 3+ Open can support any number of users, but on a practical basis it will support about the same number Novell can.) Depending on the type of work being done, networks also have limits on the number of users they can support and still give adequate performance. If your company does a lot of database work, for example, it will slow things down a lot, but the amount of slowdown will depend on the type of LAN, the number of people doing database work, and the kind of database work they do.

### Will the LAN Connect to Other Services?

If you have a company mainframe that must be accessible through the LAN, you must include that in your requirements. Likewise, you must include any requirements the LAN must support: dial-in remote access, video as well as data, or working with a remote database service. While a LAN can handle any of these items, they work better when they're designed in from the beginning.

### Where Will the LAN Equipment Go?

This might seem like a prosaic topic in this area, but the selection of the location for the LAN server and other support equipment can be very important. It can even dictate the type of LAN you end up getting. To some extent, the location of the equipment will be determined by the availability of space, power, and other utilities. It may also be determined by the location of key employees. Some of these items are critical. You must have sufficient power and ventilation for the LAN to function, for example.

### Where Are the Users Located?

In some cases, a company will simply wire every potential work area for the LAN. This is nearly always done in new buildings, since there's not always any way to tell who will need access to the LAN until the people move in. In buildings that have the cabling installed after the staff has moved in, this isn't always done, partly because of the expense of wiring and partly to reduce disruption as much as possible.

Advances in LAN wiring technology have made it easier to wire every office in a building for LAN access, however. Where once a LAN usually required special types of coaxial cable for its wiring, this is no longer the case. A number of vendors have developed ways to use commonly installed telephone wiring for the three most common types of LAN: ARCnet, Ethernet, and Token Ring. Of course, you must then have extra telephone wiring available to support the LAN, but most modern buildings were built with extra wiring in case it was needed, and the LAN can take advantage of this.

Figure 1 is a list of items that might appear on a typical requirements list for a LAN installation.

Figure 1.  
LAN Requirements Example

Requirements List for ACB Co. Inc. LAN:

- Must support electronic mail, shared word processing, CAD for the design group, and access to the IBM 4341
- Must support 34 users, of which 8 are CAD users
- Must support *WordPerfect* word processing, AutoCAD computer aided design and drafting
- The LAN support equipment must be located in room 4-207, and auxiliary equipment may be located in the telephone equipment closets if needed
- The LAN must make use of existing wiring where possible

## Designing the Physical Plant

In LAN terminology, the collection that includes the file server, the cable, any support equipment, gateways, or communications servers is called the *physical plant*. Normally, this isn't considered to include the work stations that are on the desks of individual users, nor the network interface card that's installed inside this computer. It might include such things as system printers, telephone wiring, and similar items, depending on who's talking to you. Here we won't include these, but if you're talking to LAN engineers and hear the term, ask to make sure their definition is the same as yours.

Once you've decided what requirements your LAN must meet, it's a good idea to do a simple schematic drawing of it. This drawing is to help you make sure you understand what's really going to be on the LAN. It isn't intended to reflect the LAN's actual physical layout. Instead, it's intended to make sure you're including all of the necessary items in the LAN specification.

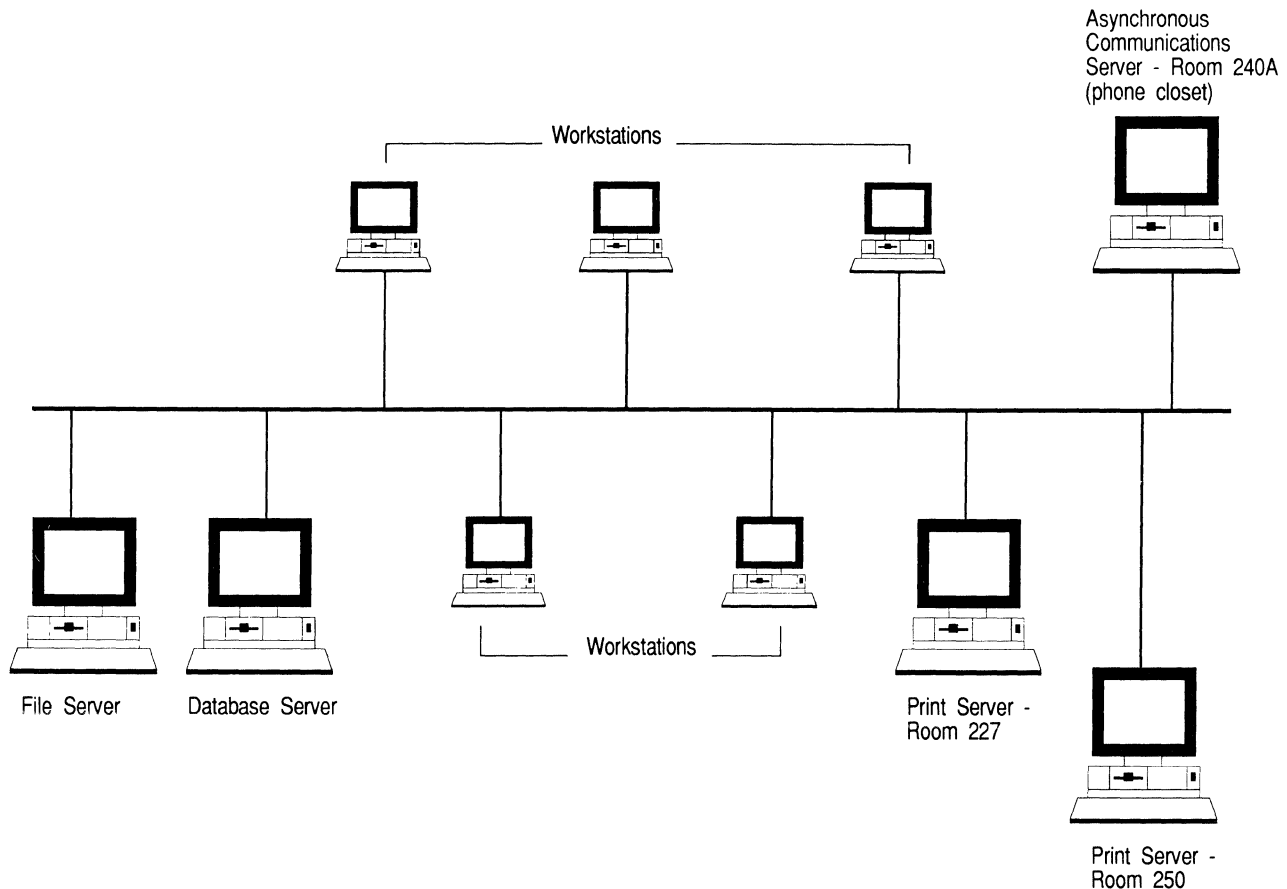
Most people make a schematic drawing that's as simple as possible. It usually represents objects attached to the LAN as boxes labeled with names such as "File Server" or "Workstation" connected by lines that represent the LAN topology. Connections to other data services, mainframe computers, gateways, and so forth are shown graphically (see Figure 2).

It isn't necessary to show everything on the LAN when you do the schematic. This means you probably should show a few examples of workstations and then note on the drawing how many there are. The same goes for such items as printers. On the other hand, you probably should show major LAN equipment such as file servers and gateways.

Once the schematic drawing is complete and checked for accuracy, it's time to start on the drawings of the actual installation. Normally, this is done with a set of floor plans for the area where the LAN is being installed. You should mark the rooms where the file server and other support equipment are located. You should indicate which offices or work areas are to receive service. If possible, you should show approximately where the LAN cable will emerge from the wall or ceiling for attachment to the workstation (see Figure 3).

Since there are certain maximum distances a printer and other equipment can be from the computer that's driving them, it's important that printer locations be drawn as close to their actual locations as possible. You should confirm that the printer is within the allowable distance. Once

Figure 2.  
Typical LAN Schematic



you've completed the drawing, it's time to start the installation process of the equipment.

At this point, you've been carrying out the same processes whether you'll be installing the LAN yourself or having someone else install it. Now's the time when you have to decide which it's to be, if you haven't made this decision already. If you're going to hire a contractor or use the services of your company's MIS department, you'll need to use the requirements and drawings you've prepared to create a set of specifications that will be used to order equipment and software and to guide the installation (see Figure 4).

If you're doing the installation yourself, it's still a good idea to draw the specifications. You might not need them yourself, but they will help you ensure that nothing is overlooked as the installation progresses. For the most part, the specifications are simply a detailed list of requirements that includes the floor plans and a thorough description of what you want the LAN to accomplish.

If you're working with LAN installers, either an outside vendor or another department from your own company, this is the time when you'll meet with them and go over the details of the installation. This will point up any questions or items that might have been missed, and it will give you a chance to express any concerns you might have about your installation. If you're doing this yourself, it's a good idea to meet with others in the department who'll be using the LAN for the same reason.

### How Do You Decide Who Does the Installation?

LAN installation is a complex process that requires the services of a number of specialists. In most cases, you'll require the services of a licensed electrician to satisfy the local authorities, and you must order cable that meets local fire codes. If you haven't done this before, it's probably a bad idea to learn by doing it yourself.

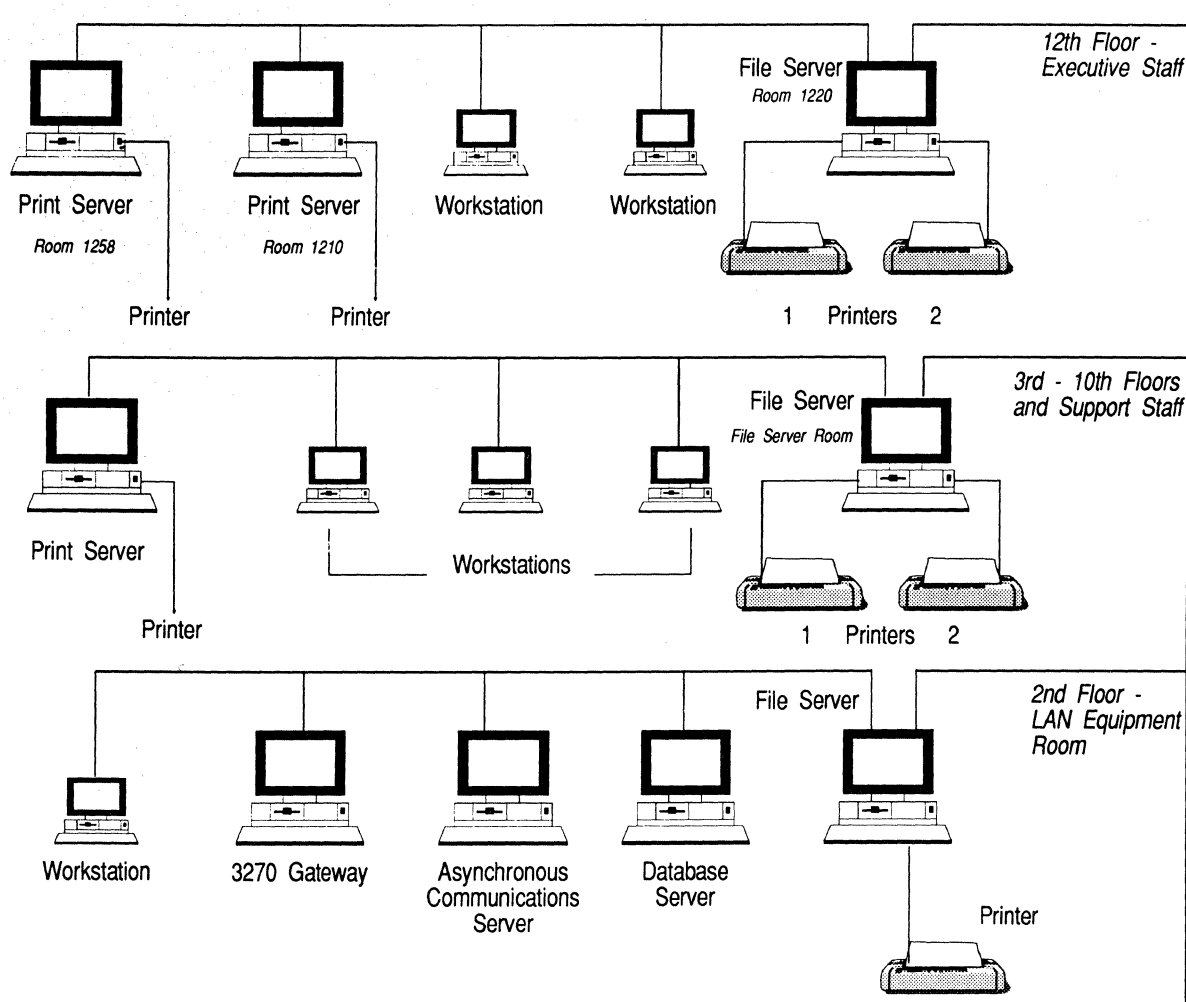
If you've done successful LAN installations before, then you probably aren't reading this report. But if you *have* installed LANs, you already know that this is a process best left to specialists. It's one thing to specify the LAN, and quite another to begin pulling cable yourself.

This isn't to say that you shouldn't take an active role in the installation. You should. That's the only way you can make sure the LAN is installed in the way you want it. You should make sure you're kept up to date constantly in the details of the installation of the equipment. Your installer should be able to answer questions about why a certain piece of equipment was chosen, or why it's being installed in the manner that it is, any time you ask. After all, it's your LAN, you're paying for it, and you have the right to find out what's going on.

### Software Selection

One of the first decisions you'll need to make with your LAN installer is to select the LAN operating system and the other software that will be used. When you do this, you

Figure 3.  
Typical Schematic for a Multiple Floor Installation



BACKBONE

must be sure that the LAN operating software you choose will meet your requirements for compatibility, security, and administration support, and that it will support the services you need. You must also be sure it will support the number of users you need on your LAN.

Normally, the installer will recommend one of the top three sellers in LAN operating software. The most commonly used system is Novell NetWare, followed by 3Com and Banyan Systems. Each of these three LAN operating packages will support most common LANs. Each has its strengths and weaknesses and, for the most part, will meet your needs, unless they're unusual.

The top selling network operating system comes from Novell. This is because it has tended to be faster than the others and requires fewer resources from the workstation. For that reason, there's some software that will work only with Novell NetWare. These differences aren't constant, however, and you might find that the other LAN operating systems will meet your needs as well or better.

### Support Software Selection

The network operating system doesn't provide all the answers. While some operating systems provide a fair number of utilities and electronic mail, you'll still need some

network management software. You'll need packages such as *office productivity software* (sometimes called *groupware*) and other support packages. Most of these packages are groups of applications sold together that are supposed to make your workgroup more productive.

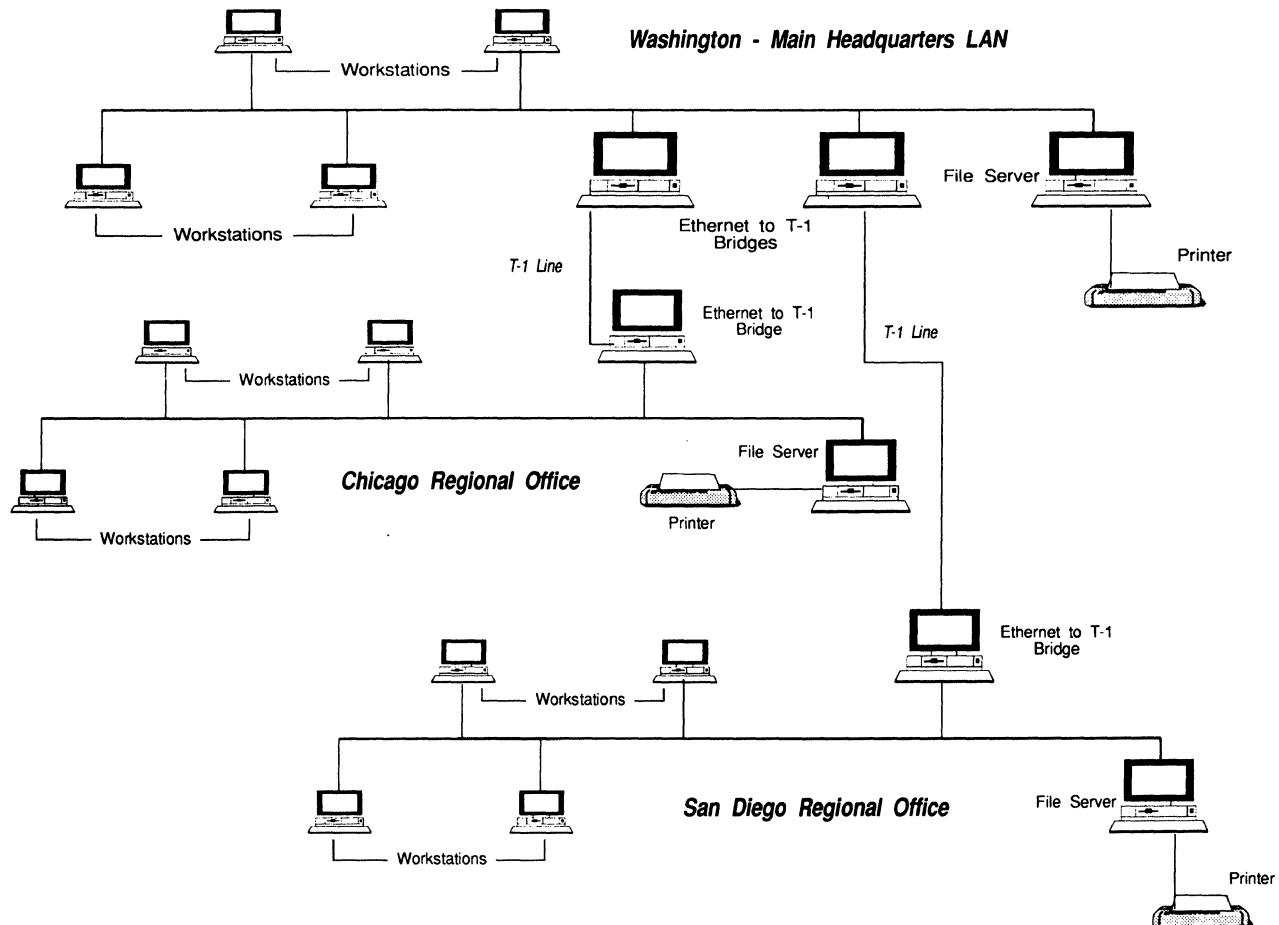
These office productivity packages, or groupware, usually consist of an electronic mail package, a calendar, a group scheduler, and usually some other software combinations. One good example of groupware is *WordPerfect Office*, which combines all of the above plus a game and an ASCII text editor. *WordPerfect Office* also has the ability to add your favorite applications into its menu system. A similar package from Enable is called *Higgins*.

Other support software might include a network management package, or perhaps a communication package such as the network version of *Procomm Plus*. None of these packages is what you bought the network to support, but they're very useful nonetheless. A few, such as *WordPerfect Office*, might eventually become the most popular software on your LAN.

### Hardware Selection

One of the major differences in software selection is in the way it affects your choice of hardware. Both 3Com and

Figure 4.  
Typical Schematic for Offices in Multiple Locations



Banyan Systems sell proprietary file servers, and vendors of their products will sometimes try to sell you those machines. Novell requires only an IBM PC/AT compatible computer with a hard disk and 2 megabytes of memory. For this reason, the Novell installation will normally be less expensive.

As it happens, both Banyan Systems and 3Com are available for standard IBM PC/AT compatible computers as well. You don't really need to buy those expensive proprietary file servers in order to use LAN operating software from either company, and you'll save a great deal of money by going with something more generic.

This isn't to say that you should buy the cheapest PC/AT clone you can find. Normally, you should make it a point to buy a machine that has a good reputation for reliability and that comes from a manufacturer that's likely to be in business for a while.

Remember, your business information is going to reside on this file server once it's part of the network. You don't want to be in a position of having to mail your file server across the country to get it repaired. Instead, you want a machine that can be serviced at your location, preferably within a few hours of when you called for service.

Most major computer manufacturers have had their products certified as being compatible with the major network operating systems. This is also true of the makers of hard disks that are designed for LAN use and for some

other LAN peripheral equipment. While there's no indication that noncertified hardware is any less likely to work, buying hardware that's certified by your software manufacturer does reduce finger pointing when there's a problem.

Once you've begun choosing your LAN hardware, remember that most commonly available computer equipment will work well in a LAN environment. If your vendor recommends something that looks like it's not really the latest in high tech equipment, he or she could only be trying to save you money. Some applications, including print servers and gateways, don't require a particularly fast computer to support them. Print servers and gateways are circuit boards that mount inside the computer, and they depend on the computer as a source of power and for a way to run some software. They don't need anything more complicated than a floppy disk-based PC/XT to work just fine.

## Protocol Selection

By now, you've probably already decided whether you want ARCnet, Ethernet, or Token Ring. If you haven't, now is when you should make that selection, because it will affect the hardware you buy, the wiring you install, and possibly the location of some of the major pieces of equipment.

More importantly, the selection of the LAN protocol will dictate some of the major pieces of LAN equipment

you'll select, since there are items that are unique to each LAN protocol. There's also the consideration that unless you're planning to use the telephone wiring in the building, you'll have to install a different type of cabling for each of the types of LAN protocol.

For most users, the choice of the LAN protocol makes no difference. Each is fast enough and has enough capacity to support the electronic mail, word processing, and spreadsheet applications the vast majority of users require. You might as well go with something that will work and not cost too much.

ARCnet is the least expensive protocol to implement, mostly because the network interface cards are inexpensive. On the other hand, it's slower than the other two. ARCnet is necessary if you plan to work with Datapoint minicomputers you already have. It's also an advantage if you have IBM 3270 terminal wiring in your area that you can use, since it can make use of the same cables.

ARCnet is a distributed star topology, and at the center of each star must be a piece of equipment called an *active hub*. This hub acts as an amplifier, and it must have a source of power. Normally, one hub will connect with eight cables, including the cable coming from the previous hubs and the cables leading to any other hubs farther down the line. Some companies make ARCnet hubs that support more than eight connections.

Ethernet tends to be slightly more expensive than ARCnet, but you get more performance. Ethernet moves data at ten million bits per second versus ARCnet's two and a half million bits per second. Ethernet also has less overhead than either ARCnet or Token Ring, so it's well suited for installations that require the ability to move large quantities of data quickly, such as in an engineering operation that's using CAD.

If you have a VAX or a UNIX system-based minicomputer in your company, you'll find that you have to use Ethernet to communicate with these machines. Ethernet operates by checking to see if there's traffic on the network and then transmitting data if there's not. This means there's a chance there will be two stations trying to transmit at the same time. As there are more stations on the network that transmit frequently, these collisions become a problem. In networks that have a large number of transmissions, such as in a database environment, Ethernet can sometimes lose some efficiency.

Token-Ring was developed by IBM to make up for the problems Ethernet had in an environment with many transmissions. Their Token-Ring protocol doesn't really transmit information any faster than Ethernet, despite what IBM will tell you, but it's better suited for an environment where there are many short transmissions.

Token-Ring is much more expensive to install than either of the other two protocols. Part of this is due to the higher cost of the network interface cards, part to the IBM wiring requirements, and part to the requirement to buy MAUs (Media Access Units) at the center of each grouping of Token-Ring computers. If you have IBM mainframe computers or the new series of AS/400 minicomputers, you might want to choose Token-Ring because these machines are also designed to work with your current computer.

## Cable Installation

Except under the most unusual circumstances, your role in the cable installation process will be to tell the installers where the cable is to go and where the faceplates are to be

in each office. It's a rare manager who is compelled to actually start pulling network cable. Partly this is because managers have other work to do, and partly because few managers know how to do the job in the first place.

Specifying the locations of the cable drops, the type of cable, and the manner of installation is important enough, though. Even more important is the decision as to whether the cabling should be done as a part of the telephone system or whether you should have separate network cabling. These choices are determined by the type of network, the distance the cable has to run, and at what point in the course of building construction the cable installation is taking place.

In a new building you have a fair amount of freedom. You can install LAN outlets nearly anywhere you want and the cost will remain relatively low, because cable installation before the walls go in is fairly easy. Once the walls are in, installation is much more difficult, and the cost goes up.

The labor cost is only one factor. The cable itself isn't cheap, and installing a lot of it can change the cost of construction significantly. IBM Type 1 cable, which is commonly used in Token-Ring installations, can cost over a dollar per foot. Ethernet and ARCnet cable is cheaper, but it's still expensive.

An answer to this is the growing popularity of using telephone cable, since many telephone installations contain a great deal of extra cable. The phone company, knowing that there are advantages to having spare cables already installed in a building, normally installs cable with several wire pairs when only one is needed. The added cost isn't great, and the benefits are significant.

Just because the phone company installed these cables for its own convenience doesn't mean you can't use them. They belong to you, since they're part of the building. Normally, you have to hire an installer who is familiar with phone systems to have the phone wiring converted to LAN wiring. This involves attaching spare pairs from each office to a wiring block in the telephone closets, and installing an additional phone jack in the faceplate in each office.

Your LAN installer will need to know which type of LAN you're planning to use, since there are minor differences in the way the three major types of networks are wired. Of course, by this point, your LAN installer is probably well aware of the LAN installation you've planned. The installer will add additional support equipment such as concentrators for Ethernet or multiple access units for Token-Ring. These will be connected to the wiring block, and the workstation wiring will mostly be complete.

There's also a type of wiring called *backbone wiring* that connects several file servers together. This may be heavy coaxial or perhaps fiber optic cable, and it serves to connect one LAN with another. In a multistory building, the backbone cable will usually run between floors.

Once you've specified the wiring, you must also specify details of the installation. This will include making sure that the cable meets all applicable fire and building codes, and that it's properly tested. Normally the cable installer will test the cable as a matter of course, but you need to make sure that it's done and that the cable all checks out as good.

## Interface Hardware

The final link between the LAN and the workstation is the network interface card. This is a circuit card that normally

installs inside your computer and attaches to the LAN cable. The general type of card is specified by the type of LAN you're installing. For instance, you must have a Token-Ring network interface card if you have a Token-Ring LAN.

There are a number of different brands of network interface cards for each protocol available. Your primary objective in selecting a card is to make sure that it will work with the LAN operating system, that it won't slow down operations, and that it is reasonably priced. If the card will work properly with your PC and transfer data fast enough, there's little to be gained by paying more for a network card. An inexpensive card will work as well for most installations as an expensive one.

The exception to this is with fast machines that are moving very large files, such as you'd likely find in a CAD office. Because computer aided design files are very large graphics files, they can take a considerable amount of time to transfer from the server to the workstation. In LAN terms, this means a second or two. During this time a great deal can happen, including collisions from other workstations, which can slow the transfer even more. For this reason, you'll want to speed up the transfer as much as you can. Since the high speed workstation can handle a great deal of data at one time, a card with a lot of throughput, such as a 3Com 3C505, can make a difference. If you're using a Micro Channel or an EISA-based workstation, a bus mastering card designed for the specific type of machine will speed up transfers even more.

Once you've installed the network interface card, the LAN software will need to be configured to accept it. With Novell NetWare, this involves a special installation process that allows you to pick nearly any card available and install it in one of several different ways. Some other LAN operating systems have less flexibility, but that's made up for to some extent by easier installation. With most network operating systems, the workstation software is included with the network software. A notable exception to this is DECnet from Digital Equipment Corporation, which charges about \$700 per workstation for its software.

The main differentiating factor between workstation software packages is the amount of memory they consume. Novell NetWare is probably the smallest of any of the major systems, taking about 45 kilobytes of memory on the average. Other packages can take significantly more, and a few such as LANtastic can take less. All of the packages allow you to address the remote disk as if it was on your computer.

## Installing the Server

By the time you're ready to install the computer you'll be using for a file server, you should have already decided where you want it to be located. There are a number of other factors that depend on the location of the server. These factors include the site for termination of the LAN cabling and the location of the system printers. Because the cabling depends on knowing where the file server will be located, you have to choose that before you can complete the cable installation.

During the final stages of the cable installation, you should be installing the file server, loading the network operating system, testing the server, and installing any peripherals that will be attached to the file server. These peripherals would include any external disk drives, printers, power supplies, monitoring cables, and communications lines. Once these are installed into the space the server will

occupy, the server testing should begin. By this time, there should be sufficient cable installed to begin testing the server in conjunction with a few workstations.

### Server Location

The location for the server can be very important. Sometimes, especially in very small LANs, the server will be located in the office with you so that it's convenient. With larger LANs, especially with LANs that have more than one file server, you'll want to locate the server in a location away from daily activity.

Ideally, the file server should be located in a room that already has sufficient air conditioning or ventilation, enough power available for the file server and all peripherals, enough space to allow the LAN administrator comfortable access, and enough security to prevent tampering. This isn't always available, so you might need to make some compromises. Generally, the only items you can't compromise are ventilation and power. A file server won't work if it overheats, and it won't work without reliable power.

An ideal location would be either the computer room, if your company has one with enough space available, or a vacant office. Either of these will probably offer security, cooling, and power. The computer room has the advantage of being convenient to the mainframe or mini computers, which makes gateway access easier. It also has a staff who can probably be trained to operate the LAN during off hours and to perform backups. A vacant office isn't quite as nice, since it usually lacks a staff, but in smaller LANs, this won't matter.

The office or computer room must be located in a place where the LAN cabling can be made available easily. This means LANs that are configured as a star (including Token Ring, StarLAN, and ARCnet) and that have a bus topology (including most broadband LANs as well as Ethernet) must have the file server located somewhere along the path of the bus, but not necessarily in the center.

Once the site is chosen, make sure it's outfitted with a sturdy table, one that's capable of holding the servers, the uninterruptible power supply (UPS), and related peripherals. In many installations, you'll want to locate all servers in the same place, both for ease of access and to make sure they have the right environment. The servers will need to have access to their printers located outside the server room.

The printers must have a physical connection to the server if they're to be system printers. This usually means each printer must have some sort of extension, since it's usually difficult to locate a printer within 20 feet unless the file server is located in the same office. You may also choose to use a *print server*, which will allow printing to take place anywhere on the LAN. Normally, all LANs will have at least one dot matrix printer that's for the LAN supervisor's use. This is used to perform printing during tests and installation, and it's usually located in the server room with the file servers.

Some organizations choose to use a rack for the mounting of the servers and their keyboards and monitors. This requires a 21-inch rack unless the server you choose is specifically designed for rack mounting. This method requires less floor space, but it makes use of the keyboard and monitor less convenient. Once the LAN is operational, this is rarely a problem, since network administration is usually performed from a workstation.

There are as many other ways to install file servers as there are organizations with unique needs. Servers have

been mounted on shelves, wire racks, the floor, balanced atop mainframe computers, as well as the more traditional desks and tables. The real key is that the server be solidly in place so the hard disk won't be jarred, and that it have the ventilation, power, and the like mentioned above.

Because most file servers are simply IBM PC/AT compatible computers that are being used as file servers, they're usually mounted in the same way as a desktop computer. They're set on a table with the keyboard in front and the monitor on top. This works fine.

### Beginning the Installation

You begin the actual installation by opening the computer and installing the network interface card. You may also need to install a disk controller and some extra memory. This depends on the machine you started with and the configuration you want when you finish. There may be more than one network interface card and more than one disk controller. Once you have the items installed that fit inside the file server, you close it up. Next attach the disk drives (if they're mounted externally), the printers, and the network itself. You will, of course, already have connected the keyboard and the monitor.

Once this is done, you'll load the network operating system, if this hasn't been done already, and begin testing. The initial tests include making sure the file server boots up and runs, the LAN cable is attached, and that the UPS works like it should. Once this is completed, you'll need to try to sign onto the network from a workstation. If you can do this, the installation is either correct or nearly so. You can continue setting up the LAN.

### Utilities

You'll need to have some basic utility services available to your LAN to make it operate. The most basic of these is power. You need more than just electricity, though. You'll need power that's reliable, constant, and clean. This means it must not go out frequently, the voltage must be within tolerance, and there must not be extraneous noise or voltage spikes. Normally the only way to guarantee that this will be the case is to get an uninterruptible power supply (UPS). The file server can monitor this supply to see if the voltage stays within limits and that it has voltage available at all times.

Nearly as important is ventilation or air conditioning. Normally you must have air conditioning, but if you're

planning to locate the LAN in an area with moderate temperatures and reasonable access to fresh air, then air conditioning isn't necessary. You'll need to watch the file server more carefully under these circumstances, however.

If you're planning to have telephone access to the outside world, then you'll need to have telephone lines near the communications server. Normally, this is located near the file server, but it doesn't have to be.

### Facility Support

The final area of concern when you're installing your servers is that you have the required support from whoever manages your facility. This means that the people who actually operate the building must know the LAN is there, and they need to agree to take it into consideration in their operations. You must be assured that the janitor won't turn off the power to the server while cleaning.

Most likely the facility manager and staff will already be aware of the LAN installation. The cabling was probably done with their support. You'll just need to clear with them the specific requirements for dealing with your new installation.

Fortunately, dealing with a LAN isn't a difficult task from the facility viewpoint. They simply need to let you know ahead of time if they must shut off the power or the air conditioning. Facility management will need to provide you with locks for the room the server will occupy. They should also assure you that service technicians will have access to building areas where LAN equipment is installed, in case service or additional installation is required.

The physical portion of LAN installation is probably the largest single portion of the job. It's certainly complex and it can be costly, but it's easily analyzed and it yields to careful planning. In reality, planning is the secret to having a successful installation, whether you're doing it yourself, or having a LAN consultant or contractor do it for you. If you're having it done, you need to ensure that you still do your planning, but you also need to know that the contractor plans carefully, too.

Thoughtful planning will lead you through the installation process in a logical manner, but you must make sure it's done carefully. Map out your steps as they're presented here, carefully determine what your approach will be, and write it all down. Make certain you do this in enough detail that you'll still remember what you need to do when you're deeply within the chaos of installation six months later. ■



---

# Your First LAN: Do It Yourself?

---

## Datapro Summary

Although the process of planning for and implementing a LAN may seem intimidating to most business users, it is possible to install a LAN yourself with the proper resources. Cost savings is a logical reason to undertake this process by yourself, but be aware that if the planning and implementing phases are not well-conceived, it may cost your business more through the life of a LAN.

---

One of the mistaken impressions that business computer users have is that LANs are only for large businesses. Many business users think that something as complex as a LAN can be designed and installed only by a company that specializes in such things. While this may be true in the case of large corporate LANs, the fact remains that, in many cases, it's entirely possible for a small company to design and install its own LAN.

A great deal depends on your ability to work with computers and your willingness to spend time on LAN installation rather than (or perhaps in addition to) your normal line of work. Not everyone is a candidate for a do-it-yourself network. In some cases, you can do part of the work, though, and still have a role in the installation.

A good example of a business that succeeded with a do-it-yourself LAN is Metropolitan Helicopters in Manassas, Virginia. The president of Metropolitan is Dave Carter, a knowledgeable computer user. Dave runs one of the largest flying services in the Washington, D.C. area and performs much of the work with a trio of PC clones

from Tandy. These computers do everything from keeping the books to downloading aviation weather from the Federal Aviation Administration. Unfortunately, they're also widely separated from each other.

While he could have used the computers the way they were, life would have been easier if the accounting functions were available from the back office as well as at the front counter. Swapping disks would never do, since there was no way to be sure that the accounts would stay the same while they were being used in two different locations.

Dave started off with a LANtastic starter kit and ran the cable himself. The design was limited to figuring out where the cable runs would be, and in a couple of hours, Dave and Metropolitan were networked.

The resulting benefits were immediate. The accounting work was significantly more productive, and the load on the staff was reduced. The LAN clearly made a difference.

Not every business is as easy to network as Metropolitan Helicopters, and not every business has someone who knows as much about computers as Dave Carter does, so not every business is a candidate for a do-it-yourself LAN. Likewise, not every business owner is as willing to work nights, as was the case here. So, how can you tell if you should do all of the job, part of the job, or

---

This Datapro report is a reprint of "Your First LAN: Do It Yourself?" by Wayne Rash, Jr., pp. 105-110, from *BYTE*, Volume 16, Number 9, September 1991. Copyright © 1991 by McGraw-Hill, Inc. Reprinted with permission.

none of the job? The answer is that you must evaluate the capabilities of your business.

## Making the Decision

The following points will help you decide if designing and installing your own LAN is for you. Remember that these are just guidelines and that there may well be additional factors that will sway you one way or the other. Remember also that while the cost of LAN installation is usually one factor in looking at do-it-yourself LANs, a lot of factors constitute costs over the life cycle of a network. If you don't know what you're doing, the money you save on installation will be extracted from you many times in the future as you try to recover from the results.

- *How much do you know about your computers?* While a simple LAN installation doesn't require formal education in computer science, you should be comfortable opening up your machines and adding or removing expansion boards. You should also be familiar with computer documentation and be willing to try a process several times before you get it right.
- *How much downtime can you tolerate?* When you're installing your computers, you may be without them for as much as a day or so. A lot depends on how well the installation proceeds, and that depends on your experience. Professional installers can have your machines out of operation for only a few minutes apiece. If you can't live without your computers for a while, you might want to avoid doing it yourself.
- *How are you at construction techniques?* Installing a LAN involves running cable to several offices and may require you to install junction boxes in walls, install conduit, and maybe install electrical power. If you aren't familiar with these skills, and if you don't have a license in areas where one is required, you will need to hire someone for this part, at least.
- *How much free time do you have?* You need to have the time to do the installation properly, including the part about reading the manuals. If you're already perpetually minus on minutes, you might want to reevaluate installing a LAN yourself. Now did Dave Carter manage? Things can get pretty quiet at a flying service when the airport is socked in.
- *How big is the project?* Success with do-it-yourself LANs drops as the complexity increases. Most users without an MIS staff should contract out even a 10-user LAN. Doing a 100-user LAN yourself without a trained staff is folly.
- *Are you connecting anything besides personal computers?* Unless you have the right training or experience, adding gateways, bridges, routers, and multiple servers is the province of skilled contractors. Many large companies, even those with skilled staffs, farm out this work because it costs too much to use a support staff for new installations.

Ultimately, the decision to install a LAN yourself is a business decision like any other. If the net cost (including all factors) is lower to do it yourself, then you should consider that option. Just remember to count things like downtime in the cost column of your comparison.

## Doing Part of the Job

Let's say that you've looked at the options and it's now clear that you're not really in the market for a do-it-yourself LAN. If that's the case, you may be a candidate for doing part of the LAN yourself. Frequently, this is both less expensive and more effective than contracting the whole job. It's less expensive because you do part of the work, and more effective because you have more involvement in the process, so the resulting installation closely reflects your needs.

In the case of doing part of the LAN for yourself, you have to decide what skills you can provide and whether you can provide them more cheaply or more effectively than the LAN installer. You might have your company maintenance staff (if you have one) do the basic cable installation, for example. You might produce the initial drawings of the LAN, showing the locations of workstations, cable connection boxes, and LAN support equipment. You might even install some of the simpler hardware, such as the network interface cards. Even doing something simple, such as attaching 3M Post-it notes to the walls to indicate connection box locations, can save money and reduce installation time.

The best way to decide how you can help install your own LAN, and maybe save some money in the process, is to make an inventory of your own capabilities. Of course, it helps a lot to know what capabilities are required:

- *Are you good with graphics software?* If so, you can save a lot by using a program like AutoSketch from Generic Software to create the drawings for the LAN. This company is a subsidiary of Autodesk, which makes AutoCAD, the leading CAD package. AutoSketch is easy to use, and there's an optional Symbols Library for LAN design. This means you can create your own physical LAN layout.
- *Do you have a staff electrician?* If your electrician (or some other qualified person) can do your data cable installation, you'll speed up the LAN installer's job. You may find that you'll have to make arrangements for additional power in any case, so you can do both jobs at once.
- *Can you plan where the equipment should go?* Making a survey of your office spaces and deciding where you would like to have the wall boxes placed for the LAN connections, and deciding where to place equipment such as file servers, power supplies, and the like will ease installation and lower prices.
- *Do you have a skilled staffer who can help with the installation?* Many LAN installers will let you supply some of the labor for things like installing network interface cards. This means you'll pay the installer less and at the same time will learn more about your own LAN than you would if your staff didn't help.

Keep in mind that a great deal of the process of helping the LAN installation along can be negotiated with the company doing your installation. In fact, this is a good way to help select a LAN contractor if you're interested in keeping prices down and in getting the process done as quickly as possible. You should remember, however, that you'll be responsible for any work you do yourself. This means that if your cable was installed incorrectly, you'll have to fix it or pay to have it fixed. Likewise, if your staff installed the network interface cards, and they are not set up properly, your staff will be responsible for making them work. This

is why you have to make sure you have skilled people doing the work in the first place.

## The Lay of the LAN

One area that nearly every LAN customer can help with is the physical layout of the network. As mentioned in the checklist above, this can be done with existing blueprints or with software that lets you create your own.

First, get a set of blueprints of your office spaces. In some cases, finding an existing set is impossible and you'll have to create your own drawings. In new buildings, a set of blueprints is usually available from the building manager or the owner, or maybe from the company that built the structure. If you have a choice, you'll want to ask for the drawings that have the electrical and HVAC components illustrated.

If you don't have and can't get drawings of your offices, it's time to get a copy of AutoSketch and do your own. Make a rough sketch on graph paper, making sure to draw to scale using actual measurements. You might have to pick up a 100-foot tape measure at the local building supply store to get accurate measurements of large spaces. You'll also need to measure ceiling height and note what kind of ceilings you have.

Once you have drawings in hand, you need to indicate the proper placement of the LAN hardware, as far as your requirements are concerned. To do this, you should visit every office and note where the LAN connection box should be located. This is an electrical junction box about the size of similar boxes that already contain phone connectors and electrical convenience outlets. Generally, it's a good idea to put the LAN connection in the general vicinity of the other boxes so that the occupant of the office knows where to find things.

You will also need to decide where the LAN equipment should be placed in your office. Normally, this equipment will include a file server, an uninterruptible power supply, and perhaps one or more hubs to which you will attach the workstation cables.

The LAN equipment should be placed out of a traffic area, but in a location that has good climate control. You will need to be sure that you have a room that can be locked if your office isn't very secure or if you have employees who like to fiddle with computers when they're not supposed to.

If you're helping with the layout for a larger LAN, you will need to be concerned about the location of telephone equipment closets, building backbones, and cable raceways. In this case, the installation will work better if you perform a preliminary study so that you can get a meaningful bid from a LAN contractor and then work with the contractor that you select to perform the more detailed work. In complex LANs, there are a number of acceptable ways to meet your requirements, and you will need to find out how the contractor plans to meet them so that you can work together to that end.

## Products Discussed

<b>AutoSketch</b> .....	\$249	<b>LANtastic Starter Kit</b> ...	\$699
<b>Symbols Library</b> .....	\$49.95	Artisoft, Inc.	
Generic Software, Inc.		3550 North First Ave., Suite	
11911 North Creek Pkwy. S		330	
Bothell, WA 98011		Tucson, AZ 85719	
(800) 228-3601		(602) 293-6363	
fax: (206) 483-6969		fax: (602) 293-8065	

## To Do or Not to Do?

Now that you've seen some of the ways that you can do all or part of your own LAN installation, it's up to you to decide what role to play. In some cases, such as the installation that Dave Carter performed himself, the choice is simple. LANtastic is easy for a user to install, and it works fine. The question is more difficult in more complex installations, however.

When an installation becomes more complicated than a few workstations, such as when it moves into the world of file servers, mainframe gateways, and communications servers, you need to call in people who have the required training. It may be that those people already work for you in the form of your MIS department. Approaching a complex LAN installation with untrained installers is an effective way to extract money from your business, so it's better to do it right the first time.

Finally, there's also the question of whether you want your staff to be spending its time installing LAN hardware and software. Depending on how lean your organization is, you might not want to shut down other operations just so your own staffers can load the NetWare shell onto everyone's workstation. Again, this might be a situation in which an outside contractor can do the job more efficiently, despite the fact that you have the skills you need on your staff.

Installing your own LAN or helping with the installation has some distinct benefits beyond the financial savings. When you participate in the installation, it means that you become vastly more familiar with the network than you would have been otherwise. This, in turn, means that you will have fewer problems that you can't handle and that you'll know what your LAN can do when you have to expand it later.

Ultimately, the question is more than just whether you can do it yourself. The question is also how much of a role you can play, for you should always plan to play some role. After all, it is your LAN, and it'll work better if you know what you've got. ■



# Ethernet Wiring Made Simple

## In this report:

First Through a Punchdown Block .....	2
Then to a Patch Panel.....	2
Variations on the Patch .....	2
Tricky Hubs .....	4

## Datapro Summary

Ethernet wiring can present potential problems due to the quantity of wires involved. The last thing you want is a jumbled mess of kinked wires. This report provides a step-by-step guide for simplified connections, using punchdown blocks and patch panels.

Some wiring closets are so frightening that a few skeletons would be a welcome relief. What follows is a step-by-step guide to help you thwart the attack of the killer wiring spaghetti, otherwise known as Ethernet.

Suppose you have new offices, all open and pristine, waiting for the department to move in and get to work. And you have to wire it up for a local area network, in this case 10BASE-T Ethernet. On the face of it, this is a fairly straightforward job. There will be an outlet at each desk position and all the other places where you'll be putting servers, printers, gateways and other common-service machines. All told, there may be scores of these outlets. And the unshielded twisted-pair cables will all be converging on a hub in a wiring closet.

But before you begin the job, consider the potential difficulties you're likely to encounter. You might have 40, 60, even 80 or more sets of wire (two pairs per outlet) converging in a small, airless room. All of them have to find their way from the point at which they emerge from the wall or ceiling to the hub. If you're not careful, you can end up with something that closely resembles a pile of spaghetti. And when the time

comes to determine whether a given workstation out on the floor is actually working and connected to the hub, you'll have a devil of a time tracing your way through this spaghetti. Even adding a new user could be a nightmare.

So what should you be doing to forestall these problems? The best way to answer this question is to trace through the logical and physical arrangements and identify the ways you can document your wiring plant, so that you can stay on top of the spaghetti.

What we've described below is the way we do things in our East Coast laboratory. It looks a little indirect and complicated, but the physical arrangements have all been made for specific reasons, as you will see. With smaller installations, you may find ways to simplify the setup.

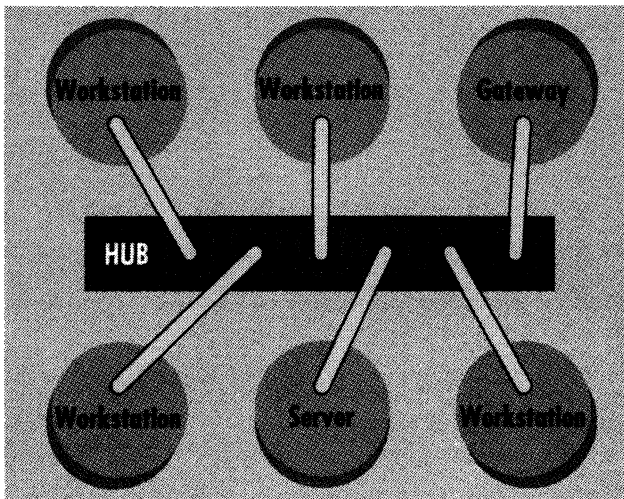
## A Hub on a Star

From a logical point of view, 10BASE-T has the features of a star-shaped topology, even though Ethernet works on an open-ended bus, as Figure 1 shows. That means that instead of one coaxial cable snaking from desk to desk and room to room, the Ethernet connection runs between the desktop and the hub, allowing the hub's control electronics to create a long Ethernet bus that consists of all the separate desktop-to-hub wire segments.

The advantage of this arrangement is that, like Token-Ring, there is one central place where you can monitor and control the behavior of all the scattered workstations, including their network adapters and

This Datapro report is a reprint of "Sorting Out the Spaghetti: Ethernet Wiring Made Simple" by Stephen Morse, pp. 88, 90-92, from *Network Computing*, Volume 2, Issue 10, October 1991. Copyright © 1991 by CMP Publications, Inc. Reprinted with permission.

Figure 1.  
Ethernet on an Open-Ended Bus



associated wiring. If one of these misbehaves and threatens the continued operation of your network, you don't have to go to the culprit's actual desk or room. Instead you can isolate the problem from the network at the central hub, fix it and then re-establish its connection.

### First Through a Punchdown Block

How does the connection get from the desktop to the hub? The answer, as Figure 2 shows, is that it takes a series of hops from the wall plug (an RJ45 jack) to the punchdown block, from the punchdown block to the patch panel and finally from the patch panel to the hub. Behind the wall plug are four wires, one pair of which transmits the data from the hub to the adapter card (the receive pair), and the other pair of which transmit from the card back to the hub (the send pair). These four wires emerge from the wall in the wiring closet and connect to four adjacent locations on the punchdown block.

The punchdown block is your first level of flexibility. It is a 50-conductor (25-pair) matrix of V-shaped connections that holds the individual wires punched down into the Vs. It has an input set of Vs and an output set, allowing for a wide variety of cross-connections. This is a semi-permanent setup that allows you to ration the patch panel or hub connections if you don't at first have enough of these connections to handle every outlet (which you might not when you first set up your network), and then reallocate later as you install more connections. Also, if your department should move to a different part of the floor, or even to a different floor, you can still bring the new wiring back to the same wiring closet by simply hooking through the punchdown block.

### Then to a Patch Panel

The four wires that began at the outlet proceed through the punchdown block connections on to the patch panel. The usual way of making this connection is by means of a 25-pair cable that runs from a standard 25-pair connector on the punchdown block to a similar connector in the rear of the patch panel.

The patch panel is an array of RJ45 jacks (identical to the outlets at the desks) that gives you the easy flexibility you need to group your network nodes into one or several hubs, which may or may not be interconnected. Each non-connected hub or hub group represents a different physical network. By changing the connections between the patch panel and the hubs, you can move nodes from one network to another, remove problem nodes from the network and combine or separate networks.

The patch panel-to-hub connections can create a tangle of cables. Once you've connected 40, 60 or 80 patch cables between the patch panel and the hub, there will be so many of them that there is almost no way to trace them visually. That's why it's vital to document your wiring on the cables and outlets themselves.

Figure 3 shows how we do it for our own labs. Throughout the lab there are 10BASE-T wall outlets, each of which has its own number. This same number appears on the corresponding patch panel outlet. (That's why the punchdown block is considered semi-permanent. If you change a floor location on the punchdown block, you'll also have to change the corresponding patch panel label.)

Once you know which patch panel location you need to deal with, your next problem is finding the other end of the patch cable (the hub end). If your patch cables are unlabeled, this is a daunting task. The way we set things up is to put a unique identifying number at both ends of the patch cable. This identifying number could be anything that uniquely identifies the cable, but we use the number of the hub port into which the cable is plugged. This way, if you make a regular practice of connecting the patch cable to the corresponding hub port, it's easy to trace whether a given machine is active and connected. You should note which outlet number it's plugged into (every outlet should have its label clearly displayed), find the patch panel connector with the same number, note the number on the patch cable plugged into it, and that's the panel indicator on the hub you should be checking.

### Variations on the Patch

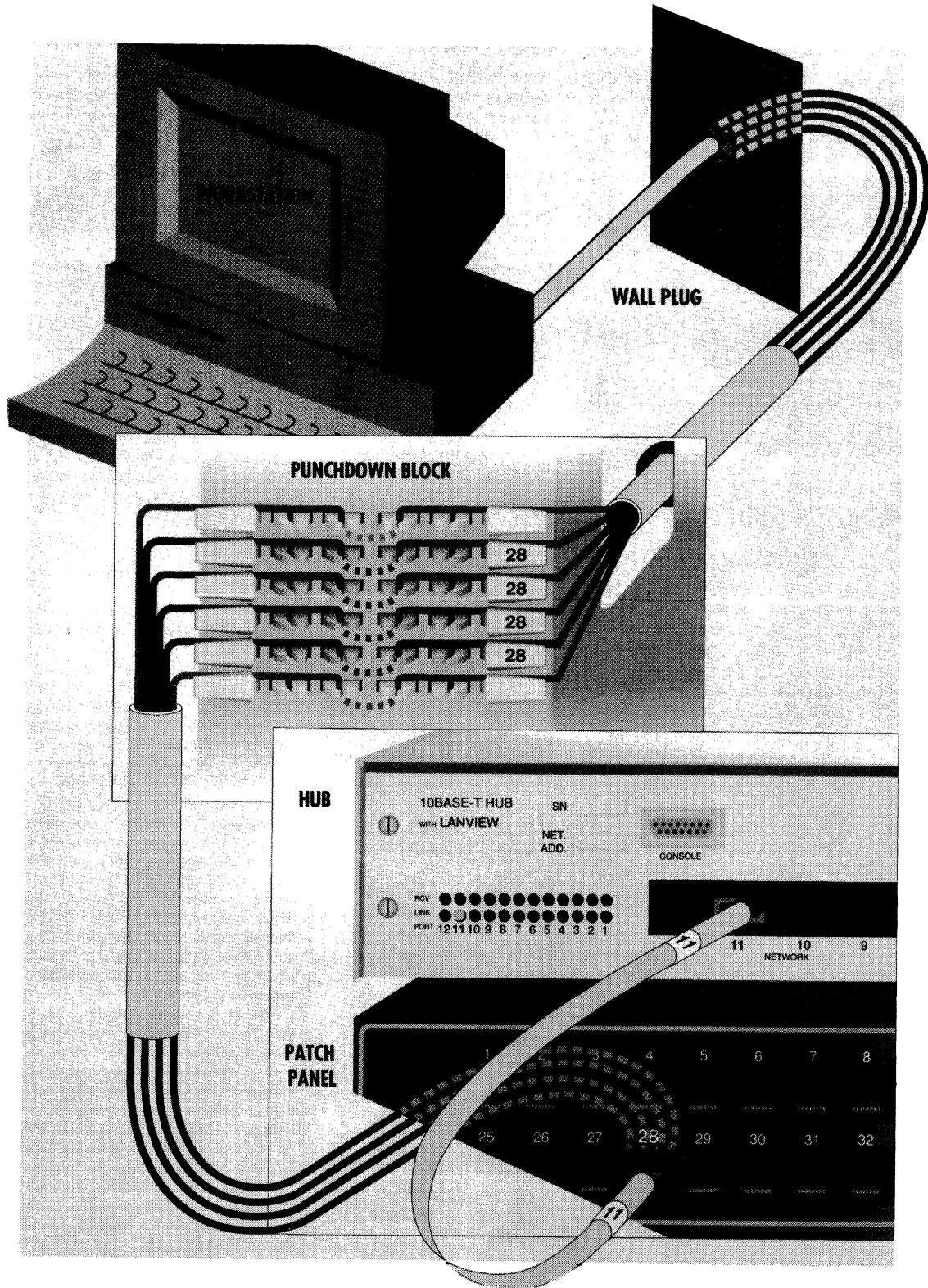
The arrangement illustrated here is, as we've already mentioned, the setup we use in our lab. We need a fair amount of configuration flexibility, and this requirement led to our decision to use a combination of punchdown block, patch panel and hub.

Don't be deterred if this arrangement seems overly complex for your needs or your budget. There are several simpler variations that will fulfill most of the same functional needs.

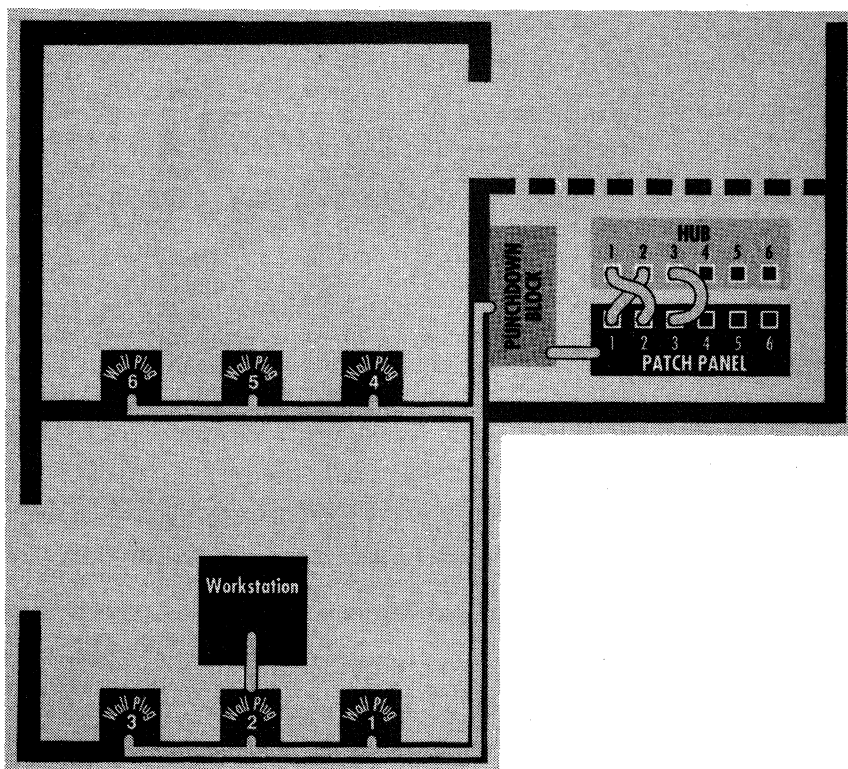
To decide whether you can use a downscale form of this setup, you first must decide on the long-term stability of your networking arrangement. If you believe that network users will remain where they are and that future expansion will be limited, you can feel justified in foregoing some configuration flexibility.

Many 10BASE-T hubs have an auto-partitioning capability, meaning that they will automatically isolate nodes that are causing trouble on the network. If the hub you are acquiring can do this, you will not be so dependent on patch panel flexibility for maintaining continuous network operation. Just make sure that auto-partitioning can handle all network error conditions, including adapter cards that can't seem to avoid collisions.

Figure 2.  
Connection from Desktop to Hub



**Figure 3.**  
**Connection to a Patch Panel**  
**Hub**



In terms of intermediate solutions, there is a variant form of the punchdown block that combines both punchdown and modular connectors. Each group of four punchdown Vs connects internally to a modular jack. This arrangement allows you to run patch cables between the punchdown block and the hub, giving you essentially the same reconfiguration control, but without the easier connection tracing capability of a regular patch panel. If you have only 20 or so patch connections, this is a reasonable compromise.

You can, of course, dispense completely with the patch panel. Most punchdown blocks and many 10BASE-T hubs can accommodate a standard 25-pair connector, allowing for easy hook-up between the two components. If you use this arrangement, you will be depending on the auto-partitioning capability of the hub, since it will not be easy to find and disconnect a malfunctioning node at the punchdown block.

Carrying things to a logical conclusion, you can even eliminate the punchdown block. You can instead connect the conductors from the wall outlet directly to the hub. This is actually fairly easy to do if the conductors converge into standard 25-pair connectors and the hub can accommodate such a connector. This is suitable for smaller installations, where flexibility is not a major factor.

If you avail yourself of any of these simplified connection arrangements, it's important to be sure that the wiring

is correctly and completely labeled by the installer. Normally this would be the case with punchdown blocks, but if you're not using a punchdown block or a patch panel, the labels must appear on the hub.

### Tricky Hubs

Be warned. While you may have set things up the way we've suggested here, using punchdown blocks and patch panels everywhere, someone may come along and move a patch cable. Everything may still work well, but when you begin to trace things and you check only the patch panel end of the patch cable, you may end up consulting a hub indicator that is no longer associated with that node. It's a good idea to find both ends of the patch cable to make sure you're checking the right indicator light.

One other hint: For file servers and other critical network nodes, mark the patch panel connector and both ends of the patch cable you're using for the server with something visually distinctive. A red stick-on dot will do the trick. Then you can tell at a glance which connections should be treated with special care.

Finally, it's worth repeating: Neatness really does count. Your wiring closet is where everything comes together. And anything you can do to keep the kinks out of the wires will help ensure error-free operation in the long run. And if the time comes when you do experience problems on your network—when users are keening in the corridors and managers are brandishing axe handles—a clearly documented wiring center will be one of your best friends. ■



# A Guide to Installing Windows on a NetWare LAN

## In this report:

Establishing Network Search Drives .....	2
Setting Up the Workstation .....	2
Working With High Memory .....	3
Using Swap Files .....	4
Managing Memory .....	5
Updated Utilities and NIC Solutions .....	5
The Master Configuration File .....	5
Tuning Network Applications .....	8
Printing With Windows .....	8

## Datapro Summary

There are salient benefits to using Windows 3.0 on a LAN—a graphical user interface, better memory management, Dynamic Data Exchange, and multitasking capabilities—the most obvious. However, Windows is very different from DOS, and that presents new challenges to network managers implementing the operating environment on a LAN. By following the proper guidelines, integrating Windows into your LAN can be a rewarding experience for all involved.

## The Benefits of Windows

Installing Windows 3.0 on your NetWare LAN doesn't have to be the horror story you've heard it can be. You can avoid problems with NICs that hang the file server and graphics drivers that generate snow instead of graphics. You can bypass problems with the setup utility, printers, and applications. This report will tell you how to integrate Microsoft Corp.'s Windows into your Novell Inc. network. Follow my advice, and your Windows installation should be virtually trouble-free.

Given the potential headaches, why should you consider running Windows on your network in the first place? Because it's worth it. For network managers, Windows is extremely adaptable: It works well with DOS-based hardware and software, and it's easy to manage when properly designed and installed. For users, Windows offers the benefits of a graphical user interface (GUI), multitasking, Dynamic Data Exchange, and better memory management. In addition, users generally find Windows' GUI easier to use than DOS's character-based interface. Both network managers and users come out ahead.

This Datapro report is a reprint of "Guide to Installing Windows" by Ethan Wilansky, pp. 45-56, from *LAN Technology*, Volume 7, Number 7, July 1991. Copyright © 1991 by M&T Publishing, Inc. Reprinted with permission.

Before you decide to try it, consider what it means to truly run Windows in a network environment. In a single file server LAN, Windows will reside as a shared program on the file server's drive. To minimize traffic in a multiple-server environment that includes routers or bridges, Windows should reside on each file server's hard disk to service that server's nodes. While it is possible to run Windows from a workstation hard disk, it's inefficient. Each Windows-based workstation would require a hard disk to store the Windows program files. Also, LAN management and Windows setup would be much more formidable tasks if each station had its own copy of the Windows files.

A single copy of Windows 3.0 is all that is necessary to provide access to Windows services for all network users. However, Microsoft requires that each user have in his or her possession an original Microsoft software license or an equivalent, which Microsoft will designate on an individual basis. If you have any concerns about running Windows legally, contact Microsoft.

## Configuring the Server Hard Disk

All right, you've accepted the challenge: Windows will run on your LAN. Now let's take the first steps in that direction. To begin, create a WIN directory for your Windows program in the shared application

area of the file server. Make this directory at the same level you store your other applications. Use the expansion utility that comes with Windows to place all the Windows files in the WIN directory. The Windows user manual has instructions for copying the decompression utility to the network and creating a batch file called EXPALL.BAT, which decompresses the Windows files and copies them onto the server's hard disk.

Using EXPALL.BAT is necessary when installing Windows on the network. Most of the Windows files are compressed to fit on five floppy disks, but a few files are not compressed. If you don't use EXPALL.BAT, the decompression and copying process will stop after processing an uncompressed file, and certain files may not be properly copied to the networked drive. In addition, you won't know which files have been processed and which remain to be decompressed and copied to the WIN directory.

After you have copied all of the Windows files to the shared program area, flag the files with the attributes "Shareable" and "Read Only." This ensures that they can be shared and read but not deleted or edited. Next, make sure that each person using Windows has appropriate rights to read and execute the Windows program files. These user rights are RF (Read Filescan) in NetWare 3.x and ROS (Read Open Search) in NetWare 2.x.

Finally, each Windows user will need to have a home directory to store Windows configuration files for his or her particular workstation. The location of the home directory will depend primarily on the presence or absence of a fast local hard disk and the file server's performance. Recent reports have suggested loading Windows user configuration files (UCFs) from a local hard disk to get reasonable Windows performance. Save your money. I have found that this is not necessary; Windows performance is much more dependent on file server performance.

At Riverbend Group Inc., we run a Tricord Systems Inc. PowerFrame 486/25 file server. Our PowerFrame contains 8 Mbytes of RAM, a 314-Mbyte hard drive, and two 16-bit network interface cards—one for Ethernet and one for Arcnet. Although many of our workstations contain hard disks, we haven't loaded the UCFs onto these hard disks. Instead, we install the UCFs under each user's home directory on the network drive. By default, Windows creates a temporary swap file on the network hard drive. Windows uses the swap file area on the network disk as virtual RAM to store information that won't fit in workstation RAM. Comparing UCF loading for 80386 workstations with and without local hard disks, we found no significant difference in Windows' performance.

Before we continue, let's consider the non-Windows user. You might be thinking, "Why should I worry about users who prefer to work in a DOS-only environment?" You shouldn't. The key to keeping your non-Windows users happy is devising a workable mapping structure. In Riverbend's network, I used Novell's MAP ROOT enhanced mapping utility to add some search mappings and to map root existing regular drives and search drives to the system login script. These new mappings only affect the Windows users.

The MAP ROOT utility is similar to the DOS SUBST command and creates a pseudo- or fake-root directory, so a drive letter may be associated with a specific directory on the network. MAP ROOT lets Windows' File Manager look down through the network and local drive directories to see where you have mapped your drive letters. Otherwise, no matter where you map a drive, Windows will always start at the volume level, or true root, of the directory

structure. And remember: When using MAP ROOT, convert global batch files and menus to this MAP ROOT standard for proper batch file execution.

---

## Establishing Network Search Drives

Unlike other applications running in a LAN environment, Windows requires two search directory areas on the server's hard disk. The first search drive points to each user's Windows initialization file area, and the second points to the shared Windows application area. This order is important since .INI files with the same prefix but different contents will exist in these two search areas. The UCFs in the user's initialization file area must be read first, since they are customized to each user's environment.

The same .INI files reside in the Windows application area for several reasons. First, when a new Windows user is set up, some of the .INI files in the WIN directory will be copied to the user's initialization directory. Second, changes to the .INI files in the application area can easily be made to the .INI files in the users' initialization area. For example, if I load a new Windows application such as Microsoft Word on the network, that application automatically writes configuration information to WIN.INI. The network manager can then cut and paste the new information to individual UCFs of users who want to use the new application.

Only two regular network drives need to be mapped for the Windows users—one must be map rooted to each user's home directory area, and the other is mapped to the root or volume level above the applications your users will execute. Pointing the second drive to the root accommodates utilities that may be above the application level of your structure, such as the network utilities found in Novell's PUBLIC directory. The figure provides an example of the necessary mappings.

When writing these map commands, consider using the MAP INS ROOT syntax. The INS parameter inserts a search path into a path statement without deleting search paths that already exist. This is especially important for users with local hard disks and path statements in their AUTOEXEC.BAT files. Their current search mappings to the local disk will be appended to the end of the network search mappings if you include the INS parameter. Another alternative is to append the network searches to the end of your users' search paths by using the MAP ROOT S16 command. As long as fewer than 16 search drives exist, all searches will be kept intact.

---

## Setting Up the Workstation

The Windows files are now in the shared application area on the server's hard disk. You have set up the directory structure and created the necessary mappings to run Windows. Now, take an inventory of the machines that will run Windows. Determine each machine's processor, RAM size, NIC, and graphics card and monitor. (See the sidebar, "Upgrading Graphics Cards and Monitors for Windows," for more on these components.)

Note the presence or absence of a local hard drive and the access speed of that hard drive. On an 80386 workstation with a fast hard disk, you might consider using the hard disk for loading the UCFs, Windows' SMART-DRV.SYS disk caching program, and a permanent swap file. As noted earlier, a local hard disk may be somewhat faster and reduce network traffic, but it does not improve Windows network performance substantially. This report

## Upgrading Graphics Cards and Monitors for Windows

Everyone knows Windows looks good, but *how* good depends on the type of graphics card and monitor at your workstation. The most commonly used graphics standards for running Windows are Hercules monochrome, EGA, standard VGA, and HIRES VGA. The quality of monochrome, EGA, and standard VGA cards and monitors are fairly consistent throughout the industry, and Windows contains the graphics drivers needed to run these three standards.

However, even standard VGA screen resolution is barely adequate for a good view of your Windows resources. For instance, unless you are in the draft view mode in Microsoft Word for Windows, you won't see the full width of a standard 8.5- x 11-inch page. For a better view, move up to a higher resolution graphics package. This upgrade requires a graphics card and monitor that will support high resolution display and a Windows

high resolution driver specifically designed for that card. The driver will be provided with Windows or your graphics card. If the card manufacturer supplies the driver, you will find it on a floppy disk accompanying the card.

The most common VGA high resolutions are 800 x 600 and 1024 x 768. Because these resolutions are industry standards, the graphics drivers to run these resolutions should be standardized as well, right? Wrong. All Windows graphics drivers are not created equal. In fact, they're often created quite poorly. As a result, some card manufacturers must significantly revise their drivers until all the bugs are ironed out.

After testing a few cards and their related drivers, I have found a couple of mature, debugged drivers. The ATI Technologies Inc. Wonder cards, the Orchid Technologies ProDesigner cards, and the Headland Technology

Inc. V-Ram card all have reliable graphics drivers. I'm sure there are many other excellent drivers for other cards, but I have found that these cards are reliable. Note the cards mentioned above: Some of the companies' lower-end units have not lived up to their high-end counterparts.

One common problem I have encountered with other cards is severely narrowed horizontal screen margins when running in high resolution mode. Another problem is snow appearing on the screen after a few minutes of Windows operation. For trouble-free displays, make sure that the third-party Windows graphics drivers are mature and debugged. Also, make sure that the card was designed to handle high resolution modes without sacrificing your screen view.

Another item to look for is a graphics monitor and card that can handle non-interlaced screen refresh. Non-interlaced-compatible monitors and graphics cards are fast, so they can update the monitor screen in one continuous pass, or write, to the screen. Interlaced monitors and cards are slower, so they must update the screen by refreshing every other horizontal line per pass. Thus, the entire screen is

refreshed in two passes rather than one. For an instant, there are two images on your screen.

Many people don't readily notice the difference between interlaced and non-interlaced screens. How do you know which screen type you need? Work on an interlaced screen for several hours until your eyes begin to fatigue. If you begin to see the screen blinking at this point, it may be wise to purchase the more expensive, non-interlaced monitor and card technology.

You must look at both the card and the monitor when determining your final graphics capabilities. For example, a card that can output 800 x 600 resolution can only do so if the monitor can display 800 x 600 resolution as well. If you decide on a high resolution card, make sure it is at least a 16-bit card. This will provide adequate screen creation and refresh speed. Also a minimum of 256 Kbytes of video RAM will allow for 16-color VGA display. While Windows can run on just about any monitor on the network, an upgrade of your graphics resources will give your users the best possible Windows view.

presumes that users will download their UCFs into workstation RAM directly from the file server.

The workstation configurations I discuss provide the minimum acceptable Windows performance on a network. If your workstation contains an 80286 CPU, you can operate in a Windows environment with impressive speed using task switching, a process whereby two applications are swapped between conventional memory and extended memory. However, you should invest in additional RAM to bring these workstations up to 4 Mbytes.

When loaded before starting Windows, the RAM-DRIVE.SYS program provided with Windows creates a virtual disk drive in the 80286 workstation's RAM. Windows uses this RAM disk area to task switch between itself and non-Windows applications. With 4 Mbytes of RAM, you can allocate 2 Mbytes to the virtual disk drive and 2 Mbytes to Windows applications. (Windows applications don't use the RAM drive area for program memory.) Next, set the temporary Windows directory that is normally created by SETUP.INF (and directed to a networked hard drive) to the virtual disk drive RAM area to make task switching lightning fast.

### Working With High Memory

You will also need to load Windows' HIMEM.SYS program, which manages the high memory area in your workstation RAM. To accommodate the use of high memory with Windows, the CONFIG.SYS and AUTOEXEC.BAT files might contain the following settings:

CONFIG.SYS:

```
Files = 60
Buffers = 30
Device=A:\HIMEM.SYS
Device=A:\RAMDRIVE.SYS2048 /e
```

AUTOEXEC.BAT:

```
C:
Md Temp
Set temp=C:\temp
```

These settings are for a diskless workstation with A: as the boot drive and C: as the next available free drive for RAM-DRIVE.SYS to use. The file and buffer settings in CONFIG.SYS may vary. If there is a local hard drive, create a

**Figure 1.**  
**Map Rooted Drives**

*The drives defined here have been map rooted to accommodate Windows in a NetWare environment. The spaces before the last slash in each drive and search drive path are a result of the MAP ROOT command.*

DRIVE F: = FILESERVER\WORK: \DATA\USERS\USERNAME \  
Drive F: points to the user's home directory area on the network.

DRIVE G: = FILESERVER\SYS: \  
Drive G: points to the root level, so Windows can point to any application on the network through Program Manager's File-Properties option. This decreases the need for extra search drive pointers.

SEARCH1: = Z:. [FILESERVER\SYS: \PUBLIC \ ]  
Search drive Z: lets non-Windows users point to network utilities within the public directory.

SEARCH2: = Y:. [FILESERVER\SYS: \APPS\OS\V3.30 \ ]  
Search drive Y: lets the workstation find the DOS command interpreter and related files for operating DOS and refreshing workstation RAM when exiting applications.

SEARCH3: = X:. [FILESERVER\WORK: \DATA\USERS\USERNAME\WIND \ ]  
Search drive X: points to the UCFs for each user accessing Windows. X: could be set up for a group of Windows users, so users would not see these mappings unless they were members of the Windows group.

SEARCH4: = W:. [FILESERVER\SYS: \APPS\WIN \ ]  
Search drive W: points to the shared Windows application area. W: could also be set up for a Windows user group, so only group members would see these mappings.

temporary directory in AUTOEXEC.BAT based on the virtual drive that RAMDRIVE.SYS finds on boot-up. Change from the temporary drive back to your local boot drive to load any additional files and connect with the network. AUTOEXEC.BAT will also contain the necessary calls to connect with the network.

If you decide to run Windows on an 80386-based diskless workstation, all you need to load into memory is the HIMEM.SYS program. In the 386 Windows enhanced mode, both Windows and non-Windows applications are handled in the same way. It is unnecessary to create a virtual RAM disk drive for the standard, DOS-based applications since they will not be removed from your first 640 Kbytes of memory in the way an 80286 machine removes them. Instead, all applications run in virtual 8086 machines created in system RAM by the 80386 processor.

Naturally, the more RAM you have, the more 8086 virtual machines you can create, but 4 Mbytes of RAM is a good starting point. Create a temporary swap file to a network drive if you need more virtual RAM space while in 386 enhanced mode.

To accommodate a 80386 diskless workstation running in enhanced mode, make the following settings in the CONFIG.SYS and AUTOEXEC.BAT files:

```
CONFIG.SYS:
Files = 60
Buffers = 20
A:\HIMEM.SYS
```

```
AUTOEXEC.BAT:
Set temp=F:\temp
```

These settings are for a diskless workstation with A: as the boot drive. The temporary directory shown in the AUTOEXEC.BAT file sits on the same level as the WIN directory, which is below each user's home directory. F:\ points to the user's home directory via MAP ROOT. AUTOEXEC.BAT will begin with the necessary calls to connect with the network.

Along with AUTOEXEC.BAT and CONFIG.SYS, all workstations require another startup file, SHELL.CFG. The NetWare shell reads this file, which should be placed at the root of your boot disk. You can create this file with

any standard text editor. At a minimum, SHELL.CFG will contain the following settings:

```
SHELL.CFG:
Show dots = on
File handles = 60
```

The first setting allows you to see the double dots, which represent parent directories, in the Windows directory structure. If you can see the double dots, you can move around conveniently and easily in the directory. The second setting increases your 40 default file handles to 60. File handles let you open files simultaneously, and you can easily exceed the default of 40 when opening multiple files in Windows.

Note that this setting matches the "Files =" number in the CONFIG.SYS file. These two settings should be the same. If an application requires a higher number for "Files =" in CONFIG.SYS, make the corresponding change to "File Handles =" in SHELL.CFG.

## Using Swap Files

Experienced Windows users may notice that I haven't added SMARTDRV.SYS to the CONFIG.SYS files. This program cannot cache memory to the server's hard disk. If your workstation does not have a hard drive, SMARTDRV.SYS does nothing except tell you that there are no hard drives on your system. If you have poor local hard drive performance, you will want to avoid loading SMARTDRV.SYS; using it could reduce Windows' speed. For more speed when task switching in 80286 workstations, use RAMDRIVE.SYS.

If you have a local hard drive and run in 386 enhanced mode, you can create a permanent swap file on your workstation hard disk with SWAPFILE.EXE. This Windows program reserves a portion of your hard disk and lets Windows write to it directly. This program makes it appear to Windows that you have more workstation RAM to load applications than you actually have. However, a permanent swap file to your local disk will do little to improve Windows' performance if you truly run Windows on the network.

Alternatively, Windows will automatically create a temporary swap file for a diskless workstation in the 386

enhanced mode section of your SYSTEM.INI file, the Windows file that stores workstation hardware device settings. When you choose the HELPABOUT option from the Windows Program Manager screen, you will see much more memory than is actually available in your workstation.

Although it sounds like a good idea to set up a temporary swap file on the server's hard disk, I don't recommend it all the time. Network traffic will increase as applications loaded in the server's swap file move between the server's swap file and the workstation's RAM. If you run out of RAM space too quickly, you can either create a temporary swap file or buy more RAM for your workstation. If you have a lot of network traffic already, it is wiser to add workstation RAM instead of setting up temporary swap files. Also, opening an application that has been placed in the swap file area takes nearly as long as opening an application through standard disk I/O.

## Managing Memory

The HIMEM.SYS program was placed in the CONFIG.SYS files for both workstations described above. You could use a different memory manager, such as Quarterdeck Office Systems' QEMM-386 for 80386 workstations. QEMM-386 includes files that let you load DOS resources, such as files and buffers; TSRs; and even your network driver into high memory. Using a third-party memory manager that lets you load TSRs into high memory frees up more of the first 640 Kbytes of conventional memory than HIMEM.SYS. However, if you have 4 Mbytes of workstation RAM, the first 640 Kbytes do not have to be left empty.

I recommend removing unnecessary TSRs and sticking with HIMEM.SYS. It provides excellent performance on both CPU platforms. In the NetWare world, HIMEM.SYS loads all but 6 Kbytes of the XMSNETX.EXE shell into the extended memory area above 1 Mbyte to free up some conventional memory.

One TSR that is necessary to run Windows in a NetWare environment is the DOS shell—either conventional or XMS. The NetWare shell for DOS serves two functions: It finds the nearest file server upon workstation start-up, and it directs calls to the DOS command interpreter and the network operating system.

Novell updated the standard shell, NETX.COM, and provided a new shell, XMSNETX.EXE, to take advantage of extended memory and to accommodate Windows. If you need to run the expanded memory management shell, you will need a LIM 4.0-compatible expanded memory manager to load EMSNETX.EXE into the high memory area.

One major drawback to using EMSNETX.EXE for Windows is that Windows will only run in real mode, and your workstation will perform like a single 8086 machine. If you can run in 386 enhanced mode with HIMEM.SYS, you can emulate expanded memory with extended memory, so running in straight expanded mode is not necessary in the 386 environment. Although the extended memory shell is larger than the conventional memory shell, NETX.COM, all but 6 Kbytes load above the conventional memory area.

If you do not have the XMSNETX.EXE shell, it's available in Novell's DOS/Windows workstation update kit. This update kit also includes other critical files for running Windows in a NetWare environment, such as IPX.OBJ. IPX.OBJ is linked to your NIC driver to create your IPX.COM file. If the upgrade kit doesn't include your NIC

driver, call the NIC manufacturer for its Windows-compatible driver or check CompuServe's NetWire forum.

If possible, do not set IPX.COM to IRQ2, IRQ9, or higher to run Windows. In enhanced mode, Windows may come in conflict with these settings. If you must use IRQ2, IRQ9, or higher, replace the VPIC Windows file with VPICDA.386, which is available on NetWare. You must also alter the 386 enhanced mode section of SYSTEM.INI to accommodate this new driver.

## Updated Utilities and NIC Solutions

The DOS/WINDOWS workstation update kit also includes some updated public and system utilities such as MAP.EXE and BINDFIX.EXE. MAP.EXE has been enhanced to allow you to map root your network drives. You must update all of the utilities included in the DOS/Windows workstation update kit, or you may encounter devastating results when using the older utilities with certain Windows settings. For instance, the SHELL.CFG workstation startup file contains the "Show Dots = on" setting to accommodate Windows. If you run the old BINDFIX.EXE system utility file, the utility can get caught in a loop of deleting files and directories that you don't want erased.

There have also been several revisions of the shells to provide patches for bugs, so make sure that the NetWare shells you are using are 3.01 Revision E. One serious problem that the 3.01 E shell addresses relates to dynamic memory usage. In NetWare 2.x file servers, earlier Windows shells would hold on to dynamic memory until there was none available for operating the network, and the file server would eventually hang.

Last March I attended a seminar where I overheard the following comment: "You're using Arcnet cards? Don't even think about running Windows—the throughput is so slow you will fall asleep between command requests to and from the file server." This statement is simply not true. I ran both an 8-bit Arcnet card and an 8-bit Ethernet card at a workstation, and the difference in speed was nominal. If you are using Arcnet cards, see if you can replace the standard Arcnet driver with a turbo driver such as Novell's TRXNET, which lets you send four packets rather than one in an Arcnet environment. When network traffic is heavy, using IPX.COM generated with the turbo Arcnet driver will improve workstation performance.

I have tested both 8-bit and 16-bit Arcnet and Ethernet cards and have come to this conclusion: It is much more important to have 16-bit or even 32-bit NICs in the file server—if your bus architecture can support it—than in the workstations running Windows. As mentioned earlier, excellent server performance is critical to Windows' network performance. If you are having problems running your workstation NIC with Windows, make sure you have the latest version of that NIC's driver to bind to the latest version (3.02) of IPX.OBJ.

## The Master Configuration File

SETUP.INF is the text file used by Windows to set up each Windows user. This file, when properly altered, can be a Windows network administrator's dream. One of my co-workers accurately described it as the Windows file you configure to your network before using it to configure your Windows users' workstations.

Configuring this file to your network means making it aware of all the resources on your network, including printers, NICs, graphics drivers, and network applications. You

## Setting Up Users With SETUP.INF

SETUP.INF is a Windows file that contains information about your network's resources. The file also contains information that will not pertain to your network. For instance, you won't use the Banyan Systems Inc. information in the network section if you do not run VINES. You can enhance SETUP.INF by adding any necessary resource information not in the original file. Manufacturers place this information on floppy disks that accompany their products. You can also enhance SETUP.INF by rearranging the code to place your network information at the front of each resource section.

Once SETUP.INF is modified, you can easily set up your users. Portions of the Windows 3.0 code as displayed in SETUP.INF are shown below. The order of the code has been rearranged to accommodate Riverbend's network. Brackets represent the headings for each section of the SETUP.INF file. The Riverbend changes to this file are in bold letters, and the italicized text describes changes to the section it precedes. The file contains representative selections from both the original SETUP.INF file and our modifications.

*The modification below will display the message shown in the Windows caption boxes upon executing the Windows setup procedure.*

```
[dialog]
caption = "Riverbend
Group Windows Setup"
```

*You can change the DEFDIR to the directory where you are storing the user's .INI files. F:/ is map rooted to each user's home directory and the WIND subdirectory contains the UCFs.*

```
defdir = F:/wind
```

*You can add any video driver that is Windows 3.0-compatible to the list. If a third-party video driver is not on the video setup list, you can load it via floppy disk. The disk contains the program's drivers and setup information that will be added to the user's SYSTEM.INI file. By adding the information to the list below and copying the corresponding graphics drivers to the Windows application area, you can accommodate any graphics card that has its own Windows graphics driver.*

```
profile = driver, description
of driver,
resolution, 286 grabber,
logo code, VDD,
386 grabber, ega.sys, logo
data
```

```
8514 = 1:8514.driv, "8514/
a",
"100,120,120",
3:vgacolor.gr2,
2:vgalogo.lgo,
4:vdd8514.386, 4:8514.gr3,,
2:vgalogo.rle
cga = l:cga.driv, "CGA",
"200,96,48",
3:cga.gr2, 2:cgalogo.lgo,
4:vddcga.386,
4:cga.gr3,, 2:cgalogo.rle
plasma = 1:plasma.driv,
"Compaq Portable
Plasma", "100,96,96",
3:cga.gr2,
2:cgalogo.lgo,
4:vddcga.396,
```

```
4:plasma.g3,, 2:cgalogo.rle
egahires = 1:ega.driv,
"EGA", "133,96,72",
3:egacolor.gr2,
2:egalogo.lgo, x:*vddega,
4:ega.gr3, 3:ega.SYS,
2:egalogo.rle
egahibw = 1:egahibw.driv,
"EGA black and white
(286 only)", "133,96,72",
3:egacolor.gr2,
2:cgalogo.lgo,,, 3:ega.SYS,
2:cgalogo.rle
```

```
v760016 = 1:v760016.driv,
"Video Seven VGA
800x600 with 512K",
"100,96,96",
3:vgacolor.gr2,
2:vgalogo.lgo, x:*vddvga,
4:v7vga.gr3,,
2:vgalogo.rle
w31600 = 1:w31600.driv,
"800x600x16 for
CHIPS Super VGA
82C451",
"100,96,96",
3:vgacolor.gr2,
2:vgalogo.lgo,
x:*vddvga, 4:vga.gr3,,
2:vgalogo.rle
vga54 = 1:win3-54b.driv,
"Mode 54h (800x600
16-color) display driver
V2.01",
"100,96,96",
3:vgacolor.gr2,
2:vgalogo.lgo,
4:vddvga.386, 4:vga.gr3,,
2:vgalogo.rle
vga55 = 1:win3-55b.driv,
"Mode 55h (1024x768
16-color) display driver
V2.01",
"100,120,120",
3:vgacolor.gr2
2:vgalogo.lgo,
4:vddvga.386, 4:vga.gr3,,
2:vgalogo.rle
vga61 = 1:win3-61b.driv,
"Mode 61h (640x400
256-color) display driver
V2.01",
"100,96,96",
3:vgacolor.gr2,
2:vgalogo.lgo,
4:vddvga.386, 4:v7vga.gr3,,
2:vgalogo.rle
```

*Originally, Novell came last in the SETUP.INF list of networks installed. Since Novell is our primary network, I moved it to the first position. This alteration was specifically made for convenience when I ran SETUP.INF without hardware detection, following the setup procedure with the /I parameter. This parameter was used on workstations with Arcnet cards.*

```
[network]
;Prof Driver, Description,
HelpFile,
;OptFile, WininiSectName,
SysiniSectName,
;VDD, VDD, n...
;
novell = 2:netware.driv,
"Novell Netware 2.10
or above, or Novell Net-
ware 386",
4:netware.hlp,
4:nwpopup.exe, Novell_net,,
x:*vnetbios, 4:vnet-
ware.386, 4:vipx.386
3comopen = 2:msnet.driv,
"3Com 3+ Open LAN
Manager (XNS only)",,,,
3Com_net,
x:*vnetbios, x:*dosnet,
4:lanman10.386
3comshare = 2:msnet.driv,
"3Com 3+Share",
,,,3Com_net,x:*vnetbios,
x:*dosnet
banyan = 2:msnet.driv,
"Banyan VINES 4.0",
,,,Banyan_net,x:*vnetbios,
x:*dosnet,
4:baninst.386
```

```
lanmanlx = 2:msnet.driv,
"LAN Manager 1.x
(or 100% compatible)"
,,,,x:*vnetbios,
x:*dosnet, 4:lanman10.386
```

*In this section, I added the .INI files for Adobe Type Manager and Microsoft Word, so they are automatically copied to the UCF area. NORMAL.DOT is the macro file that we modified for Word and copied to the UCF area.*

```
[net]
2:CONTROL.INI, "Windows
```

can rearrange SETUP.INF's existing code, so you don't have to scroll through five screens to get to your specific resource code. I have found that you can also add third-party .INI files to each UCF directory, add graphics and

printer drivers, and customize almost all parts of this file. (See the sidebar "Setting Up Users with SETUP.INF.")

You can change the configuration of SETUP.INF to match your network, such as pointing to network search

User Files"  
2:WINVER  
2:ATM.INI  
2:WINWORD.INI  
2:NORMAL.DOT

The eight names below represent the group windows that will be created for each user upon setup.

```
[progman,groups]
WinApps,1
Std
LanWork
Extras
NetWare
Tools
Control
Games
```

Each group window is titled to represent the beginning of that window's contents. The description of each icon follows. On the same line of each description is the path to the executable file or to the PIF file created to call the executable. Drive G: has been mapped to the root of the application directory. In *SETUP.INF*, the same line includes the file path and a path to an icon or icon library. (Here, we've placed the icon path below the file path.) If an icon library is called, a number is specified to point to a specific icon in the library. (This icon library does not come with Windows.)

```
[WinApps]
"Word", g:/apps/win/
wrw/winword.exe
"Designer", g:/apps/win/
mgx/designer.exe
"Excel", g:/apps/win/
xcl/excel.exe
```

```
[Std]
"tnt", g:/apps/win/
pif/tnt.pif,
g:/apps/win/icn/
tnt.icn
"cc:Mail", g:/apps/win/
pif/ccm.pif,
g:/apps/win/icn/
iconlib.exe,107
"WordPerfect",
g:/apps/win/
pif/wp.pif,
g:/apps/win/icn/
```

```
wp50.icn
"Lotus 123", g:/apps/win/
pif/123.pif,
g:/apps/win/icn/
iconlib.exe,1
```

```
[LanWork]
"BulBrd", g:/apps/win/
wrk/wwboard.exe
"Extension", g:/apps/win/
wrk/wwext.exe
```

```
[Extras]
"Paintbrush", g:/apps/win/
pbrush.exe
"Terminal", g:/apps/win/
terminal.exe
"Recorder", g:/apps/win/
recorder.exe
"Cardfile", g:/apps/win/
cardfile.exe
```

```
[NetWare]
"Syscon", g:/apps/win/
pif/n_sys.pif,
g:/apps/win/icn/
earthsun.icn
"filer", g:/apps/win/
pif/n_flr.pif,
g:/apps/win/icn/
folder.icn
"Session", g:/apps/win/
pif/n_ses.pif,
g:/apps/win/icn/
iconlib.exe,94
"Map", g:/apps/win/
pif/n_map.pif,
g:/apps/win/icn/
map-us.icn
"Salvage", g:/apps/win/
pif/n_sal.pif,
g:/apps/win/icn/
doorway.icn
"Help", g:/apps/win/
pif/n_help.pif,
g:/apps/win/icn/
lightblb.icn
"ListDir", g:/apps/win/
pif/n_lstdir.pif,
g:/apps/win/icn/
iconlib.exe,92
```

```
"Ndir", g:/apps/win/
pif/n_ndir.pif,
g:/apps/win/icn/
harddriv.icn

[Tools]
"Notepad", g:/apps/win/
notepad.exe
"PIF Editor", g:/apps/win/
pifedit.exe
"V4.00 DOS", g:/apps/win/
pif/dos40.pif,
```

```
g:/apps/win/icn/
cprompt.icn
"V3.30 DOS", g:/apps/win/
pif/dos33.pif,
g:/apps/win/icn/
dos3.icn
```

```
[Control]
"Panel", g:/apps/win/
control.exe
"Print Q", g:/apps/win/
printman.exe
"WinSetup", g:/apps/win/
setup.exe
```

```
[Games]
"Solitaire", g:/apps/win/
sol.exe
"Reversi", g:/apps/win/
reversi.exe
"Eyes", g:/apps/win/
icn/eyes.exe
"Horse", g:/apps/win/
icn/horse.exe
"Tetris", g:/apps/win/
wep/tetris.exe
```

You must be cautious with the changes you make to the system section since they directly affect the creation of the *SYSTEM.INI* file. Placing semicolons in front of the lines you change will temporarily remove them from the file. To allow for Adobe Type Manager's font management, these changes are copied to the *SYSTEM.INI* file when you run setup.

```
[system]
; The various
SYSTEM.DRV, SOUND.DRV,
; COMM.DRV
;
; These are the drivers
which may vary
; from system to system,
but are selected
; only by the [machine]
menu — they do not
; have special menus for
their selection.
system = 1:atmsys.dr
system = 1:system.dr
sound = 1:sound.dr
comm = 1:comm.dr
hpsystem =
1:hpsystem.dr
```

When *SETUP.INF* is executed, the three printers in bold letters show up first in printer

selection through Windows Control Panel. Choose the printer drivers that correspond to the printers you have on your LAN(s).

```
[io.device]
; (printers, plotters, etc.)
; The filename is followed
by the
; descriptive string which
will appear in
; Control Panel and in
WIN.INI. One or two
; strings indicating the scal-
ing for this
; device.
; There may be more than
one line for a
; driver, corresponding to
different
; printers.
```

```
5:HPPCL.DRV, "HP Laser-
Jet [PCL / HP LaserJet]",
"DEVICESPECIFIC"
5:PSCRIPT.DRV, "TI Omni-
Laser 2108 [PostScript
Printer]",
"DEVICESPECIFIC"
5:TOSHIBA.DRV, "Toshiba
P351", "100,180,180"
5:TTY.DRV, "Generic / Text
Only",
"DEVICESPECIFIC"
5:PSCRIPT.DRV, "Agfa
9000 Series PS
[PostScript Printer]",
"DEVICESPECIFIC"
```

Once it's set up for your network, *SETUP.INF* goes a long way toward automating the creation of user configuration files. You save a lot of time and effort by cutting and pasting information from this master file to each user's file. The alternative is creating the UCFs with *SETUP.INF* and customizing each UCF later with floppy disks that contain the code not found in *SETUP.INF*. When one user needs code for several resources not found in the original *SETUP.INF*, or when several users need the same piece of code not in the file, the extra floppy-disk step can be significant. So, spend some time tweaking *SETUP.INF* to your network and make it work for you.

drives, adding high resolution graphics drives, and rearranging network lists. When configuration is complete, running *SETUP.INF* on each workstation becomes a much less formidable task than making configuration changes to each workstation from scratch.

Use your imagination when altering this file, but modify *SETUP.INF* carefully. Place a semicolon in front of the line you wish to edit to temporarily remove a statement

from the file without erasing it completely. The semicolon functions within SETUP.INF as the DOS REM statement does within a batch file.

Write your change and run SETUP to be sure it works before deleting the original line. Since this file is in the Windows application directory, make sure it is flagged for editing before you begin to modify it. If you have a set of identical machines, run SETUP once and copy the UCFs to each identical workstation.

Look out for a few things when you run the setup procedure. First, make sure the user mappings do not point to existing .INI files other than those in the Windows application directory. Second, create the UCF directory with SETUP before you get started, so the user can map to that directory. Make sure to type SETUP /N to let Windows know that you would like to copy over only the necessary files to the UCF directory. If you don't follow SETUP with /N, the procedure will copy all files for running stand-alone Windows to the UCF directory.

Finally, if you are going to use Arcnet cards, turn workstation hardware detection off by typing SETUP /I /N. Windows' hardware detection conflicts with any NIC that uses IRQ2, IRQ9, or the RAM address 2E0H, and the workstation will hang during the setup procedure.

The /I parameter keeps Windows from attempting to determine the type of hardware you have. If you use this parameter, you will have to read through the machine settings in the setup procedure and change them from the default configuration to match the workstation's actual hardware configuration.

To allow for easy installation of Windows applications, you should keep a WIN.INI file in the Windows application directory. Many Windows-based applications need to write information to a WIN.INI file, so you can allow this to occur in the application area.

Once the WIN.INI file has been appended by the application, you can append the personal WIN.INI files of all users who may use the new application by cutting and pasting with Windows' notepad or some other text editor. Make sure the global WIN.INI file is flagged Read/Write before installing a new Windows application that needs to append WIN.INI.

## Tuning Network Applications

To run non-Windows programs on either an 80286 or 80386 workstation, I advise setting up program information files. PIFs reside on the file server and tell Windows exactly how to run non-Windows applications. Each application has defaults that are automatically copied to each individual's UCFs, but PIFs fine-tune network applications and optimize their performance under Windows. Place all PIFs in a PIF subdirectory of the WIN directory for central or global management by the designated Windows administrator.

Thoroughly read the section on PIFs in the Windows 3.0 user manual before creating these files, and note two important considerations. First, if you set parameters in the PIF, the parameter line within the PIF file will not transfer between the 386- and 286-mode PIFs. So, if you set PIF parameters in 386 mode, switch the PIFs to 286 mode and set the parameters in these PIFs as well. Now you will have one PIF with two groups of settings, one for 286 mode and one for 386.

Second, if the DOS application you are running closes itself out without a command to do so, deactivate the "Close Windows on Exit" box in the PIF screen. Now the user will see the screen output when executing the program from its PIF before the PIF returns control to the Windows screen. For instance, if you execute a directory command without deactivating "Close Windows on Exit," the directory will flash on the screen and close out before you have a chance to read it. This PIF feature is especially important when you want to run command line-driven network utilities.

## Printing With Windows

Printing through Windows is possible. In fact, it's not as difficult as you might think. The first thing you must do in a NetWare environment is to capture your printers to a queue or a job configuration.

Because Print Manager gets the job from the Windows application, the application is released from the print job when the job has been completely spooled. There is no reason at this point to use a TIME-OUT setting or the NO AUTOENDCAP setting for printing. Instead, keep AUTOENDCAP enabled (this should be your default capture setting), and the TIME-OUT will be automatically set to zero.

This suggestion is contrary to the suggestion to use capture flag NA for NO AUTOENDCAP in Novell's *NetWare Application Notes*, January 1991. However, this setting has worked well for Riverbend, even for printing complex graphics. No form feed or tabs should be set, since Windows handles form feeding and tab formatting automatically. I recommend using the following capture statements:

```
CAPTURE Q={queue name} NFF NT AU NB L=1
```

I set the NB flag to keep a banner from printing, but this setting is entirely up to your preference. The L flag can be set to one, two, or three. Use a different number for each capture statement. You are limited to three printers captured at a time. This limitation is network-, not Windows-, specific. Start with three printers that are common to the company, regions, or groups.

Through the Printers window on Windows' Control Panel, each user can capture other queues on the fly if necessary. If you use job configurations, most of your capture statement flags would be contained within the print configuration. In this case, your capture statement would look like this:

```
CAPTURE J={job configuration name}
```

Make sure to invoke your captures before running SETUP. By taking this initial step, you can associate the printer drivers to the capture statements without having to connect each printer to the network through the Windows printer setup procedure.

I have found that the native NetWare printing services provided with NetWare 3.x and now ported into NetWare 2.15c and later are fine for Windows network printing. So, don't buy a third-party printing utility for Windows' sake; it doesn't need it.



Once you're up and running with Windows, beware of the phenomenon I call scapegoating, which manifests itself in remarks such as, "It must be Windows' fault." From now on, users will be able to conveniently blame any mysterious network problem on Windows. While it is a wonderful productivity enhancement tool, Windows 3.0 is also radically different from the DOS world, so the revolutionary environment naturally leads to product mistrust.

Of course, Windows will take some getting used to for network users and managers. With careful planning, testing, and the right information, you can make integrating Windows into your LAN an exciting and enjoyable experience. ■



---

# Installing Modems and Software

---

## In this report:

Modem Installation .....	2
Software Installation .....	4
Technical Details .....	4

## This report will help you to:

- Understand how modems and related software function.
  - Know the steps involved in planning for modem installation.
  - Select and install modems and software that provide effective communication.
- 

### Introduction

Setting yourself up to communicate using a PC isn't easy, but it's not hard, either. It's just that nothing seems to work properly the first time you try it. Your communications hardware and software are designed to work smoothly, and they will. It's only when something is slightly out of kilter that you might run into a snafu. As emphasized throughout this report, this is a common problem; you're dealing with something more complicated than a fax machine, after all. But take heart, and remember the power and usefulness of your new system.

Think about the car you drive. You probably are not intimately familiar with all aspects of the car's

operation, but you know the fundamentals well enough to operate it. It's possible to also take this approach when you are setting up your PC, modem, and software: You don't have to be an expert to get everything working and get on line. Nevertheless, all the technical details are included in this report and if you encounter a problem, this information can help guide you to a solution.

To telecommunicate using a PC, you need specific equipment—both hardware and software. The hardware consists of a modem. The software is a program that will command the modem, and perform the various other functions of *terminal emulation*. These programs, called telecommunications or comm programs, range from the inexpensive to the very expensive. They're built into integrated programs such as Microsoft Works and even some of the more powerful utility packages such as PC Tools. They're available as powerful stand-alone commercial

---

This Datapro report is a reprint of Part 1, Chapter 3, "Installing Modems and Software" by John Dvorak and Nick Anis, pp. 35-46, from *Dvorak's Guide to Desktop Telecommunications*. Copyright © 1990 by McGraw-Hill, Inc. Reprinted with permission.

programs such as Crosstalk. Comm programs are among the very best shareware offerings, including such bestsellers as Procomm, Qmodem, and Telix. And many modems come with comm programs such as Bitcom or simple, proprietary "modem drivers." It's a buyer's market and if you can't find a comm program that's right for you, you're probably not looking very hard.

Let's start with some basics. For specific details, read the appropriate sections in your modem and communications software manuals.

---

### Asynchronous Communication Software

Many large computer systems consist of a CPU (central processing unit) with terminals attached. To use an application on the computer, a user communicates with it via a terminal. The computer in this case is referred to as a *host*.

A terminal is a dumb device consisting of video display, a keyboard, and the circuitry required to communicate with the host. It is capable of only two things: sending characters you type on its keyboard to the host and displaying on the screen, characters it receives from the host.

Terminals are still in use, but many have been replaced by PC systems running terminal emulation software. This system has many advantages. Among these are the ability to "capture" data received from the host, to exchange information, and to emulate more than one type of terminal without making a single hardware change. This arrangement also allows you to download information from the host and process it locally.

Many communication programs feature a programming language that can be used to automate a variety of tasks, such as logging in to the host. A couple of programs have more sophisticated languages that give you substantial power and versatility.

There are also specialized PC communications programs that allow you to take control of another PC at a remote location. The uses for this type of communication are many, and include remote application software support.

---

### Asynchronous Communication Hardware

Asynchronous communication hardware for PC systems is either built into the PC or added with an expansion card. The type of interface provided by this hardware is known as an RS-232 interface, commonly referred to as a port, a COM (communication) port, or a serial port. IBM originally called this a serial adapter, and sometimes it's called the serial card or serial interface card.

The original IBM PC standard defines two communication ports, COM1 and COM2. Most PC expansion cards can be configured as one of these two ports. PC systems that have a built-in serial port define it as COM1, and it usually cannot be changed. The new PS/2 systems define eight ports, COM1 through COM8. Most of the newer machines say they can address, or "talk to," these extra ports too. At this point, you don't need to be concerned with how many ports you have.

---

### Modem Installation

Telecommunication using a PC involves using an internal or external modem. The differences between the two styles will be covered later in this report, but first, let's take a look at some of the features found on both modem types.

A modem (whether internal or external) performs the function of converting information from your PC into a form that can be transmitted over a telephone line. It also converts the signals it "hears" on the phone line into information that is sent to your PC.

Most modems have two telephone jacks, one for the line that connects to the wall plug (usually marked "line" or "telco") and one that connects to a telephone. Your modem was probably shipped with a phone cord that connects it to the phone jack on the wall. If you have a spare phone, you can connect it to the other modem jack. This allows you to use the same line for both voice and data calls, although you can't do both at the same time. It is helpful, however, if you can't connect to a remote modem by direct dialing because you may first have to ask an operator to connect you to an extension.

Most modems are equipped with a speaker that lets you monitor the progress of a call. Volume level on some modems is set with a knob; others use a command.

External modems usually have status lights to indicate what the modem is doing. Internal modems do not have lights, but some software packages make up for this by simulating these lights with a clever screen display.

### Preparing for Installation

Before beginning a modem installation, it's a good idea to take an inventory of the expansion cards you have in your PC. For your purposes, you are really only concerned with the number of serial devices in the system, but it's handy to have a complete inventory for general use. You'll need to take the cover off the PC to do this; if you need instructions for this, refer to your PC user guide.

Start the inventory by writing down the types of cards you have in your system. If a card has a serial port on it, check its settings to see what COM port it's assigned to, and write that down too. (Refer to the expansion cards' documentation to determine their settings.) If you can't figure out the cards and their settings, take the computer to a dealer or friend who can run tests to determine what ports are used. The card inventory will be used in both modem and communication software installation.

*Caution:* If in the process of making the inventory you find that more than one device is configured for the same COM port number, you'll have to change a few switch settings. In a PC, each serial port must have a unique COM port number. If not, things won't work properly.

If you're using an internal modem, you'll need to know how many serial ports already exist in your PC, so that you can configure the modem. If you're installing an external modem, you'll need to know which serial ports (if any) you have available, and the COM number for which they're configured, because you'll be hooking the modem directly to the connector of one of these ports.

Your modem will likely have to be configured to work with your software. Some modems have physical switches you can set; others have "soft switches" that are set by the software in its configuration process. This isn't something to fret over—most Hayes-compatible modems usually work right out of the box, with the factory settings intact.

If you're installing an internal modem, look in its documentation to see if it has physical switches to set. The most important ones are the ones that define the COM port used, and the one

telling the modem if it is hooked to a multiline phone. The other settings are not as important, and the factory (default) setting will usually work. In any case, don't be afraid or annoyed if you have to take the modem out of the computer and change some switch settings more than once.

External modems also have either physical or soft switches. Since the modem is external, it can be configured at any time.

### Internal Modems

An internal modem is an expansion card that occupies a slot in the PC. Like any other adapter card, it gets its power from the PC. Internal modems require no cabling. All that is needed is the phone cord to connect the modem to the phone line.

Since many PC systems already have one serial port, and it's usually COM1, most internal modems are shipped from the factory configured as COM2. If you have only one serial port, and it's configured as COM1, then installing your modem card is just a matter of plugging it into an empty slot. If you have a COM2 port but no COM1, just change the modem's switches and/or jumpers so that it's configured for COM1.

If you find that you already have both a COM1 and COM2 in your system, you have the following options. You can disable one of the existing ports, and configure the modem to take its place, or you can check to see if the modem can be configured as COM3 or COM4. You'll also need to make sure that the communication software you plan to use supports this new setting.

Installation of the modem is explained in detail in the modem's manual, so refer to it for more information and to look up settings.

### External Modems

External modems are packaged in enclosures, ranging from very nice aluminum cases to inexpensive plastic ones. Use of an external modem requires a serial port in the PC, cabling from the port to the modem, and a power supply.

The power adapter provided with most external modems is a small cube that plugs into the wall, and has a wire that plugs into the modem. Power adapters of this type can sometimes be a bit bulky. Fortunately, a few modem manufacturers now use an adapter that sits on the floor and has two wires—one that extends to the wall plug and one that goes to the modem.

## Cabling

A cable is a bundle of wires which, in this case, is used to carry electrical signals from the PC to the modem and vice versa. Cables are manufactured for just about every PC need.

The type of connector used for a PC serial port and most modems is known as a D-type connector (it vaguely looks something like the letter D). There are two types of D connectors: male (with pins) and female (with sockets). D-type connectors are labeled so that each pin or socket has a number from 1 to 9, or from 1 to 25. A serial connector on a PC is usually male. Some have 9 pins, others have 25. Modem connectors are usually female and have 25 pins. The 9-pin connector is also known as a DB-9; the 25-pin is called a DB-25.

Computer stores can supply you with a cable manufactured specifically for connecting a PC serial port to a modem. There are two types of cables: ribbon and shielded. Shielded cables are preferred because they prevent radio frequency interference with other equipment such as televisions, radios, and VCRs.

A properly manufactured off-the-shelf cable should work with your setup. If it doesn't, check it against the connection information provided in "Technical Details" later in this report. Fifty percent of the problems people have with modems have to do with the cables. Some printer cables, for example, have the pin assignments switched around to work only with certain printers. Since the printer cable looks like a modem cable, it is often used as one. The result, of course, is that your modem won't work.

## Connection to the PC

Now that you have determined which COM port you're going to use, you have the correct cable, and you have your external modem, it's time to connect everything together. This is a simple matter of plugging one end of the cable into the PC, and the other end into the modem. If your cable has screws, tighten them just enough to make the connectors fit snugly.

---

## Software Installation

There are many communication programs available, each with unique commands and operating characteristics. For this reason, this report includes only general directions for software setup. Some of

these products will be easier to configure than others. Fortunately, most packages have a program for installation or configuration, or both.

The first thing to determine before you install the program is the port number of the internal modem or serial port you will be using. (You should already know this if you've read the "Preparing for Installation" section of this report.) Some programs can concurrently use more than one device. If your program has this ability, and you plan to use it that way, you will also need to know the port number of the additional devices.

When you install communication software on your PC, the program must be configured for the type of modem you have and the COM port to which the modem is attached. Some software packages accomplish this as part of the installation program; others do it as part of a separate configuration step. The latter is preferred, since this method allows you to define a different setup should you change the modem or configuration of your PC.

Most communication packages have a method of storing a configuration for each service you call (CompuServe, a local bulletin board system, and so on). It's a good idea to set up a configuration for any service you plan to use. You'll then be able to test your setup by using one of these configurations to place a call.

---

## Technical Details

As mentioned previously, you probably won't need the information in this section. It is useful, however, if your system setup is a little out of the ordinary, or if you simply wish to know some of the technical details.

### COM Port Addressing

Think of the communication (COM) port definition as the *identity* of a port. This identity is the combination of an address and an IRQ (interrupt request number). Each port in a PC must have a unique identity; that is, no more than one port can have the same address and IRQ combination.

The *address* is what tells the software where the port is. When the software needs to communicate through COM1, for example, it knows the address for that port, and sends the information there. It's really very much like the street address

of your house; when someone needs to send information to you, they use your address.

Various parts of a PC system use signals known as *interrupts* (or IRQs) when they have something to say. For example, a port generates an interrupt when it has received a character. This tells the software that it needs to address the port to get the character and process it (put it on the screen, for example).

The original PC systems defined only two ports, COM1 and COM2. As PC users needed more serial ports, hardware and software manufacturers responded by creating their own definition for two more ports, COM3 and COM4. Since there is no official standard for these ports, however, not all software and hardware products support COM3 and COM4. The current PS/2 systems define eight ports, COM1 through COM8.

Most programs refer to the communication ports in your PC as COM ports. You do not usually need to know the actual address and IRQ of a port to make things work, but it can be handy to know the standard settings. The following chart lists the standard addresses used for PC and PS/2 serial ports.

Port	PC Address/IRQ	PS/2 Address/IRQ
COM1	03F8h/4	03F8h/4
COM2	02F8h/3	02F8h/3
COM3	03E8h/4*	3220h/3
COM4	02E8h/3*	3228h/3
COM5	n/a	4220h/3
COM6	n/a	4228h/3
COM7	n/a	5220h/3
COM8	n/a	5228h/3

\*Remember, although they are the most common, not all hardware and software products support these address/IRQ combinations for COM3 and COM4.

**The UART**

The UART (universal asynchronous receiver transmitter) is a specialized IC (integrated circuit) that handles asynchronous communication. It is responsible for sending and receiving data, and handshaking with the attached device. Older PC systems use the 8250 UART. AT-class and better machines (those with 80286 and 80386 processors) use the 16450 or 16550 UARTs. The older 8250 design has had problems at high speed. The newer 16450 and 16550 devices offer better performance, and fewer problems.

There are several “support” chips for the UART, but you need not usually be concerned with these. It can be useful, however, to know which UART is in your system. If you don’t know what you have, look at the expansion card for a large, 40-pin IC labeled with one of the numbers listed above. If you can’t find the chip or a number that matches, look it up in the documentation or contact the hardware manufacturer.

The interaction between your communication software and the UART is largely transparent. The software tells the UART to send and receive characters, and to start and stop the flow of data (flow control). Some programs actually detect the type of UART with which they are working and make appropriate adjustments.

**Serial Port Cabling**

Although serial connectors have as many as 25 pins, only 9 of them are actually used to carry signals. These 9 signals are listed in the following table.

Mnemonic	Name	Source	Purpose
TXD	Transmitted Data	PC	Carries characters from the PC to the modem.
RXD	Received Data	Modem	Carries characters from the modem to the PC.
RTS	Request To Send	PC	Used for flow control.
CTS	Clear To Send	Modem	Used for flow control.
DSR	Data Set Ready	Modem	Used for flow control.
SG	Signal Ground	N/A	An electrical reference point for the other signals.
CD	Carrier Detect	Modem	The modem makes this signal active when it has connected to another modem. Some software programs monitor this signal to know when they’re on line.

Mnemonic	Name	Source	Purpose	Signal	PC Pin	Modem Pin
DTR	Data Terminal Ready	PC	Some programs make this signal active when they are about to place a call, or are in "local" mode.	TXD	3	2
RI	Ring Indicator	Modem	Becomes active when a ring is detected. Can be important when answering calls.	RXD	2	3
				RTS	7	4
				CTS	8	5
				DSR	6	6
				SG	5	7
				CD	1	8
				DTR	4	20
				RI	9	22

There are two categories of serial devices: DTE (data terminal equipment) and DCE (data communications equipment). Terminals and PCs are DTE devices; modems are DCE devices. This gives you some idea of how the equipment interconnects. Generally, if you're connecting a DTE device to a DCE, and both have 25-pin connectors, a straight-through cable can be used.

The following sections show the connections for a 25- to 25-pin cable and for a 9- to 25-pin cable.

### 25-Pin PC Connector Signals

A cable of this type is said to be wired "straight through." This is because pin 2 on one end connects to pin 2 of the other, pin 3 goes to pin 3 and so on. Most cables used for the purpose of connecting a 25-pin PC port (DTE) to a modem (DCE) are wired this way.

Signal	PC Pin	Modem Pin
TXD	2	2
RXD	3	3
RTS	4	4
CTS	5	5
DSR	6	6
SG	7	7
CD	8	8
DTR	20	20
RI	22	22

### 9-Pin PC Connector Signals

The 9-pin AT-type serial port pinout is quite different from that of a standard DTE port's. For starters, there are only 9 pins. Also, the signals are on different pins from what you might expect. This table shows the correct wiring for a 9-pin AT-style port to a 25-pin modem.

### How Software Communicates with the Modem

Each modem has a set of commands. The most widely used one, the AT command set, was originated by the Hayes Microcomputer Company when they began manufacturing their Hayes Smartmodems. When a modem is said to be "Hayes compatible," that usually means it supports some part of the Hayes command set. Because of these command sets, it is important for your software to know the type of modem you're using, so that the software knows how to "talk" to your modem.

Most modems available today are programmable to an extent. Usually a modem will function just as it is, fresh from the factory, but there are many settings that you may wish to adjust for better performance in your particular environment. *Modem registers* are used to configure a modem's operation. For example, registers set the modem's dialing speed, and how long it waits to get an answer after dialing a number. Your modem's manual lists the modem registers and explains their use.

To set a register on the modem, you need to talk directly to the modem. This is usually done by telling the communication software that you want to "go local." Instructions for this procedure are available in the software's documentation. You can use the software's local mode to give commands to the modem.

The dialing process is usually somewhat transparent. When communication software commands the modem to dial, it sends three pieces of information (strings): a dialing prefix, the telephone number, and a dialing suffix. You usually don't see this data going to the modem, although the software will report that it is placing a call.

The dialing prefix usually contains the actual command to dial. In the case of a modem that uses



the AT command set, this is usually *ATDT*. The *AT* gets the modem's attention, the *D* is the Dial command, and the *T* tells the modem to dial using tones. If you need to pulse dial (that is, use a rotary dial instead of push button phone), you would use a *P* instead of a *T*. The telephone number is the number of the remote host's modem, and was entered by you when you set up the software to call the host. The dialing suffix, usually a carriage-return character, terminates the dialing command.

After the modem dials, it will wait to "hear" another modem answer the call. After the modems connect, the software makes you aware of this, and you're on line.

*Note:* Once on line, the modem can still respond to the AT commands when typed in as, say, a message. It is therefore a good idea to familiarize yourself with commands such as *+++ATH*, which hangs up the phone (terminates the connection).

You should now be on line and communicating when you command the modem to dial out to another number. Even if there is only garbage coming over the screen (meaning you haven't yet properly set the software parameters), you have at least made the all-important connection! You are now ready to do some serious work. ■



# Turning an Older PC into a 386 CPU-Based Machine

## In this report:

Motherboard Upgrades .....	2
Sources of Replacements.....	4
Accelerator Board Upgrades .....	7

## Datapro Summary

Upgrading to a 386-based PC does not necessitate abandoning your old system entirely. This report explains how to make your present IBM PC XT, AT, or compatible computer into a 386 machine. It outlines, in tutorial fashion, the installation of a motherboard and an accelerator card with 386 microprocessors. The report details the advantages and disadvantages of both motherboard and accelerator card upgrading.

## Pros and Cons of Upgrading

Installing either a new motherboard or an accelerator card has advantages over buying a complete 386 PC. The decision to upgrade or to purchase a complete system essentially comes down to a trade-off between cost and performance.

### Cost

The most notable advantage of upgrading is cost. Because you're not paying for a new case, keyboard, drives, ports, and the other essential parts of a computer, you save the expense of those components. The cheapest stripped-down, mail-order 386 SX microprocessor costs about \$1,000. To keep the price low, manufacturers often complete the basic configuration of 386 machines with

the least powerful components they can get away with. A recognized brand-name PC with a 386 microprocessor costs more. The street price (the cost you're likely to pay, as opposed to the list price) of a motherboard with a 386 microprocessor ranges from \$400 to \$1,500. The street price of an accelerator card with a 386 microprocessor ranges from \$400 to \$600. Your savings increase if you've already invested in high-end peripherals, such as a fast, large hard drive.

Although theoretically the cost of an entirely new PC can be offset by selling your old machine, anyone who has actually sold a used PC knows that you won't get much. Used PCs depreciate faster than cars. (That isn't surprising. Think about how little most cars are improved from one model year to the next, and then consider the real benefits incorporated into new generations of personal computers. A 10-year-old car generally will get you from one point

This Datapro report is a reprint of Chapter 4, "Turning an Older PC into a 386 CPU-Based Machine," pp. 61-79, from *Intel's Official Guide to 386 Computing* by Michael Edelhart. Copyright © 1991 by McGraw-Hill, Inc. Reprinted with permission.

to another as quickly and as efficiently as a new one. A 10-year-old PC is a museum piece.) If you have a use for an older PC—such as giving it to junior employees or your children—purchasing an entirely new 386 PC is a better option. On the other hand, if your old machine is likely to wind up in a closet, you should consider upgrading it instead.

By upgrading, you also get added convenience. Your hard drive stays in place. You don't need to back up its contents to floppies and then restore them to a new hard drive. You can keep the keyboard you've become accustomed to.

### Performance

There are, of course, drawbacks to either a motherboard or accelerator card upgrade. Installing either is perhaps more of a bother than backing up and restoring a hard drive. Installation, however, is a one-time affair. A more serious disadvantage of upgrading is that the performance of your retrofitted PC may not equal that of a new 386 machine. Of the two most significant performance issues, one is more applicable to upgrading an XT-class PC with an accelerator board, a topic that will be discussed in detail later in this report. The other performance issue applies to either type of upgrade: how well are your old components matched to your new processor?

A 386 PC without a hard drive, video, and other components that can match the speed of the 386 chip doesn't live up to the processor's full potential. A hard drive that was a speedy match for a 286 processor can be a drag on a 33MHz 386 chip. Does this mean that upgrading an older PC is a bad idea? No, not if upgrading is the only affordable option. If your old PC is a 286 machine and its hard disk is an ESDI drive with an access time in at least the 30-millisecond range, upgrading is sensible.

Even if your old machine is not a perfect candidate for upgrading, it's still an option you should consider. You can always upgrade the other components as your budget allows, beginning with a faster drive. In the meantime, while upgrading an older machine may not produce the overall speed of a new 386 PC, speed is not the only advantage

of a 386 processor. The upgrade will give you access to the 386 chip's memory management, important with a program such as Windows 3.0, and will let you run programs optimized for 386 microprocessor instructions.

If you decide that upgrading is the route you should take, you must then decide between the two upgrade options—a new motherboard or an accelerator card.

---

## Motherboard Upgrades

A PC's motherboard is the most crucial part of a computer. The motherboard is the site of the component that most defines the character of a computer, its microprocessor. The motherboard also holds the ROM BIOS chip, which defines a PC clone's compatibility with the IBM family of personal computers. All signals among the other components of the PC are routed through the motherboard, which also provides slots for expansion boards. Some motherboards also provide many of the functions that ordinarily would be on expansion cards, such as parallel or serial ports. If you remove the motherboard, what you have left is a collection of inert, unusable computer components. Install a motherboard that contains an Intel 386 processor and you have a new computer.

### Advantages of a New Motherboard

A new motherboard has two advantages over upgrading via an accelerator card. One is that you automatically upgrade the data path if you're installing the motherboard in an XT-class machine. The second is that you get a new ROM BIOS chip that maintains your PC's compatibility with new machines.

### The Data Path

The original PC and the XT have an 8-bit bus. That means those computers can move data among peripherals and through the motherboard only 8 bits at a time. The 386 DX microprocessor is a 32-bit chip that internally handles data 32 bits at a time and is capable of transferring data to and from the motherboard 32 bits at a time—if the motherboard can accept data with that wide a path. Only MCA and EISA buses have 32-bit data paths to peripherals, although some ISA motherboards provide 32-bit paths to RAM.

If the PC you're upgrading is an XT-class computer, a replacement motherboard is the better way to move the machine to 386 microprocessor status. The XT's 8-bit path puts a serious damper on the 386 chip's performance. If, however, you are upgrading an AT-class machine with its 16-bit path, there is usually no performance gain in data transfer rates involving the replacement motherboard's bus, which is likely to be the same ISA design. There were no replacement motherboards with 32-bit data paths at the time of this writing. That situation may change by the time you read this. If you can find a 32-bit EISA or MCA replacement motherboard, it is an option you should consider to obtain the best performance and the longest usability from your PC.

At first, swapping in a motherboard with an ISA bus seems like a serious limitation on the 386 microprocessor. It is, in theory. But in pragmatic terms, an ISA replacement motherboard is no worse an option than you get with many off-the-shelf 386 PCs. Many 386 computers have the older, 16-bit ISA bus.

There are several reasons why manufacturers use a bus narrower than the 386 chip's data path. ISA buses are less expensive to produce. While the number of expansion cards that can take advantage of a 32-bit bus is growing, 16-bit cards are more plentiful and less expensive. For a PC powered by a 386 SX microprocessor, the 16-bit bus is a perfect match because the SX chip itself, while it handles data internally 32 bits at a time, only has a 16-bit data path to the outside world.

The new motherboard, however, should have a 32-bit data path to RAM. Communication with some peripherals, such as a serial port, printer, or drive, is slow because those devices are slow; a wider data path gives you no advantage in communicating with them. You should use the fastest RAM your motherboard will support, and the processor should be able to access memory in 32-bit chunks.

What it comes down to is that you shouldn't be concerned if a replacement motherboard only has a 16-bit bus as long as it has a 32-bit data path to RAM. Presumably, your budget is one reason you'd consider a new motherboard in the first place. If that's the case, while you might be able to stretch the budget to cover the cost of an off-the-shelf ISA machine powered by a 386 processor, you probably can't afford the more expensive

MCA or EISA machines. Conversely, if you can afford one, that's what you should get. When you can afford it, a 32-bit bus is a better choice for a 386 microprocessor configuration in the long run.

### ROM BIOS

The second advantage of getting a new motherboard is that you also get a new ROM BIOS at the same time, which isn't true if you upgrade with an accelerator card. This is to ensure that there are no incompatibilities between your new processor and your old BIOS.

If your machine is a relatively new AT-class PC, the BIOS may not be an issue. But with older, XT-class computers, a new BIOS is essential if you want to have a 386 PC with all the functionality to which you are entitled. An older ROM BIOS may not support the expanded 12-function-key keyboards or 3½-inch floppy drives. It is possible to buy BIOS upgrade chips for major brands of PCs, but if you have a no-name PC clone or a discontinued model of a known compatible, there's no guaranteed source of BIOS upgrades.

### Disadvantages of a New Motherboard

In addition to the consideration that your hard drive and other peripherals may be a bottleneck to a 386 chip's performance, the primary disadvantage of a new motherboard is that selecting and installing it can be a daunting experience. If you are replacing a board in an IBM computer or another popular brand, you're likely to find one that fits inside your old case easily. With less common PCs, particularly "baby" ATs (machines with little room for expansion cards) and others with a small footprint (designed so that they don't take up much desk space), finding a replacement that fits correctly can be a problem.

Even if you have the right motherboard, installing it is not as simple as slipping in a new expansion board. You'll have to remove most of the components inside the case because motherboards are invariably mounted on the bottom, below everything else.

In some instances you'll also need to replace the power supply. Many older PCs are equipped with 65- to 85-watt power supplies. The new motherboard may require a 120-watt power supply. While replacing the power supply is neither difficult nor expensive (about \$70), it's a factor in your upgrade decision.

If you have the original keyboard for an 8088-based computer, you'll still be able to use it with the new motherboard. If you want the functionality of a 12-function-key keyboard, add another \$60 to \$70 to the cost of your upgrade.

### Buying a Motherboard with a 386 Microprocessor

Motherboards are not generally stocked by PC retailers. The demand for new motherboards is not great, and retail shops would prefer to sell you an entirely new PC. You may be able to find a motherboard locally at a shop that specializes in electronic components. Computer "swap meets"—flea markets where individuals sell and swap hardware—are another possible source of an inexpensive motherboard. Unless you regularly attend the same swap meet and know the person from whom you're buying the motherboard, however, don't try this. If the motherboard doesn't fit or doesn't work, you may not be able to find the person you bought it from to return it.

If you can't find a local source of motherboards, your only choice is to purchase a motherboard by mail order. The classified ads in a magazine such as *Computer Shopper* or *PC Sources* are good places to start looking. In addition, Table 1 lists the names of several companies who manufacture motherboards; you can contact them either to purchase a board directly or to find out the name of a local store that sells their products. Although mail order has a bad reputation, most mail-order, or "direct-purchase" firms, as they like to refer to themselves, are respectable. Follow these guidelines, and you can be confident about a mail-order purchase.

- Ask the firm if it charges a restocking fee. This is a fee—from 5% to 15% of the purchase price—that the company retains if you return your purchase. The possibility of a motherboard you buy sight unseen not fitting your case is too great to risk a restocking fee.
- Use a credit card to make your purchase. This gives you certain legal rights, and the credit card company gives you some clout that you wouldn't have if you used a check or money order.
- Ask the mail-order firm when it will ship the product. A reputable firm will ship within 48 hours of the purchase, which, if you're using a

---

## Table 1. Sources of Replacement Motherboards and Accelerator Cards

---

### Motherboards

Hauppauge Computer Works, Inc.  
91 Cabot Court  
Hauppauge, NY 11788  
(516) 431-1600  
Fax: (516) 433-3198

Orchid Technology, Inc.  
45365 Northport Loop West  
Fremont, CA 94538  
(800) 767-2443; (415) 683-0300  
Fax: (415) 490-9312

### Accelerator Boards

AOX, Inc.  
486 Totten Pond Road  
Waltham, MA 02154  
(617) 890-4402  
Fax: (617) 890-8445

Applied Reasoning Corp.  
86 Sherman Street  
Cambridge, MA 02140  
(617) 490-0700  
Fax: (617) 492-7908

Cumulus Corp.  
23500 Mercantile Road  
Cleveland, OH 44122  
(216) 464-2211  
Fax: (216) 464-2483

Intel Corp.  
3065 Bowers Avenue  
Santa Clara, CA 95051  
(408) 765-8080  
Fax: (408) 765-1821

MicroWay, Inc.  
P.O. Box 79  
Kingston, MA 02364  
(508) 746-7341  
Fax: (508) 746-4678

Sota Technology, Inc.  
559 Weddell Drive  
Sunnyvale, CA 94089  
(800) 237-1713; (408) 745-1111  
Fax: (408) 745-1640

---

credit card, should mean 48 hours after you get off the phone. Less reputable firms use your money to buy their own stock before they ship, or they simply sit on your money to draw interest from it before they deliver the product.

- Check with the Better Business Bureau in the city where the mail-order firm is located. If the

Bureau has received complaints and if the company has not responded to the complaints satisfactorily, don't do business with the company.

Regardless of the type of firm that you do business with—local or mail-order—ask for a guarantee that the board will fit inside your case. Motherboard replacements are generally sold in sizes for the PC, XT, and AT. Some others are classified as baby AT size, and still others are designed to fit inside a *tower* chassis, which is made to stand vertically on the floor.

Each type differs in size and in the location of bolt holes. If you have an obscure brand, open the case and measure the current motherboard. Look for the bolts that attach the motherboard to the case and measure their positions. Although it isn't necessary for all the bolt holes in the new motherboard to line up with all the bolt positions in the case, at least half of them should align. If that's not the case, you'll either have to look for a better-fitting motherboard or drill new holes in the case. Some bolts use *standoffs*, small plastic or nylon tubes through which the bolts fit; standoffs hold the motherboard away from the case. Measure the clearance between the board and the case, and check that your new board will fit properly with either your old standoffs or with new ones supplied with it. If the motherboard must fit under drives or other components, measure to make sure you have enough clearance there, too.

Replacement motherboards are available with 386 DX and 386 SX microprocessors, and chip speeds range from 16MHz to 33MHz. Shop for the best match of price and speed. Make sure that the board provides a 32-bit data path to memory, and try to buy a board that does not use proprietary memory chips, which are more expensive than generic RAM.

### Installing a New Motherboard

The only tool you must have to install a new motherboard is a screwdriver or socket driver, although you may find needlenose pliers handy for such chores as positioning standoffs and holding nuts. Some PCs, such as those from Compaq, use a special tool called a *star driver*, which you can buy at hardware or electronics supply stores. The installation process should take less than an hour.

### Getting Ready

Before you begin, prepare an open area in which to place the components that you remove from your PC. Ground yourself to prevent any static discharges that could harm chips, either by touching the computer's case as it is still plugged into a grounded outlet or by purchasing a *grounding strap*, a strap you wear around your wrist and attach to some object you know is grounded. If you have the antistatic bags that your expansion cards came in, lay them out in your work area and use them to hold the boards as you remove them. If you don't have the bags, place sheets of aluminum foil on the work surface and lay your boards on them.

If you are upgrading an AT-class PC, run your PC and go through its setup routine before opening its case. (Usually you reach the setup screen by pressing a certain key combination, but some computers run the setup program from a disk file; consult your manual.) When the setup screen appears, either make a note of the drive information or, better yet, hit the PRTSC key to make a hard copy of the information. You'll need this information, particularly the type of hard drives you have, when you set up your new motherboard. An XT-class computer has no setup routine. You'll need to make a note of any information attached to the hard drive after you have opened the case. This may include a drive type, the number of cylinders, and the drive's capacity.

Now turn off the PC and all other peripherals attached to it, such as the printer and monitor. Unplug the power cord, and disconnect the keyboard, monitor, printer, phone, or any other cables leading to it.

Open up the case by removing the screws mounted along the edge in the back; there are usually five. If you have a flip-open case, open it by pressing the two release buttons on either side.

As you disassemble the computer, make a sketch showing the location of each expansion board. (A few boards are particular about which slot they are installed in.) Also sketch which wires are attached to which components. Note the colors of wires and the shapes of connectors. It may all look logical when everything's attached, but once you have a nest of similar-looking wiring, it's hard to remember what goes where. As an added precaution, as you remove each connector, put a small

piece of masking tape on the connector on the cable and on the connector on the motherboard that it was attached to. Label each set with something descriptive, such as "power supply," "reset switch," and so on.

### Disconnecting Components

Disconnect the following cables:

- The power supply connector from the motherboard
- The speaker cable from pins on the motherboard
- The cables leading from the floppy and hard drive controllers to the drives. (You only need to remove them from the ends that attach to the expansion board.)
- Any other cables that lead from the motherboard to other components, such as the keyboard lock switch, reset button, turbo mode switch, and LED

Now remove all expansion boards. Unscrew the bolt that attaches them at the rear of the case, and then gently rock the boards back and forth until they come out. Don't rock or bend them sideways.

### Removing the Power Supply

If you are replacing the power supply with a more powerful one (or if the power supply gets in the way of removing the motherboard easily), you'll have to take it out. Disconnect the power cables leading to the drives, labeling them as you do. Unscrew the bolts that anchor the power supply to the back or bottom of the case. Some power supplies are also attached by two metal tongues along the bottom of the case. Slide the supply off the tongues to free it.

### Removing Drives

In some configurations, you may be able to slide the old motherboard out from beneath your drives without removing them. In other cases you'll have to remove at least one drive to get to the motherboard. If the drives are grounded to the chassis with a small wire, slip the wire off the grounding connection. The drives are attached to the drive bays by bolts that enter from the sides. Remove the bolts and slide the drives forward. Some under-sized drives are attached to shims within the drive

bay. You can generally leave the shims attached when you pull out the drives.

### Removing the Old Motherboard

Motherboards are generally attached to the case in one of two ways. Some are attached to standoffs through which bolts are attached. Others have plastic standoffs that snap onto the motherboard and slide onto runners on the inside bottom of the case.

In the first instance, there could be up to nine or ten bolts holding the motherboard in place. Remove them with a screw driver or socket driver. Then lift the motherboard slightly and slide it out from beneath the drives far enough to raise it free of the case.

In the second circumstance, loosen the screws that hold the motherboard, slide it toward the back, and then lift it to remove it from the chassis.

### Installing the New Motherboard

Before installing your new motherboard, make sure that it is completely configured. Install any memory chips or BIOS chips and the math coprocessor chip, if you have one. After installing any chips, check against illustrations that should accompany the board to confirm that they are installed in the right order and facing the right direction. Make sure none of the chips' pins are bent under the chip or sticking outside the socket. Finally, give each chip a push to make sure it's seated firmly.

Set any jumpers on the motherboard. *Jumpers* are small connectors that fit over a pair of metal prongs sticking up from the motherboard. Which prongs should be jumpered depends on how much memory you've installed and other factors that will be detailed in the board's instructions. Make sure that any onboard battery is installed and that it's the proper type.

If the board you're replacing has standoffs, you may need to remove them from the old board and attach them to the new one. From the underside of the case, screw down the standoffs that align with the holes in the case. If not all of them align, don't worry about it; if half of them do, your new motherboard will be secure enough.

If you removed the power supply or drives, reinstall them. Reattach all cables to the supply and the drives; then reattach all the cables that go to the new motherboard from the power supply, reset switch, and so on. Your new board should



have a diagram that identifies the board's connectors for the power cable and other wiring. If the diagram is confusing, consult the masking tape labels you put on your old board. The connectors on the old and new boards will look similar and should be in approximately the same locations. Connectors will generally fit only one way, so don't worry about connecting something backwards.

Finally, reinstall the expansion boards. Most 8-bit boards can be installed in any slot. (The exceptions are boards, such as some mouse boards, that are temperamental about being installed in certain slots. Don't install a mouse board in the farthest slot from the power supply.) Consult your manuals if any of the boards don't seem to work once you start up the computer. Sixteen-bit boards can only go into longer, 16-bit slots or 32-bit slots if you have an MCA or EISA bus. Bolt the boards to the back of the case and reattach any cables.

Replace the case top and plug in the cables for the power supply, monitor, keyboard, and any other peripherals.

### Checking the New Motherboard

When you start up your "new" computer, you'll have to set it up to tell the CMOS memory chip the time and date, what kinds of drives you have, how much memory is available, and the type of display. You may also have the option to use high memory as shadow ROM for faster input/output operations. (If you do, choose that option.) On some 386 microprocessor boards, you must boot from a special setup floppy. On most, however, the setup routine is included in the BIOS chip. When you turn on the PC, the bootup routine will look for a proper setup, and if it doesn't check out, the BIOS will automatically display the setup screen. Using the information you obtained from the setup screen for your old motherboard or the information you copied from labels on the drives, answer the questions the setup screen asks you.

If your new motherboard or any of the components don't work, go through the steps again to make sure everything was done properly. Particularly check for bent chip pins, improperly cabled connections, or improper jumper settings. If that turns up nothing, reinstall your old motherboard. If a component that didn't work with the new one doesn't work with the old motherboard, then the problem lies with it. (It may have been jolted by static during the procedure.) If everything works

with the old board, reinstall the new one, paying particular attention to the setup screen. When all else fails, call the manufacturer.

---

## Accelerator Board Upgrades

The alternative way to upgrade your present PC into a 386 machine without swapping the motherboard is to install an accelerator board, sometimes called a *turbo card*.

### Types of Accelerator Boards

There are three kinds of accelerator boards. One, typified by the PC Elevator 386, plugs into an expansion slot like any other add-in card. Its circuitry replaces many of the operations normally handled by the motherboard. It includes its own BIOS and operates essentially as a coprocessor, working in tandem with the processor already in your computer. The new processor handles code instructions and calculations and leaves the old processor to handle the routine chores of communication among the peripherals.

A second type of card also plugs into an expansion slot. But unlike a coprocessor card, the second type uses a cable that runs from the accelerator card to a plug that fits into the socket that otherwise would hold your old 8088 or 80286 microprocessor. Your old processor is sometimes then installed on the new accelerator card; with some boards you can discard the old processor entirely.

A third, less common form of accelerator plugs directly into the old microprocessor slot and doesn't use an expansion slot at all. Instead, it fits entirely on a circuit board slightly bigger than a credit card.

Each type of accelerator has its advantages. The coprocessor card can more easily be maximized for optimum performance because it includes most of the operations of a self-contained computer. However, because the combined operations of old and new processors are more complex, the design is more difficult to use successfully. Theoretically, you can install several coprocessor boards and achieve amazing performance through parallel processing, but there is no software written for parallel processing—you would have to write your own.

The type of board that plugs into your old processor's slot, such as the Sota 386si, tends to

work better with your peripherals because it communicates directly with them. It may sacrifice some performance if it uses your old memory, which is not designed to operate at the speed of a 386 chip. Some of these boards have a connector to which you can connect a *daughterboard* that contains 32-bit RAM.

Boards that forgo a slot to plug directly into the old processor's socket, such as the Cumulus 386 SX, also work more smoothly with your other components, and they can work with either ISA or microchannel buses. However, these boards rarely have any provision for adding essential 32-bit memory.

An accelerator card of any type is the easiest and cheapest way to upgrade a PC to a 386 machine, but for technical reasons is it not the most ideal.

#### **Advantages of an Accelerator Card**

The chief advantages of an accelerator card are that, like a motherboard with a 386 microprocessor, it gives you five to eight times the speed of a standard PC and access to the memory management features of the 386 chip, with all the abilities that can be built into software using that memory management. The primary advantage an accelerator card has over a motherboard upgrade is that it is less expensive—\$200 to \$1,000 less than a motherboard, depending on the type of 386 chip and the manufacturer. It is also easier to install an accelerator card than a new motherboard. The only component you must remove to install a card is your old processor.

As with a motherboard upgrade, you don't have to back up your files from an old hard drive and restore them to a new drive in a totally new machine. Unlike a motherboard upgrade, an accelerator card may not require you to run through a setup routine. The CMOS setup is a function of the BIOS chip, which doesn't change when you install some types of accelerator boards. You don't have to worry about whether the accelerator card will fit inside your case as you do with a new motherboard. It is the standard size of any other expansion board.

#### **Disadvantages of an Accelerator Card**

The primary disadvantage of an accelerator card is that, while you will definitely see a performance improvement, in many cases an accelerator board

will not provide as fast throughput as a comparable motherboard with a 386 microprocessor. Depending on the machine in which you're installing it, you may be left with your old, narrower data path. This is particularly true in the case of XT-class machines. While data may be manipulated inside the 386 chip in 32-bit chunks, the processor can communicate with RAM, the drives, video, and other peripherals only 8 bits at a time. Many accelerator boards for XT-class or AT-class computers provide some sort of RAM caching to overcome the limitations of narrow data paths, but RAM caches do not improve the performance of all operations. The performance of a PC equipped with a 386 microprocessor that still has an 8-bit bus is so far below optimum that you shouldn't consider this combination unless it is the only tactic your budget will permit.

If you are adding an accelerator card to an AT-class computer, throughput with peripherals is not worse than it is with the many replacement 386-microprocessor motherboards that also have the 16-bit AT bus. However, an accelerator card upgrade to an AT-class computer doesn't always give a 32-bit path to RAM. Most new motherboards, on the other hand, include their own memory connected by a 32-bit path even if they continue to communicate with peripherals over a 16-bit bus. It is possible for some accelerator cards to provide 32-bit memory with daughtercards, but the additional price of the memory boards eats into the savings of getting an accelerator card.

The combination of a new processor board and your old BIOS can lead to compatibility problems. When accelerator cards were first released, many of them could not run OS/2 or Windows/386. Although those particular incompatibilities have since been largely resolved, you cannot install the advanced software written specifically for the 386 chip and confidently expect it to run. An accelerator card in an XT-class machine may produce hardware incompatibilities, too. Intel's AboveBoard memory expansion board, for example, can be configured as either expanded memory or extended memory. In an 8-bit bus, however, the memory board cannot work as extended memory. Even though you have a processor on your accelerator card that can handle extended memory, the AboveBoard sees the 8-bit bus and assumes that

you have an XT-class PC that won't handle extended RAM. This can cut you off from the large number of programs that use extended memory, such as Windows 3.0.

If you have a PC powered by an Intel 8086 processor, such as the Compaq Deskpro, you have less choice in accelerator boards. Although Sota Technology's accelerator card for the 386 microprocessor can work in either an 8088 or 8086 computer, most accelerator cards are designed with the more common 8088 machines in mind.

If you are upgrading an older PC with a power supply of about 63 watts, the manufacturers of some accelerator cards recommend that you upgrade your power supply to at least 100 watts.

Finally, with any of the accelerator cards that plug into an expansion slot, you give up that slot for any other use. With a motherboard, you usually wind up with the same number of slots. With some motherboard replacements, you may even wind up with *more* either because they come with more slots than your old motherboard or because they incorporate some functions usually handled by an expansion board, such as serial or parallel communications.

### Buying an Accelerator Card

Because accelerator cards are less expensive and easier to install than a motherboard, you have a wider choice and you're more likely to find them at your local computer retailer. You'll find more models to upgrade an AT-class computer than those that work with XTs. The disparity makes sense; there's a better market for AT upgrades because they don't burden the 386 chip with an 8-bit bus. Many owners of 8088 PCs sensibly prefer to buy a complete 386 machine that gives them the wider data path rather than upgrade a machine that can never keep pace with its new processor.

Whatever machine you plan to upgrade, make sure the board you get is designed to work with it, particularly if your present computer has an 8086 processor. Even if you don't plan to buy a memory daughterboard when you make the upgrade, it's better to purchase an accelerator card for which you can buy a memory option later. Without 32-bit memory, your 386 PC isn't living up to its potential and you should leave open a way to get that memory later. Most accelerator boards provide a socket to install an Intel 387 math coprocessor, but ask to make sure. If you already have an 8087 or

80287 math chip with your old board, it will not work with your accelerator card. You may not have a need for a math coprocessor now, but as with 32-bit memory, it's best to have the option later when you want to push your 386 machine to the limit.

If you buy your accelerator board from a mail-order outlet, follow the same warnings given earlier in this report for buying a motherboard from a mail-order firm.

### Installing an Accelerator Card

Even those who may find replacing a motherboard too forbidding will have a relatively easy time installing an accelerator card. Before you begin the installation, take the same precautions that you would when installing a new motherboard. Make a record of your CMOS setup, unplug your computer, and take care to ground yourself before you open the case. If the card you're installing is designed to work with a combination of 8088, 8086, or 80287 PCs, you'll have to set either jumpers or dip switches to tell it what kind of computer you're putting it in.

Once you have the case open, you won't have to remove the power supply, disk drives, or cables as you must to remove an old motherboard. If you are installing a coprocessor accelerator card, all you have to do is to put the card into any open expansion slot.

If, however, you are installing the type of accelerator that connects to the socket for your old processor, you may have to remove one or two expansion cards so that you can reach your old processor. Use a \$1 tool called a *chip extractor* to remove the old processor. The extractor looks like a large pair of tweezers with tips bent inward. Grip the old chip lengthwise, hooking the bent tips under the ends of the chip. Pull and rock the chip until it's free.

If the accelerator board does not use an expansion slot, simply plug the self-contained accelerator into the old processor's socket; the hardware installation is done. If the accelerator plugs into both an expansion slot and the old processor's socket, choose an expansion slot located near the old processor and slip the accelerator card into it. A ribbon cable about six inches long will run from the board to a connector that has pins like the pins on your old processor. Plug that connection into the processor socket. The ribbon cable is typically stiff, and the hardest part of the installation will be

making it reach the processor socket without getting in the way of other expansion boards. In some instances, it may be necessary to run the cable under other boards.

In addition to installing the accelerator hardware, you may have to put a line in your system's

CONFIG.SYS file so that your computer recognizes the new board. The CONFIG.SYS line may also need parameters to tell the accelerator board how to handle RAM or disk caching. If you are upgrading an AT-class computer, you'll have to run the setup routine to update the CMOS. ■

---

# PC Maintenance Guidelines

---

## In this report:

Maintaining a Proper Environment.....	2
"But I Did Everything I Could . . ."	3
What Are My Service Alternatives?.....	4
What Do I Look for in a Service Company?.....	6

## Datapro Summary

Personal computer users expect their machines to work forever. However, when the systems break down for some reason, we panic. This report is a practical guide for maintaining a proper PC computer environment, with emphasis on preventive maintenance. It also explains industry-wide service alternatives, such as carry-in and on-site service. The report also identifies several key sources of PC maintenance, including self-maintenance, OEM, and reseller/VAR services.

---

Has this ever happened to you? That computer that you could never understand the need for has become an essential element in your life. And whether you're using the PC for financial analysis, personal organization, recipe filing or games, a "crash" can reduce the most stalwart user to the point of tears. As personal computers have become an indispensable (albeit not always beloved) part of our lives, the potential impact of computer failure has taken on epic proportions.

While microcomputers can and do break down, frequency of failure tends to be less a function of brand than of particular application or usage. In general, PCs are extremely

reliable, with system failure occurring in roughly one-half of one percent of all computers. It is the components with moving parts which experience significantly higher incidence of breakdown; in order of frequency of breakdown, computer equipment can be ranked (from least to greatest frequency of failure): microcomputer, keyboard, disk and tape drives, and printers.

While it is impossible to prevent a computer-down situation—with its attendant anxiety and lost productivity—there are a number of steps you can take to protect your system and resolve a failure with a minimum of pain and suffering. Preventive maintenance (PM) is critical to enhanced—and extended—computer life. The most important elements of computer PM are the most obvious, and therefore the most overlooked or neglected:

---

This Datapro report is a reprint of "Taking Care of Your Computer," by Jim Rosen, pp. 7-14, from *PC Today*, Issue 3, Volume 5, March 1991. Copyright © 1991 by Peed Corporation. Reprinted with permission.

1. Read the operator's manuals for all components of your computer system to ensure that you fully understand the system's operation; and
2. Familiarize yourself with the manufacturer's recommended suggestions for care of the equipment and comply with those directions.

### Maintaining a Proper Environment

Like the sensitive electronic equipment it is, your computer system will operate longer in a conducive environment. If you pile books on top of your monitor, spill the morning's coffee into the keyboard or plug your system into the same outlet as your refrigerator, you have flunked the "conductive" criteria.

Your PC requires a clean, dry environment with a proper power supply. Following are some rules of thumb:

#### Your Work Area

Keep the area around your computer system free of excessive dust, dirt or moisture. Any materials such as books, magazines and printouts should not be placed where they might obstruct air flow. Similarly, equipment should never be enclosed, and the ventilating system must always face an open area where warm air can rise freely. Once you decide on a location for a computer, do not keep moving it around; frequent movement can cause a number of problems, particularly with the read/write heads of your floppy and hard disk drives.

Do not eat, drink or smoke around your computer. These activities will invariably put Murphy's Law into play: Crumbs, sauces and coffee have a tendency to find their way into sensitive components. If you spill a beverage into your computer or keyboard, take it to a reputable service center immediately for professional cleaning; don't try to do it yourself. Tobacco smoke will also cause problems for your system. If you must smoke, exhale away from the computer. The tar and dust particles in smoke can harm delicate components, and a hot ash can melt plastic.

#### Static Electricity

The components of your computer are extremely sensitive to static electricity, and static build-up can result in anything from monitor failure to data loss. Static electricity is particularly common in

carpeted areas, or in environments with dry heating systems, such as forced hot air. Build-up can easily be minimized with properly grounded static mats and/or regular application of anti-static carpet spray as required. In addition, a humidifier should be used to maintain adequate humidity if you have a forced-air heating system.

#### Power Supply

A proper AC power supply with grounding as specified must be maintained for your computer system. Provide dedicated outlets, or make sure that other electrical devices such as motors, heating or cooling appliances, or fluorescent or blinking lights are not plugged into the same outlet/circuit.

At one time, UPS (Uninterrupted Power Supply) systems to protect the integrity of the power supply—and therefore the computer's operation—cost in the tens of thousands of dollars and were only available for huge mainframe installations. You can now purchase a UPS for home or office use for under \$100, and a simple surge protector, to prevent equipment and data damage due to spikes (irregular jumps or drops) in voltage, can be purchased for personal computers for as little as \$25-30.

#### Leave It Turned On

While you might think that the opposite would be true, leaving your computer with the power on during the entire normal working day will actually improve its reliability and extend its life. The "POWER ON" cycle activates all of the components in your computer with a sudden surge of power, and the thermal expansion and contraction can stress solder joints, as well as connections within the integrated circuits. Therefore, use your on/off switch sparingly.

#### Cables

Cable-related problems are an obvious, but often overlooked cause of computer failure. Make sure that you are using the appropriate cables for all components of your system. When your computer is installed, all cables should be dressed (bundled and secured with cable ties) to minimize contact with desk drawers or passing pedestrian traffic. The cables should be in good condition, free of kinks or tight bends which might damage them. Check to ensure that all connections are secure.

### Your Monitor

The screen of your monitor should be cleaned as needed with an antistatic fluid, not ordinary window cleaner. Use the product recommended by the manufacturer or an equivalent.

It is vitally important that you prevent image etching, a visible mark which is indelibly burned into the phosphor-coated screen of your monitor. Turn down the monitor's contrast when you are not using the computer (many users will simply turn off the monitor while leaving the computer itself turned on), and set the contrast only at a comfortable level for viewing while in use. Having the intensity turned up all the way for an extended period of time can cause etching or a degradation in the general quality of the display. Etching may also be caused by leaving one image on the screen for an extended period of time. Be aware that some monitor warranties will not cover burn marks on the screen.

### Protecting Your Printer

The majority of printer problems can be traced to either the paper handling or the printing mechanisms, and are generally the result of not complying with manufacturer specifications. Paper of a type or weight not designed for, or compatible with, your printer can cause serious damage. Paper should be stored in a clean dry place and stacked properly behind or under your printer so that it will "feed" into the printer easily and without obstruction. If paper is fed from a tray, "fan" the paper before loading the tray, and pull out any sheets which are curled, ripped or otherwise damaged.

The ribbon of your dot matrix printer should be changed according to manufacturer recommendations. Worn ribbons will not provide good print quality, and fibers from worn ribbons can clog print heads. You should also avoid using "hard strike" printer settings as much as possible to reduce wear-and-tear or damage to your printing mechanism. For laser printers, use only toner cartridges and other components manufactured or recharged to manufacturer specifications; laser toner recharging has become a fast-growing cottage industry, and not all rechargers may be of the same caliber.

### Disk and Tape Media

Whether your computer uses floppy diskette drives, hard drives, tape or some combination,

data storage media are vital to your computing and should be treated accordingly. The drives utilize read/write heads to record and then retrieve data encoded in the magnetic coating of the disk or tape. Most manufacturers recommend that the heads of your floppy diskette or tape drive be cleaned periodically with a specially designed cleaning diskette or tape. They are impregnated with a chemical which will remove any dust or magnetic particles which may have accumulated with normal use on the heads.

As mentioned earlier, read/write heads are particularly sensitive to movement, so the user should be careful never to drop or shake the computer, or move it at all while it is in operation. If you are standing your computer on its side (e.g., for use under a desk), make sure that it is secure in its position. Most computer retailers or catalog outlets sell a pedestal base which will hold the box tightly and evenly upright.

A note about magnetic media: Buy the best you can. Quality varies widely among floppy diskettes and tapes, and the quality of the media has a clear and definite impact on the quality of data storage. The better the quality of the product, the better and thicker magnetic media which is applied to it, and the more accurately data will be encoded and subsequently retrieved.

Floppy diskettes also require special handling to prevent damage to the data stored on them. Avoid using diskettes which are damaged (folded, creased, etc.) or those which have been contaminated with fluids such as coffee or soft drinks, or abrasive particles such as metal filings or dust. Handle the diskette only by the edges, and store it in its envelope when not in use. Keep the diskette away from paperclips and any kind of magnet, and do not expose it to excessive heat or sunlight.

---

### "But I Did Everything I Could . . ."

Despite your best efforts at preventive maintenance, your microcomputer may malfunction one day—probably at the most inopportune possible moment. The smartest thing you can do is assume that a breakdown will occur—while hoping that it doesn't—so you won't be caught unprepared should your system fail. In addition, remember that while your computer system is replaceable (although that's not a pleasant thought), your data may not be.

Protect critical data from loss or damage by backing it up on an ongoing basis and storing the backup copies in a safe area—perhaps a fireproof box or safe—away from the computer system. In addition to protecting you in the event of computer breakdown, this will also secure your data if there is a theft of equipment or natural disaster such as fire or flood. Keep in mind the statistics from the University of Texas which indicate that 43% of all companies which suffer from a disaster and do not have a plan for recovery never reopen for business.

### What Are My Service Alternatives?

Whether you use your computer at home or in the office, if it's important to you, it should be protected by some type of service program. And the program should be in place BEFORE you have a problem with the equipment. After all, you spent a great deal of time and money to select the appropriate microcomputer products and software; your service decision is equally important.

There are several different types of service, and you should consider carefully the best alternatives for your particular situation:

- Carry-in (also called depot) service is a cost-efficient option if you are located within a reasonable distance of a service center. This option can also yield a quick turnaround on your repairs, but requires that you devote the time to bring in the computer yourself. Some service companies will offer a pickup-and-delivery service at an additional charge.
- Ship-in service is practical if you are some distance from a service center, and if you can do without your system for a longer period of time (assuming that at least 2-3 days must be added to the repair time for transit back and forth). You must also consider the repacking and cost of shipping. If you plan to use this option, remember to keep all the original boxes and packing materials from the various components of your system.
- On-site service is the most convenient option and will generally offer the fastest repair time. The technician, upon diagnosis of your problem by a help desk, will arrive at your location with the parts necessary to get your computer up and running in the minimum amount of time.

Because on-site service requires a technician's time to come to your location, it tends to be more expensive than the other options.

Computer service and support is big business—it's currently estimated at over \$40 billion and is expected to double over the next five years. There are a multitude of service providers and competition is stiff for your service business. That translates into more choices for you, the end user. Competition in the service market has resulted in generally higher quality service, more service programs and options available, and more aggressive pricing. There are several key sources of maintenance for your microcomputer:

### Self-Maintenance

From owners of single machines to corporations with thousands of systems, many end users maintain their own computer equipment. Most PCs are, in fact, fairly easy to work with, and an increasing number of people are becoming less intimidated about removing the computer's cover and installing upgraded memory, add-on boards and so forth. Companies which maintain their own computers will claim better control, faster response and repair times, and lower costs than by using an outside service provider. They will cite the hourly labor rate of their own service staff, compared to \$100 or more per hour for a service call.

In fact, the disadvantages of self-maintenance will usually outweigh the benefits. First and foremost, the end user should ask him or herself, "Do I want to allocate my (or my company's) time and resources to service computers, or are those resources better invested in our primary business?" The following are just a few factors to be considered by the potential self-maintainer:

- Salary and benefits for technical and dedicated managerial staff
- Hardware and software capabilities and training of the technical staff
- Coverage during vacations, illness, training, etc.
- Management and administration of a help desk to receive and process service calls from users
- Costs associated with hardware and software necessary for call tracking and reporting
- Costs associated with spare parts planning, purchasing and storage



- Costs associated with diagnostic software, tools and test equipment
- Billing and chargeback considerations
- Facility requirements for personnel, inventory and repair
- Additional resources needed to resolve difficult problems
- Access to documentation from the manufacturers, as well as warranty repair reimbursements
- Inflationary cost increases which might be avoided with multiyear contracts with third-party maintainers

### **OEM (Manufacturer) Service**

There is no question that the manufacturer of your computer or your peripheral device will be the most competent in its service—the manufacturer has the experience, the documentation and the spare parts already in place. In addition, the OEM is the first line of contact if service is needed while the system is still under warranty. If you are a large corporate end user, such as a data processing or micro products manager, you will probably also be able to leverage future product upgrades and sales to be guaranteed the highest levels of service from the manufacturer.

However, there are several key disadvantages to dealing with OEMs: First and foremost is that, while they can service their own products, they rarely have the capability (or willingness) to support environments with equipment from multiple manufacturers. They may either subcontract service on products not manufactured by them, or else simply provide preferential services to their own systems. You must keep in mind that the manufacturer wants to sell you products, not service them, and will therefore tend to offer service primarily as a means of maintaining account control for the purpose of future product sales. Finally, manufacturer service tends to be less flexible and more expensive than other service programs.

### **Reseller/VAR Service**

If you are the average home computer user, odds are that you bought your system in a computer store. Larger companies which purchase hardware and software bundled together for specific applications may have worked with a VAR, or Value-Added Reseller. This dealer adds value to

hardware purchases by configuring customized microcomputer solutions. In either case, the reseller is well-suited to provide in-warranty service, and would probably be the first person you call when your system fails. Like the manufacturer, you can probably “bargain” for the best service delivery by leveraging it against your hardware purchases.

As with the manufacturers, the primary disadvantage of reseller service is that it is interested in selling you product, not support. Furthermore, many resellers have limited resources or geographical coverage to support a comprehensive maintenance program, particularly in the case of a larger company or one with installations outside the reseller's immediate area. Among franchised resellers, many end users find that the quality and levels of service delivery vary widely from store to store.

### **Third-Party Service**

Third-party (also called independent) service providers are those companies not associated with manufacturers or resellers which are set up solely for the delivery of service and support. The primary advantage of this is that there is no bias toward selling products; the TPM (third-party maintainer) can devote his or her full resources to providing you with the most cost-effective service solutions customized for your particular needs and budget. Third parties generally support a wide range of products and manufacturers, so end users with products from multiple manufacturers can find a single source for service. TRW Customer Service Division, for example, the leading independent provider of microcomputer maintenance and support services, has alliance relationships with over 125 manufacturers, covering more than 2,500 different product offerings.

The disadvantage of third-party services arises from their independent status: Manufacturers may be resistant to giving the TPM documentation for its most current computers, particularly in the case of larger systems, or may not authorize the TPM to provide in-warranty service. IBM and Apple Computer, for example, will only provide authorized warranty service through their network of dealers.

However, some end users who wish to use a third-party service organization may get permission from the manufacturer to allow the TPM to

support the warranty; this must be done on an account-by-account basis.

---

### **What Do I Look For in a Service Company?**

Because you have such a great investment in your computer systems, there are several criteria you should use in selecting a provider of maintenance and support services:

- Can the company support the products you have installed?
- Can it grow with you, meeting your future computing needs?
- Will the company offer support for your hardware, software, upgrades, installations, moves, deinstallations, asset management and reporting requirements?
- Can it custom-design a service program which meets all of your service needs, as well as your budget?
- Is the company's pricing competitive?

Most importantly, you must make an objective assessment of your own computer maintenance needs to determine what you really need. Recalling the three service options mentioned earlier, you should ask yourself if your system needs on-site service when carry-in or ship-in might be equally suitable. Similarly, if you do not use your computer for applications critical to your company, you probably should not pay the additional expense of more rapid response times, e.g. four-hour versus next-day response. For larger companies, levels of service delivery may be adjusted according to types of equipment or location, based on criticality of use.

Wherever possible, standardize your micro-computer hardware and software products and configurations to simplify maintenance and support. Use a single service provider to take advantage of discounts based on your total service volume. And, of course, train yourself and your other computer users in basic PC operation and troubleshooting; this will save you the cost of NTF (no trouble found) service calls. ■

# A Guide to Keeping LANs Running Smoothly

## In this report:

Maintaining Network Data/Security .....	3
Managing an Expanding Network .....	5

## Datapro Summary

As companies implement critical business applications on their local area networks (LANs), keeping networks up and running becomes increasingly important. Network downtime is very costly in terms of lost productivity. Major corporations report capital losses resulting from network crashes. Studies report an average loss of productivity resulting from network problems to be in excess of \$3 million per year. The average network is completely or partially disabled twice a month, and the average duration of network inavailability is more than half a business day. Thus, it is apparent that network downtime is a serious problem. It is incumbent upon LAN administrators to take the necessary steps to minimize network downtime, ensure security, and ease network growing pains in order to keep the network running smoothly.

## Minimizing Network Downtime

### Documenting the Network

The single most valuable weapon in the war against downtime is documentation. During local area network (LAN) installation, the network administrator should create a manual file for each component of the LAN (hardware and software). Establishing a standard form of documentation makes it easy for the network administrator to locate vital information when troubleshooting the network or contacting technical support. Network documentation should include hardware, software, and user information, along with the physical configuration of the LAN cabling, network addresses, and devices (such as routers, bridges, gateways, and servers).

—By Amy Force and  
Ben Mayberry  
Business Systems Group, Inc. (BSG)

### Hardware Files

For each piece of hardware, create a file that includes copies of the purchase order, packing slip, invoice, warranty information, and serial number. When repairs are performed, add related documentation to the file. Organize the filing system by the hardware serial number or by a unique internal asset number. You may also want to keep electronic records of hardware devices. There are several tools available to track hardware; these include PC Tracker (RG Software Systems, Inc.), NetManager (Brightwork Development, Inc.), and Micro Resource Manager (Computer Associates International, Inc.). Even when on-line records are available, however, hard copies of the documentation should be maintained.

### Server and Workstation Configuration Files

Document each server and workstation configuration. File server configuration sheets should include general diagnostic information. The leading network operating systems include software utilities that report this information; for example, these statistics can be found using FCONSOLE

for NetWare networks. Include the number of file server processes; statistics on file server and network utilization, availability, and memory; number of files open; and number of users. Workstation configuration information includes the workstation model, memory, network address, drives, fixed disks, LAN adapter configuration, config.sys, shell.cfg, autoexec.bat, and workstation location.

### Software Installation and Configuration Files

Files should also be established for each software package installed on the network. Include copies of the purchase order, packing slip, invoice, completed registration form, and warranty information (if applicable). It may also be helpful to keep track of who is responsible for the manuals. Maintaining current software information can be very valuable because manufacturers often require detailed purchase information from users that want to take advantage of upgrade offers. (For example, Lotus Development Corp. offered free upgrades to release Lotus 1-2-3 3.0 for anyone who had purchased release 2.01 after September 1988.)

Develop a software configuration sheet for each software package on the network. Each software configuration sheet should include the directories where software was installed, trustee rights for those directories, file, flags, default configuration. Also include information on how to modify defaults and access software. Document users' configuration files and search mappings, along with any special hardware or software required to use the package (e.g., a plotter).

Finally, create a general information sheet containing specific information about the software and its required files. Include the application name, executable filenames, overlay files, hidden files, or temporary files used by the software. The protocols required, along with the logical drive definition, print queues, and job definition, can also assist the network manager.

### User Configuration Files

Adding, deleting, and changing user privileges are routine tasks for the network administrator. Maintaining files documenting users, groups, privileges, and login scripts can simplify the process and provide insights into the creation of logical user groups.

### Physical Connectivity

Documenting the physical schema of the LAN can prevent crawling around in the ceiling space looking for the end connection of an unmarked cable. The first step in documenting the physical layout of a LAN is establishing a naming and numbering convention. Develop number schemes to identify MSUA, including closet number and sequence. Assign meaningful names to identify routers, bridges, servers, and gateways. If possible, assign your own locally administrated addresses (LAA); the address location can quickly be isolated to a building and segment.

### Graphical Representation

A graphical representation of the physical connectivity of the LAN is extremely beneficial, especially when the network covers more than a single office. The easiest way to create a picture of the network is to use a physical network management system. Physical network management systems, like Isicad's Command cable management system, provide network information in both graphic pictures and reports. Physical network management systems combine a graphical front end with a standard relational database.

The database supplements the graphics with detailed information on network components, and database changes are reflected in the graphics. Reports that can be generated by a network management system include work orders for moves, adds, changes, and repairs; reports on equipment schedules; cable schedules; cable tray accommodations; and bills of materials. However, the primary reason to maintain a physical diagram of the network is knowledge of the exact network configuration for troubleshooting.

### Change Control Procedures

A LAN is an ever-changing creature. To keep changes under control, a change control system and standard procedures should be developed to help organize and minimize change. A log book of anything that is "done" to the server can be a useful tool and can provide an audit trail in the event of server failure. A change control system will also help the LAN administrator organize daily requests for adding or upgrading software, users, and workstations. Change control forms should include the user's name and location, details on the equipment, and network and physical location. This information can also provide a history of devices on the network to aid in diagnostics and troubleshooting.

### Maintenance Agreements

Maintenance contracts (or arranged service contracts) are beneficial in LAN environments where it is impractical to maintain a team of skilled technicians and spare inventory. There are three common types of service contracts: time and material contracts, term maintenance contracts, and hot spare contracts.

### Time and Material

Time and material (T&M) contracts are the least comprehensive. Parts and labor are provided on an as-needed basis. T&M contracts usually specify service costs, such as set rate per hour, technician qualifications, trip costs, minimum billed hours per call, equipment covered, response time, and warranty repairs. The areas to pay particular attention to are the equipment covered under the contract, the response time promised, and how warranty repairs are handled. Some contracts add a fee for repairs performed under warranty.

### Term Maintenance

Term maintenance contracts cover a specific piece of equipment—for example, the file server—over a specified period of time. A term maintenance contract establishes specific contractual performance requirements, such as scheduled preventive maintenance, spare parts on reserve, and turnaround time guarantees. Term maintenance contracts differ from T&M contracts in that they pertain to a discrete piece of equipment and provide preventive maintenance as well as a guaranteed maximum length of downtime during repairs.

### Hot Spare

Hot spare contracts are the most costly, guaranteeing that a failed device will be replaced immediately by a backup. This type of contract is generally maintained only on critical devices such as file servers, but should also include the device's peripherals (e.g., disk drives). Backup equipment should be tested regularly to ensure that it will work properly if the primary device fails.

## Troubleshooting

Preparing for trouble by documenting the network and establishing maintenance procedures and contracts proves its value when the network fails. Once the network is down, the administrator's job is to identify the source of the problem and fix it.

### Troubleshooting Methodology

Like network documentation, the key to effective troubleshooting is establishing a baseline reference in advance of actual problems. The baseline helps indicate normal network utilization during the day, dominant applications, protocols, and other network performance characteristics. Network interfaces, including gateways, repeaters, bridges, and routers, should be examined, and their normal performance documented before problems occur.

Troubleshooting requires knowledge of network operation and of the relationship between the symptoms experienced by users and possible causes. First, observe the problem symptoms. Then ask the following questions:

- Are only a few users affected or are all network users having the same problem?
- Are there particular times when this problem manifests itself (e.g., during peak hours)?
- Has anything changed lately?
- Are all the release variables and versions experiencing the same problem?

These questions should provide clues to the range and scope of the problem, the percentage of time the problem occurs, and whether a change in the network may have triggered the problem. The change control log can be invaluable in determining recent changes in the network.

Next, consider how the network differs from the baseline discussed earlier. Could an increase in users affect the performance of the network? Then develop a theory and a method to test that theory. This is where a good understanding of network protocols and the applications running on the network comes into play. Once the iterative process of theorizing and testing is completed, conclusions can be drawn concerning the cause of the network problem.

The most common problems encountered in networks involve connectivity and configuration. Connectivity can be disrupted by breaks or shorts in cables and malfunctions in hardware. Configuration errors usually occur across bridges, routers, and gateways.

Certain network configurations are more conducive to troubleshooting. Token-ring networks use active and passive status monitors to provide feedback in the form of MAC frames. Token-rings can recover from some failures by "wrapping around" the faulty node. Token-ring adapters are also self-checking; when they fail they remove themselves from the network without taking the network down.

The network operating system (NOS) can also do part of the troubleshooting. Some NOSs provide tools that log errors as network events and confirm that critical network resources are present. Most network operating systems have built-in utilities that interact directly with the disk drive subsystem. NetWare, for instance, provides services such as COMPSURF, VREPAIR, and DISKED.

## Troubleshooting Tools

Three types of tools commonly assist in network troubleshooting: physical layer tools, network monitors, and network analyzers.

*Physical Layer Tools:* These include time-domain reflectometers (TDRs), oscilloscopes, breakout boxes, and power meters. Physical layer tools help identify cable opens, cable shorts, unterminated cables, and poorly functioning connection hardware. Perhaps the most useful of these is the TDR. TDRs send signals along the physical medium at regular intervals. The returning signals' reflections provide a representative waveform which shows the placement of network devices and location of any cable problems.

*Network Monitors:* These monitor all or selected portions of network traffic. They help compile statistics on network utilization, including packet type, number of packets sent, and packet errors. These statistics are useful in establishing baseline performance and identifying problem areas. One of the best features of network monitors is that they can watch the network 24 hours a day. Network monitors provide relatively low-cost error detection facilities, and integrate easily into the network management schema.

*Network Analyzers:* These assist in locating network problems and testing solutions. Many network analyzers now incorporate on-line troubleshooting guides that provide probable causes for network problems. They provide extensive, detailed information through realtime network traffic analysis, including packet capture decoding and transmission statistics. Things to consider in selecting a network analyzer are the number of protocols supported, collection capabilities, transit capabilities, maximum capture data rate and sustained data rate, ease of use, and format of display.

## Summary

A checklist to help network administrators minimize network downtime should include the following steps.

- Keep detailed documentation on hardware, software, users, and network configuration.
- Document the physical schema of the network including servers, routers, and gateways.
- Implement a change control procedure to manage changes as well as to provide detailed documentation.
- Consider signing a maintenance agreement, especially for critical hardware.
- Develop "normal" statistics for reference.
- Develop a troubleshooting methodology to diagnose and treat network problems.
- Consider purchasing network troubleshooting tools such as TDRs, network monitors, and network analyzers.

---

## Maintaining Network Data/Security

### Network Security

Users depend on network administrators to maintain the confidentiality, integrity, and availability of network data. This includes keeping data safe from human disaster, as

well as natural disasters such as floods, fires, and earthquakes. The tasks required to ensure the security of the network are:

- control access to network data by unauthorized users
- maintain data integrity using data mirroring and backups
- developing disaster recovery plan

### Securing Sensitive Data

The first step in determining LAN security needs is to determine the sensitivity of the data and develop a suitable network security policy. The company's business needs determine the security approach that is most suitable for the network environment.

Using the "common sense" approach, network users follow some basic security precautions. These include locking office doors and desks, changing passwords periodically, avoiding easy-to-guess passwords (like a spouse's name), and prohibiting two users from having the same password. Banning automatic logon and ensuring proper logoff, as well as automatically logging users off the network after a set period of inactivity, helps keep unauthorized people from gaining access at an unattended computer. Restricting access to the LAN server and removing the server's keyboard and monitor prevents authorized or unauthorized personnel from attempting to "fix" servers themselves. Changing factory default passwords and settings in the network software is important when installing the network (make sure documentation is kept on these changes). These are low-cost ways to secure a network and can be implemented without much work on either the network administrator's or users' part.

If the network requires more confidentiality or contains especially sensitive data, employing data encryption and user identification authentication may provide the security level needed. Data encryption protects against breaches in security like wiretapping and unauthorized file access at the system level. The encrypting process encodes and decodes messages using an algorithm and a key. There are two types of encryption systems commonly used: symmetric and asymmetric. The symmetric system requires the sender and receiver of the message to use the same algorithm and key. In the asymmetric system, the algorithm and key for encoding the message to a specific person are distributed, while the decoding key is known only to the message recipient. Using encrypted data adds a great deal of responsibility to the system administrator to track and assign keys and algorithms.

User authentication can replace or enhance the logon password. Several devices are available for authenticating users. One low-cost device uses tokens (small plastic cards containing personal user identification) that are inserted into readers at each workstation. More expensive devices read fingerprints, retinas, voiceprints, or other unique biological characteristics.

### Securing the Network

Restricting physical access to network facilities is key to securing the network. Access to wiring closets, servers, bridges, and gateways should be limited to authorized personnel only, since these are the major "hubs" of network communication. The threat of unauthorized personnel accidentally or deliberately causing network failure is greatly reduced by limiting access to network devices. Restricting access to the server includes removing input/output devices such as the monitor, keyboard, and disk drives. If a

disk drive is required to reboot the server, it should be disabled once the server is up and running. The addition of workstation locks to prevent physical start-up, block physical start-up, or mask the disk drives from the start-up poll can prevent access to workstation files. Diskless workstations can prohibit copying of network files to diskette.

A number of products provide workstation security. Programs such as *Fastlock* (Rupp Corp.) and *The Gatekeeper* (International Data Security) encrypt the disk partition table and require a password when the computer is switched on. *Certus* (Certus Intl.) is a TSR program that compares programs that the user is trying to run against an approved programs list. Other products like *Net/Assure* (Centel) establish user privileges at the workstation level and create audit trails for each workstation. Some network operating systems automatically disconnect inactive workstations.

### Data Integrity

Mission-critical data requires special security considerations to prevent corruption of files or records in the event of network failure. Two popular methods of insuring data integrity are disk mirroring and disk duplexing.

Disk mirroring (also called disk shadowing) uses two disk drives simultaneously. The second (backup) drive duplicates the information on the first. The disks are not identical, since each can be expected to have unreadable portions in different physical locations, but they contain exactly the same information. In the event of a disk failure, the surviving disk takes over.

Duplexing is an extension of disk mirroring. The difference is that when data is retrieved from the server, the disk that has the information most readily available is the one the server reads. This enhances server performance as well as providing data protection. Duplexing provides resistance not only to disk failure, but to failure of the disk co-processor board and/or disk controller.

### Backup and Recovery

Backing up network files is vital. There are three steps to a solid backup strategy.

1. Determine the frequency of data backups.
2. Choose a backup procedure and backup media.
3. Test the backups to ensure that the data is retrievable.

The frequency of data backups depends on a number of factors, including how critical the data is to the business, the volume of data, and the resources available to perform backups. Mission-critical data may change several times during a business day, so even daily backups may not provide enough protection. Servers that contain mission-critical data should be disk duplexed or mirrored. The most practical and popular backup solution for LANs is a tape backup system. During the period that the LAN is used least (usually at night), network data files are automatically copied onto magnetic tape. The tapes are then archived and stored off-site. Once a backup schedule has been established, using multiple tapes in rotation can help protect against tape failure. Digital Audio Tape (DAT) backup drives and tapes are becoming increasingly popular because they can store approximately 1.1GB of data on a single tape. Other backup systems include Write Once Read Many (WORM) drives and optical drives.

A number of automated data management systems are available to assist the network administrator in maintaining backups. *ARCserve* (Cheyenne Software) combined

with *LaserStor* (Storage Dimension), and an erasable optical drive, is a software/hardware combination that performs backups. ARCserve performs full and incremental backups interactively using the NetWare queuing services to schedule and dispatch to the backup device. LaserStor provides random access to archived data. In the event of a drive failure, LaserStor can be used as a direct access device and as an emergency replacement for the server's hard disk.

*The Network Archivist* (Palindrome) is a rule-based expert system that combines the activities of backup, archiving, restoration, and file system maintenance. Restoration can be accomplished on selected files, which is a key feature during any recovery aside from complete disaster recovery. File maintenance features maintain the automatic migration of unused files to tape using rules specified by the LAN administrator.

In addition to backing up data files, NOS files containing the system configuration should be backed up. Backing up network files, such as the bindery files in Novell NetWare, allows the files to be used not only in case of failure, but when a server needs to be downed in order to load a new process. Network system files can also be used to endow new file servers with the same configuration of users, groups, and print services.

### Planning for Disaster

While most LAN administrators faithfully back up their networks and use backups to recover files accidentally lost by users, many administrators do not plan for a disaster. A study performed by the University of Texas estimates that 43% of all companies that do not prepare for a disaster and suffer one never reopen, and 90% that do reopen are out of business within two years.

The first step in minimizing network downtime following a disaster is preplanning—determining risk and the systems necessary to cover critical business functions and support operations. A company is at extreme risk if it copies damaged files, does not store backups off-site, or has not scheduled backup operations correctly.

The cost of a disaster, such as an office fire, flood, or earthquake, can be calculated as the loss of revenue for the estimated downtime. That downtime may be a day, a week, or longer. This price far exceeds the costs of implementing a disaster recovery plan. Disaster recovery plans are similar to insurance, and there are a number of consultancies that specialize in developing disaster recovery plans for businesses. There are also software packages available for the smaller businesses. For example, Chicor's *Total Recovery Planning System (TRPS)* software provides a step-by-step program for larger companies that costs approximately \$10,000. Chicor also offers *Disastar* for smaller LAN environments, at a cost of approximately \$3,000.

### Summary

Maintaining network data and security involves the following steps.

- Encourage network users to take a personal interest in security by locking desks and doors, and changing their passwords.
- Ban automatic logins and easy-to-guess passwords.
- Time-out inactive machines.
- Remove input/output devices from servers and other critical network equipment.
- Use encryption schemes for critical data.
- Restrict physical access to important network hubs.
- Establish a backup strategy.
- Establish a disaster recovery plan, or review the existing plan.

## Managing an Expanding Network

Minimizing LAN downtime through comprehensive documentation, effective troubleshooting, and solid backup/recovery procedures is key to keeping a LAN running smoothly. Another component of a well-managed LAN is the ability to efficiently accommodate expansion. This includes handling increases in users, creating a productive network environment, and updating network software.

### Managing User and Group Information

Grouping users functionally according to similar duties and responsibilities significantly simplifies user administration. Groups include users that require similar access to network resources. It is easier for a network administrator to provide privileges for a group once, then add users to that group, than to provide privileges for each user individually.

The leading NOSs include naming services which assist the administrator in adding both users and groups to servers on the network. Prior to naming services, the LAN administrator had to add users to each of the network's servers individually. Before naming services, each user had to be added four different times, and each server would keep duplicate information.

Most network naming services are based on the principle of domain-based naming. A server is placed within a group of servers (i.e., a domain). A user logs on to the domain instead of a single server. This allows the user to share the resources of a number of file and database servers without different login IDs. Each of the leading NOSs handle naming services differently, so it is important to know how the network responds when applications and other network resources are moved. For instance, moving an application to a different server may require a change to all of the users' drive mappings.

Several products are available to track user information such as access privileges, rights, group membership, account restrictions, login scripts, and user requirements. *LT Stat+* (Blue Lance) and *Lomax Utilities 3.0* (J.A. Lomax Assoc.) provide tools for documenting the LAN environment.

### Creating a Productive Environment

An often-overlooked way of keeping the LAN running smoothly is to create an atmosphere in which users feel comfortable and confident. Developing a menu system is an easy way to do this. It provides users with a list of available services on the network and makes it easy to access programs and files. The leading NOSs provide menu-building systems, and a number of third-party products are available for DOS and Windows environments. Graphical user interfaces (GUIs), such as Microsoft Windows 3.0, also make users more comfortable when using the network.

### Upgrading Smoothly

Network users are becoming more aware of the software packages available for the LAN environment and when new versions or upgrades are available. However, it is the

job of the LAN administrator to determine when an upgrade is warranted. Upgrades involve costs (in both money and time) over and above the purchase price. Issues that must be addressed before deciding to upgrade include the following.

- Upgrades must be installed on all servers and/or workstations; this is a time-consuming task and may result in some user downtime.
- The new software version must be compatible with previous versions.
- Users must be trained.
- The technical staff must learn new features to provide support to users.

The timing of the upgrade should also be considered, based on the requirements of the affected departments. To determine if an upgrade is warranted, balance user needs and wants with the upgrade's compatibility with, and value to, the network. If users are satisfied with the functionality of the current version and are able to do their jobs, then consider postponing the upgrade. If the users are demanding more, then proceed carefully.

---

This report was developed exclusively for Datapro by Amy Force and Ben Mayberry of Business Systems Group, Inc. (BSG). Amy Force is a consultant with BSG SI Consulting's applications development group. Ben Mayberry is BSG's director of consulting. BSG, based in Houston, TX, is a national systems integration company specializing in business solutions based on client/server, network computing technology. BSG can be reached at (713) 965-9000.

### Testing New Applications

Before installing a new software application on the network, thoroughly test and document the application. Most applications can be used in the LAN environment, even though some may be unaware of the existence of the LAN or may not specifically use the LAN. Others are for use only on LANs. Unfortunately, there are no real standards for design, installation, or configuration of LAN-based applications. Applications sometimes require multiple configuration files. Generally speaking, configuration files should be distributed to users' personal directories so they can customize the application if necessary.

### Summary

Following is a list of the key points to address when managing an expanding network.

- Establish logical groupings of users.
- Learn how to use network naming services.
- Consider software tools that help manage users.
- Install a menu system and/or GUI to make the environment more usable.
- Selectively upgrade software, balancing user needs with network restrictions.
- Carefully test applications for problems before placing them on the network for public use.

---

### Conclusion

The network administrator has a difficult job—minimizing network downtime, maintaining data security and integrity, and managing the expanding network environment. There are many specific choices to make in terms of products and procedures, and there is no single universal solution to network management. Following the basics laid down in this report, however, will create a solid foundation for keeping your LAN up and running smoothly. ■



# Control Change From the Ground Up

## In this report:

Identifying Elements to Manage .....	3
Benefits of a CMS .....	3
CMS Costs Vary .....	4
Setting Policies and Procedures .....	4

## Datapro Summary

Configuration management systems (CMSs) manage change. This report analyzes how a LAN CMS will help LAN managers track and report the status of all actions affecting network configuration. The benefits and costs of CMS are discussed.

Just as an air-traffic controller determines the course, speed, altitude, and other parameters of arriving and departing airplanes, a network manager must control the operation of LAN components. Fortunately, the principles of configuration management (CM) traditionally associated with software development on host computers are beginning to be applied in the LAN arena. As a result, a new class of products is emerging that allows network managers to keep their networks flying.

The main purpose of a configuration management system (CMS) is to manage change. A LAN CMS is defined as a system that gathers, tracks, and reports information generated from installing, initializing, and modifying the configuration parameters of network hardware and software. This report presents the significant issues facing LAN managers who are considering implementing CM on a network and the benefits that can be realized from developing a CMS. A few of the tools available for developing such a system are also discussed. These tools typically support NetWare and LAN Manager, but tools for other network operating systems are beginning to appear.

This Datapro report is a reprint of "Control Change From the Ground Up" by Maura A. Hart, pp. 26-28, 30, 32, 34, and 36 from *LAN Technology*, November 1991. Copyright © 1991 by M&T Publishing, Inc. Reprinted with permission.

CM encompasses four processes: configuration identification, status accounting, change control, and configuration verification.

Configuration identification is a matter of identifying which hardware, software, and documentation elements to include in a CMS. In identifying these elements, you also develop a naming and labeling convention.

Status accounting is the process of tracking and reporting the status of all elements; actions affecting network configuration are recorded and reported.

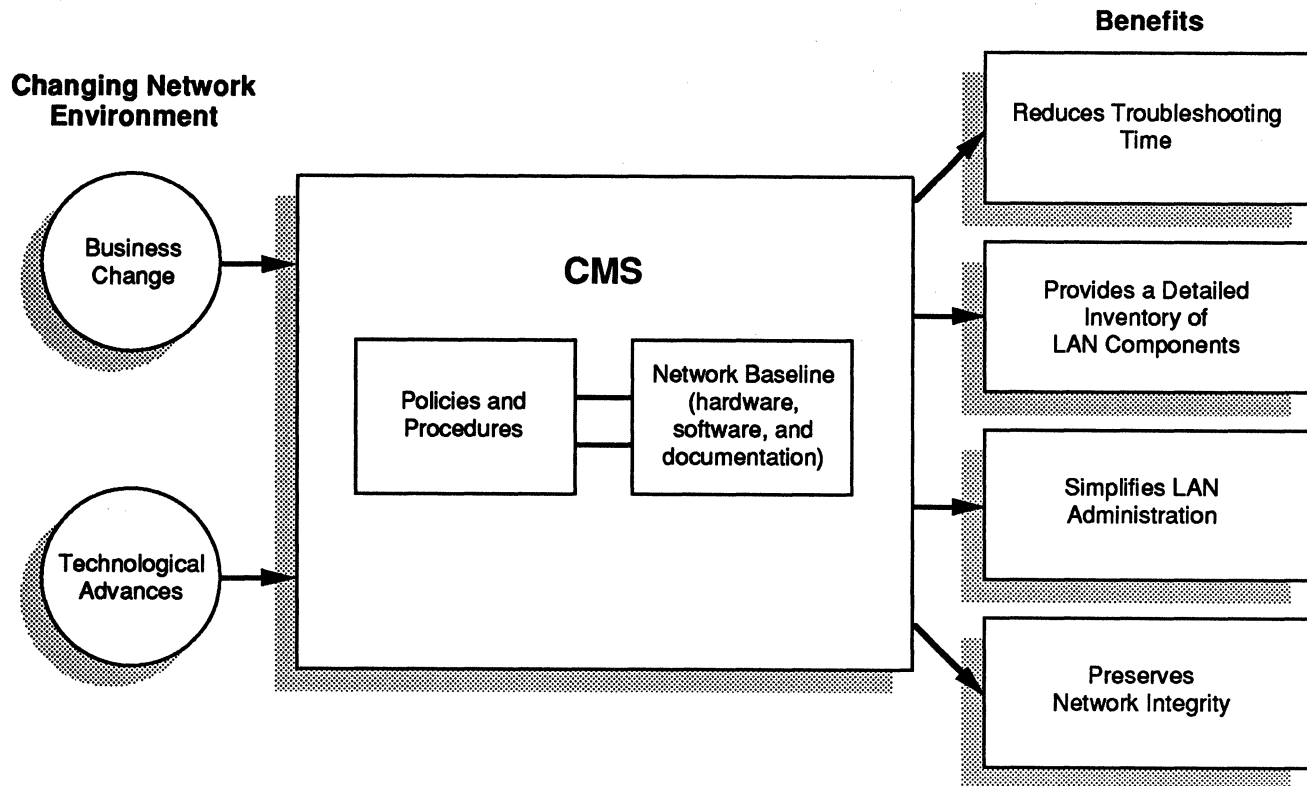
Change control consists of the systematic evaluation, coordination, and approval or disapproval of proposed changes to LAN components.

Configuration verification is a review process whereby the network manager verifies that the configurations of all elements on the network are correct and suitable for their intended function, and that the elements match the network documentation. It is used to compare an element's configuration with its approved specifications.

## Why Do You Need CM?

Unless LAN components operate in a defined way, a network can become inefficient at best and chaotic at worst. A LAN CMS is important because it assists technical management in preserving network integrity, reduces troubleshooting time, and simplifies network administration (see Figure 1).

Figure 1.  
Network Changes Drive the Need for a Configuration Management System



Once installed, a CMS can provide multiple benefits.

LAN managers must define the basic flight plan by establishing a network configuration baseline and tracking the changes to that baseline. A network configuration baseline consists of detailed information about network hardware, software, and documentation. It can include a floor plan, detailed descriptions of all hardware and related settings as well as software and installed versions, and information on the associated documentation.

Implementing CM on a network is more complex than implementing a conventional CMS for managing software development. Instead of a single host computer, LANs consist of diverse devices that are tied together. Because all of the components in a network are interrelated, performing a function in a network environment involves more components than performing it on a standalone computer. This interplay of components makes it more difficult—but arguably more important—to implement CM on a network. As networks begin to carry more of the burden of critical computing, LAN managers are turning to CM to help control network change.

Several factors make it difficult to implement CM on a network: the growth of multivendor networks; the accessibility of LAN components to individual users; and the trend toward geographically distributed LANs. In an effort to share computer resources, computers of all kinds are being networked together. Therefore, the task of tracking each system's configuration has become more complex. Each unique computer configuration increases the number of items a network manager must monitor. For example, IBM PS/2s and Compaq Computer Corp. Systempros will have different settings, all of which need to be tracked.

When configuring one LAN component, network managers need to consider the characteristics of other network components. For example, to avoid conflicts, the network manager needs to know which memory interrupts and base memory addresses are used by workstation video adapters when configuring network interface cards.

LAN components are generally accessible to network users. As a result, configurations are often altered to suit an individual's preference without the system supervisor's knowledge. The workstation is one of the easiest LAN components for an individual to alter. Anyone with a screwdriver can bring down a node by altering a variety of boards and settings. Users don't even have to open up the workstation to change network drivers located in a CONFIG.SYS file. Users can also affect the network by introducing unauthorized software acquired from colleagues or outside sources.

Implementing CM in geographically distributed environments can burden a LAN administrator. Reconfiguring network elements remotely can be taxing if the current configuration is unknown. In distributed environments, network managers may need to travel, reducing the time available for them to perform LAN administration.

Despite these obstacles, tools are emerging that bring traditional CM to the LAN environment. More importantly, makers of network products are expanding CM concepts beyond the scope of software development and using them to manage the change of network components.

## Identifying Elements to Manage

Network managers must first define the criteria for determining if an element should be identified as a configuration item in the CMS. Criteria include the element's number of operational modes, potential to disrupt the network if its configuration is altered, and frequency of configuration changes.

The criteria must also stipulate the breadth and depth of elements to be tracked. For example, a LAN manager may identify a workstation and its hardware components as configuration items, but decide not to include workstation software. The depth of the CMS maybe influenced by factors such as the head count of the LAN administration staff and the complexity of the network. Organizations should strive for a thorough system and include as many configuration items as is pragmatic.

Elements on a network can be classed as either physical or logical. Physical elements (or active devices) include wiring hubs, bridges, and routers. Logical elements include user profiles, print queues, and drive mappings. Manufacturers of active elements often provide a management system to aid with configuring and managing their devices. For example, Cabletron Systems Inc.'s Remote LANView\Windows has a CMS that enables network managers to examine and reset parameters for the company's NB20E, NB25E, NB30, and NB35 bridges. LANView\Windows lets network managers check bridge information such as name, address, number of ports, location, and status as well as perform functions such as resetting bridge counters and disabling and restarting a bridge.

Logical elements are configured and managed through the network operating system and third-party utilities. As network operating systems have become more sophisticated and robust, they have provided more CM features. In addition, there are a growing number of third-party CM products that complement the CM capabilities of network OSeS, including utility packages that can help monitor statistics and track changes to an element's baseline configuration.

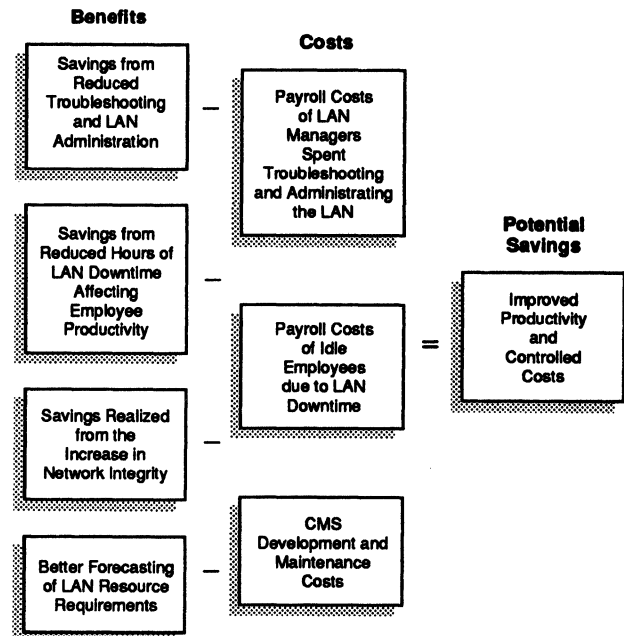
One CM tool for tracking configuration changes is Frye Computer Systems Inc.'s Frye Utilities for Networks. This package is composed of NetWare Management and NetWare Early Warning System. These utilities provide consolidated information on a NetWare server's hard disk and RAM utilization and summarize user names and trustee rights. The packages can also generate reports on server configuration and indicate drivers in use on client workstations.

Not all hardware and software elements come with their own management systems, and those that do are often manufacturer-specific. You can integrate the configuration management of different network elements through the use of products supporting the Simple Network Management Protocol (SNMP) and ISO's Management Information Base (MIB).

For example, Cabletron's Spectrum Version 2.0 supports any SNMP-compliant device. This UNIX-based management package provides information on active devices, regardless of manufacturer, and can create a hierarchical view of a network. It also has the ability to "discover" existing subnetworks and SNMP-compliant devices on the subnetworks.

A sophisticated CMS can also provide an integrated solution. Intelligent Network Applications Inc.'s NetMapper provides CM for physical network devices from multiple

Figure 2.  
The Investment Decision



These are several of the benefits and costs you should weigh when considering whether a CMS can save your organization money.

vendors. NetMapper can discover and map Sun SPARC-stations and DOS machines, bridges, and printers on NetWare and SunNet-based LANs. In addition, it monitors a network for configuration changes and will set off an alarm notifying the network manager of changes that occur.

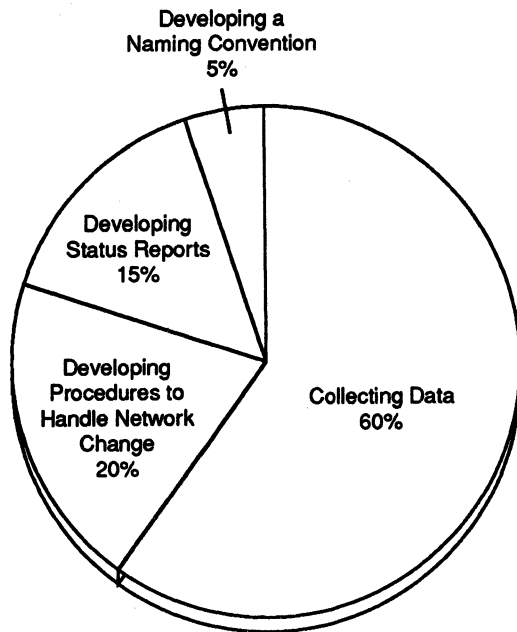
## Benefits of a CMS

CM tools save time and money by simplifying LAN administration tasks and providing accurate configuration information in a manageable format. One example of a LAN administration task that can be simplified with CM tools is the installation of a network operating system across multiple servers. This usually requires a separate installation process for each server. Preferred Systems Inc. has addressed this problem with Origen, which streamlines the installation of NetWare 2.x and 3.x across multiple servers by allowing the configuration file from one server to be copied to other NetWare bindery databases. Users, groups, trustee rights, and directories with standard drive mappings can be copied across file servers.

CM tools also allow LAN administrators to access configuration information quickly and to store the information in useful formats. VisiNet 2.0 from Technology Dynamics Inc. is based on a Windows 3.0 graphical database. It can automatically collect information known by NetWare, LAN Manager, and VINES about servers and workstations supported by these network OSeS. Once the database is set up, network administrators on these LANs can quickly access system configuration data.

VisiNet supports dynamic data exchange, so other Windows-based applications, such as spreadsheets, that

Figure 3.  
Relative Costs of Developing a CMS



are linked to VisiNet are automatically updated when changes are made in VisiNet. In addition, VisiNet has a script language; network managers can create a script to track key network parameters (such as disk utilization) and generate alerts when predefined thresholds are reached.

A CMS also preserves network integrity by helping you identify all elements on the network and their configurations. A CMS enables technical support personnel to respond more quickly and efficiently to problems because they know the configuration details of the system in question. In addition, non-approved elements and non-standard configurations can easily be detected. Providing an accurate inventory of LAN components is another benefit of a CMS. An accurate inventory helps ensure that standard configurations are maintained, thus reducing maintenance costs.

A CMS helps network managers provide comprehensive network administration. An organization that implements CM will benefit financially because properly managed networks require less on-site LAN administration time (see Figure 2).

### CMS Costs Vary

Building a CMS is time consuming. Fortunately, there are products available to accelerate the data gathering process. For example, PC Census from Tally Systems Corp. will inventory memory, mass storage, serial and parallel ports, off-the-shell and in-house software stored on local hard disks, video cards, and modems on IBM PCs and compatibles. The current version requires the manager to either create a collection disk and insert it into a personal computer or to install the collection software on the file server and execute the program from a workstation to collect data on that computer. A future version will allow data to be gathered when users run their login scripts, according to company officials.

The cost of CM tools for networks can range from a few hundred dollars to several thousand. LAN Automatic Inventory from Brightwork Development Inc., for example, lists for \$695 per server. Legent Corp.'s Endeavor for DOS is a traditional software CM system adapted for NetWare, VINES, LAN Manager, and LAN Server networks. Its price ranges from \$5,375 to \$26,875, depending on the number of nodes. Endeavor for DOS keeps track of all software development and maintenance activities for programmers working on a network. (The company is working on an OS/2 version of the product.)

Few tools provide only CM features, which makes it difficult to compare one tool to another. Factors to consider in assessing the cost of a CM tool and the development of a CMS include development, training, maintenance, and portability.

Developing a CMS, like developing any resource, takes time and effort. Building the CMS requires you to gather data, define a numbering scheme to track the configuration items, form procedures to handle network change, and develop reports to reflect the status of LAN components. The initial inputting of data to the database is the most time-consuming part of building the CMS. The relative costs for developing a CMS for a 100-node network as illustrated in Figure 3 are based on personal experience.

Training costs are associated with the time LAN administration personnel spend away from their daily duties learning the CM tools. Maintaining the CMS will require system supervisors to perform backups and timely updates. It is my experience that the annual cost to maintain a CMS is approximately one-third of the initial development cost for a 100-node network experiencing moderate change. Portability is an issue if your organization uses multiple network OSES. You may need to purchase different CM tools for different network OSES.

Potential risks and losses need to be quantified before you make the decision to invest in a CMS. An organization with mission-critical applications on a network takes a great risk if it chooses not to implement some form of CM. In addition, each organization needs to determine the level of CM that it needs. Factors that must be weighed include the network's size and complexity and its importance to the organization.

A network environment experiencing extensive growth and change can incur productivity costs from not implementing CM. Inevitably, the configurations of devices on the network change; unless you keep track of the changes, it will be difficult to reinitialize these devices should you need to. The longer it takes to reinitialize and properly configure devices, the longer they will be unavailable to users. The cost of idle employees can add up quickly.

### Setting Policies and Procedures

Policies and procedures are necessary to define the user's role and responsibilities relating to LAN CM. Users' roles and the roles of those responsible for implementing CM should be stated clearly to avoid misunderstandings and to protect computer assets. If it is corporate policy that LAN staff members are the only individuals authorized to install equipment and alter settings, then that policy should be explicitly stated and known. By establishing policies and procedures that govern CM, the CMS's integrity can be preserved.

Network managers should develop standard configurations, a change control method, a configuration verification process, and a CMS maintenance schedule. Standard

configurations for hardware and software facilitate system installations, troubleshooting, and updates. Configuration management problems can be reduced if standard configurations are a corporate policy. For example, a standard workstation configuration might stipulate hardware options and settings, software applications, and a common directory structure. A standard workstation configuration can benefit both users and management; it allows system administrators to solve problems more easily, and it permits users to be more mobile. For example, a user can cover for a co-worker on vacation by temporarily accessing that co-worker's resources.

Because networks are not static, change control procedures are necessary to outline how changes will affect the network baseline. Proposed changes to network components should be detailed in a change request form. CM personnel should evaluate the proposed change and its impact on the network. An approval or disapproval decision should be reached and documented in a change notice log. Once in effect, approved changes need to be updated in the CMS. Configuration status reports should be generated periodically to summarize change notices and to inform LAN administration personnel of approved network changes.

At times, users bypass the change control process and alter network components themselves. These changes go unnoticed and are not reflected in the CMS. The value of the information provided by the CMS is dependent on timely and accurate updates. Therefore, in order to discover these changes, CM updates should be integrated and

scheduled with routine physical inventories. Physical audits and reviews will detect unreported changes and provide a means for configuration verification. Products such as PC Census can aid in the configuration verification process by providing information on a workstation's hardware and software contents and I/O mappings. It's worth the time it takes to run this program on each personal computer to get the information it can provide.

A CMS maintenance schedule should include security and backup procedures. To ensure system integrity, the CMS should only be accessible to personnel requiring access. The ability to update the CMS should be restricted even further. Backup procedures should address the issues of backup frequency and the number of generations to be maintained. A copy of the CMS should be stored off-site to provide protection against disasters.

Developing a CMS takes time and is an ongoing effort. However, a CMS has many benefits to offer to organizations willing to take the time to develop and maintain such a system. A CMS preserves network integrity and creates a more consistent networking environment. If a change occurs, it should go through the change control process. If someone bypasses this process, the change should get picked up during an audit.

In addition, CM simplifies LAN administration, provides a detailed account of network components, and reduces troubleshooting time. Like a good air-traffic control system, a CMS can give you the visibility and control you need to keep users flying high. ■



---

# LAN Administration: A New Job Function

---

## This report will help you to:

- Understand the various functions performed by a LAN administrator.
- Plan a strategy for effective LAN administration.
- Identify trends and issues shaping the future role of LAN administrators.

---

Network management is the management of operations and assets of the network. The discipline encompasses both services and products. While part of the network management problem is handled by products that keep, say, an inventory of network components, the greater part of the same problem is developing the family of systems and procedures to collect, update, and use the data.

Network management began as a fragmented collection of services aimed at maintaining operational control. Different designs, equipment, rules, and management methodologies blocked the orderly flow of information through the corporation. The network's physical composition

was the major area of focus by technologists. The arrangement of channels, the types and quantities of network devices and the characteristics of terminals and computers are all part of this dimension.

Interactive and transactional forms of processing allowed applications such as airline seat assignment or quotation services for securities brokers and other agents. To take advantage of the processing improvement, the computer department became intimately involved with communications. The link between voice processing and data has forced data processing and communications departments together.

As LANs proliferate and tie computing and communications together, organizations are recognizing that a new network management function-LAN administration is needed to keep networks operating. LANs are distributed computing systems that combine communications

---

This Datapro report is a reprint of "LAN Administration: A New Job Function," by Howard Frank, pp. 38-40, from *Networking Management* December 1990. Copyright © 1990 PennWell Publishing Company. Reprinted with permission.

with end-user applications. As such, the skills needed to operate and maintain them span both the traditional communications and MIS functions.

Most LANs originated as small departmental systems with a few personal computers and perhaps a shared printer or file server. The administration function then involved keeping track of software releases and calling a vendor when the system failed to perform as expected. The LAN administrator was usually a computer-oriented staffer whose main job had to do with the department's primary mission. Rarely was LAN administration part of that individual's formal job description.

### **Today's Job Responsibilities.**

Now, LAN administration starts with the installation of hardware and software. A LAN administrator must:

- Regularly create system backups to lower the risk of data loss.
- Maintain and upgrade operating systems including adding, deleting, and modifying user access privileges.
- Maintain menus and applications.
- Monitor system performance.
- Maintain shared network printers.
- And since even small installations require a means to communicate with the outside world, the administrator must ensure that modems, communications lines, and gateways are functioning properly.

These tasks are all relatively straightforward for small networks and are accomplished usually in a part-time effort. However, as the network grows, support requirements increase and the administration function becomes a full-time activity.

The LAN administrator becomes the central point for reporting system problems. He/she now must spend time each day helping and instructing users and diagnosing workstation problems. The administrator coordinates with external maintenance organizations and works with suppliers' technical support personnel to diagnose problems.

Often the demands grow insidiously and the LAN administrator, originally a user with a different job function, becomes fully absorbed in LAN

operations without any official management decision to divert the administrator from his/her initial job. More often than not, the results of this "non-decision" are unhappy users, an unhappy administrator, and unhappy managers because the need to manage the LAN exceeds the capabilities of the ad hoc administrator, while the original job requirements go unmet.

LAN administration has become a major activity, with administrators managing a "help" desk where users can report trouble, request service, obtain advice, and arrange for LAN system and application training.

System security in any LAN installation, regardless of size, is a crucial element of today's environment. LAN administrators need to be involved in password and log-in control and must also track user activity on the LAN.

Security features are normally built into the network at the time of installation, often using various LAN management packages from independent software organizations or systems integrators. The LAN administrator must therefore understand network management systems and software products that troubleshoot, diagnose, and control the LAN.

The LAN administrator must also insure that the many software releases resident in the network remain compatible whenever they are upgraded and when new releases are added. This function will be time-consuming because the network is becoming the core computing resource supporting many users and applications. Maintaining software compatibility will become one of the major network management problems over the course of this decade.

There are several ways to implement effective LAN administration. In small installations, e.g., small departmental LANs or LANs installed in small businesses, administrators usually are chosen from system end-users. The choice should be a conscious decision and the administrator's job function appropriately redefined. Typically, one administrator and one backup should be selected. They should be trained at the time of installation and on an ongoing basis.

Management of larger corporate or agency-wide LANs usually falls to data processing or MIS/telecom departments. End-user departments and MIS organizations can also use external vendors for support, such as on-site administration or training.



**Future Trends**

Several trends are driving the LAN market toward an enterprise network environment.

- LAN end-user departments are finding themselves incapable of managing and administering their LANs and are also discovering the need to connect their isolated systems to other networks.
- Bridges, routers, and gateways are being sold by specialized companies making LAN/WAN links. This segment of the industry is becoming recognized as the highest growth rate business, but there are relatively few technologists who understand this area.
- Products linking LANs to WANs are being developed and introduced by WAN vendors, in particular T1 products.
- LAN vendors are discovering the need to facilitate connectivity between LANs and are therefore releasing operating systems with global addressing and interoperability capabilities.

These factors are increasing demands on network managers and administrators to become computer and software specialists as well as communication technologists. Further, network administrators will need to establish and maintain regional, national, and global directories of user IDs, addresses, authorization levels, and security information on a daily basis to allow communications in the enterprise network of the future.

**Ubiquitous Access**

To understand the magnitude of this problem, compare it to that of maintaining global telephone directories for private tandem voice networks. Because users of tomorrow's network will be everywhere that the organization operates, network connections will be more common than today's telephone connections, and network access will need to be as simple as using the telephone. The demands on network administrators will be immense.

As time-consuming as maintaining global network directories will become, the problem of document retrieval and management for the interconnected LANs comprising the enterprise networks will be even greater. Procedures and product are required to manage the distributed networks and applications growing from today's LANs into tomorrow's enterprise networks. These products and procedures must be designed for use over multiple sites within large networks, to handle the complex problems of distributing data over networks of thousands of users and many cities. For the most part, these products and procedures do not yet exist, but the need will develop whether or not vendors and users are ready.

Having a strategy for LAN administration is crucial (whatever the strategy may be) since LANs are now pervasive elements of modern information systems. Further, network managers must be prepared to extend their strategies to the interconnected LANs of tomorrow to make the enterprise network not just a technical achievement, but a strategic resource. The alternative is loss of user productivity or, even worse, the breakdown of the information infrastructure. ■



---

# LANs by the Book

## In this report:

The Network Office Manual.....	2
Messy Hard Disks.....	2
Getting Organized.....	2
Security and Access Control.....	4
Accumulated Knowledge.....	5

## Datapro Summary

A LAN that is administered in a haphazard way invites disaster. The key to good network administration is the creation and maintenance of a network office manual that lists the important information about all of the hardware components and software used on the network and the users, vendors, and support contacts. A network office manual can save time and money by helping a network administrator to solve problems without resorting to outside consultants. The manual can also help the administrator to develop and maintain a good security program. This report describes the essential information that should be maintained in a network office manual, gives sample forms that can be used to record the information, and describes administrative procedures that help to keep the LAN operating smoothly.

It's not over when your LAN is installed. After the endless headache of researching products, reviewing quotations, solving installation and cabling woes, tracking down software incompatibilities, and sustaining downtime for employee training, small business owners often find that installing a network was more than they bargained for. They're only too happy to leave behind the network administration issues.

And the users don't object. Because they see little difference between the care and feeding of their PCs and the finicky and temperamental nature of a LAN, network administration is rarely revisited.

When mysterious glitches occur, users often develop creative solutions to sidestep—not solve—the problem. At one small business, the users turned on the server with a yardstick from two feet away because whenever they stood at the desk, a loose cable grounded the system and shut

down the power supply. This had been a ritual for more than two months. After a painful installation process, it was more palatable than more downtime.

But network administration need not be painful. A few simple procedures can help you get the most out of your computing resources.

---

## The Mythical Manager

In small businesses, system administrators are partly mythical. The title is usually bestowed on the person who is most knowledgeable about computers in general. This is often the receptionist, office manager, or even a part-time accounting clerk. Many "elected" system administrators barely know enough to keep the network running. Their most challenging demands are clearing print and e-mail queues and changing passwords. If they have mastered clearing print and e-mail queues, that's fine because they have "real" jobs to do.

Though some administration is better than none, a haphazard approach invites disaster. A default LAN administrator, especially a part-timer, can be unavailable when needed. A well-meaning but untrained administrator, who doesn't know

---

This Datapro report is a reprint of "LANs by the Book" by Don Dresden, pp. 143-152, from *LAN Magazine*, Volume 6, Number 10, October 1991. Copyright © 1991 by Miller Freeman Publications, Inc. Reprinted with permission.

enough about network hardware to solve simple, but abstract technical conflicts, can create more serious problems by attempting even the simplest experimentation.

When rates for network consultants are from \$75 to \$195 per hour, it isn't difficult to calculate that a moderately trained system administrator, armed with the basic utilities and good documentation, can save the company money, reduce downtime, and not create greater problems from inexperience. How can business owners who don't have the resources to hire a full-time system administrator protect themselves against disaster and get better performance? The key is the network office manual, which stores all of the important information about its network and users.

### The Network Office Manual

Network documentation frees the small business owner from being held hostage to a key employee who just "knows" how the LAN operates. A network office manual can be organized so any employee can consult it. In many instances, it holds the answers to the simple problems a network regularly presents. The network office manual also can reduce or eliminate potentially crippling problems that result from a lack of configuration or procedural data. Such a resource doesn't necessarily demand a great deal of time; you can create the beginnings of a manual in about four hours.

The network office manual proves useful for all sorts of problem situations, including when you don't understand why someone can't log into a directory or access a program, when you want to figure out where on the 180MB of tape the latest version of a file is, or when talking with technical support staff over the telephone. Most importantly, the network office manual can help you to develop and maintain a good security program for your network.

### Messy Hard Disks

A quick look at the average, nonnetworked, hard disk will reveal a myriad of forgotten files with save dates from 10 years ago. Old configuration and batch files, miscellaneous memos, drafts, and resumes abound in every directory, if directories even have been created. Lots of hard disk systems have 2,476 files all in the root directory, but not CHKDSK.

Hard-disk management is almost impossible to administer when left to users to do on their own. Even with desktop managers such as Central Point Software's (Beaverton, OR) PC Tools Deluxe, which quickly organizes a hard disk, users rarely take the time. While disorganization might be fine on a standalone PC, messy file habits on a file server court disaster.

Server performance is proportional to the maintenance of its hard disk. Accumulated files quickly gobble up much needed disk space. The most common offenders include log files that are never closed or deleted, old backup history files, e-mail messages, and personal files in user directories.

If your software is disk-intensive (often reads and writes little bits of data to disk) as are many accounting or desktop publishing programs, your server disk is particularly subject to fragmentation. Fragmentation happens when a file's data is spread across the hard disk in many chunks.

Fragmentation contributes to the creation of small, unused, and unaccounted for clusters or spaces that appear

Figure 1.  
Master Vendor File

MASTER VENDOR FILE (Hardware, software, service, support, supplies)			
<b>HARDWARE</b>			
<b>CODE</b>	<b>VENDOR NAME</b>	<b>SALES REP/PHONE</b>	<b>SERVICE REP/PHONE</b>
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
<b>SOFTWARE</b>			
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
<b>SERVICE</b>			
_____	_____	_____	_____
_____	_____	_____	_____
<b>SUPPORT</b>			
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
<b>SUPPLIES</b>			
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____

The first page of your office manual should list your vendors telephone and fax numbers, technical support departments, and contact names.

between files on your hard disk. These spaces are lost from the total disk capacity and will eventually cause data corruption. As fragmentation grows, it can lower the disk speed by 25 percent or more. At the worst, fragmentation causes file corruption and a data loss that can be replicated.

To clean up the hard disk, determine which data and programs users need to do their jobs efficiently and keep only those files. Archive excess applications and files to floppies, local hard disks, or tape. To address fragmentation and file organization, a good network disk utility program such as NetUtils from Ontrack (Eden Prairie, MN) is an important addition to the system administrator's toolbox. With the network office manual, it comes in handy for general network maintenance.

### Getting Organized

Ninety percent of good system administration is knowing basic information and how it relates to other information. Your network office manual contains much of this data organized on simple forms. Photocopy working forms from masters. Each form contains a heading, date, and spaces for relevant information. The main components of a network office manual are: a hardware list, software list, network flowchart, user security and access control information, and backup and recovery information.

Figure 2. Hardware Unit List

**HARDWARE UNIT LIST**  
(Record one item per sheet)

BRAND: \_\_\_\_\_ VENDOR CODE: \_\_\_\_\_

MODEL: \_\_\_\_\_ NETWORK ID#: \_\_\_\_\_

LOCATION: \_\_\_\_\_ SERVER TERMINAL LOCAL PC

TYPE: (circle) Desktop Tower Laptop Notebook Portable Serial #: \_\_\_\_\_

CLASS: (XT, AT, 386, 486, PS II, etc.): \_\_\_\_\_ VIDEO (Mono, RGB, EGA, VGA, PGA, etc.): \_\_\_\_\_ MODEL: \_\_\_\_\_

HARD DISK DRIVE: Size, type: \_\_\_\_\_ FLOPPY DRIVES: 360K 720K 1.2MB 1.44MB

OTHER: Amount of ram: \_\_\_\_\_ Numeric coprocessor: Y N Extended keyboard: Y N Other: \_\_\_\_\_

NET: Network card type: \_\_\_\_\_ ID#: \_\_\_\_\_ Arcnet Ethernet Token ring Other: \_\_\_\_\_

INTERRUPTS/DRIVER CONFIG.: \_\_\_\_\_

Staple or tape printout of: Autoexec. bat file Config. sys file List of drivers Other: \_\_\_\_\_

---

SYSTEM NOTES

For each network device, list its configuration, plus a description. Staple a printout of the pertinent batch files.

The first page of your manual should be a list of your hardware and software vendors' telephone and fax numbers, their technical support departments, and contact names, if you have them (see Figure 1).

### Hardware Forms

Configuration changes, such as adding a printer or changing a batch file, can cause LAN hardware conflicts to appear where none had existed before. The first form for your hardware list should record the brand, description, and serial number of each piece of every device connected to the LAN. Include a code letter with each piece of hardware to indicate which vendor sold you the equipment.

The second form for your hardware list can be a sheet for each network device, including workstations and servers (see Figure 2). Each sheet should include the unit's configuration, including its disk size, amount of RAM, video type, network interface card and device interrupts, and information specific to that device. Staple to each form a printout of the contents of the boot disk or start-up batch file. These should include the AUTOEXEC.BAT file, the CONFIG.SYS file, and a full directory showing all the drivers used for booting the system. Keep copies of purchase invoices here too.

### Software Forms

Though it is often overlooked, every company should formulate a policy on software that's approved for company use. Employees often use databases or word processors

that the company didn't purchase because they use the application at home or are more comfortable with its features. This creates confusion when other workers need data and can bring a project to a halt if the employee leaves the company.

To achieve consistency, use the first form to record applications on the "approved" list, version release numbers, serial numbers, technical support telephone numbers, and any identification required (see Figure 3). Staple to this form a list of all files and directories on the hard disk. You can use the DOS TREE/f command or a disk and file utility to create it.

The second form (see Figure 4) in your software list can be a sheet for every major application and database used by the company. It should include the directory path of where files are kept, the file names, associated files, a brief description of what the application does, who created it and when, and where backups exist.

Users should not be allowed to create random databases. Management should develop a policy for spreadsheets and databases that specifies field lengths, types, and basic design. If such a policy exists, the occasional random database that a user creates can be imported or exported later with little fuss.

### Network Flowchart

The network flowchart is a graphic representation consisting of simple boxes and rectangles that shows the physical location of all servers, workstations, and printers, along

Figure 3. Authorized Software Packages

**AUTHORIZED SOFTWARE PACKAGES**

PROGRAM NAME	VERSION	DISK SERIAL #'s	SUPPORT PHONE	SUPPORT ID #
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____

TRAINERS	SOFTWARE SUPPORTED	TELEPHONE
1) _____	_____	_____
2) _____	_____	_____
3) _____	_____	_____

CONSULTANTS	SOFTWARE SUPPORTED	TELEPHONE
1) _____	_____	_____
2) _____	_____	_____
3) _____	_____	_____

PROGRAMMERS	SOFTWARE SUPPORTED	TELEPHONE
1) _____	_____	_____
2) _____	_____	_____
3) _____	_____	_____

NOTES

For each application on the approved list, record version number, serial number, and tech support number.

Figure 4.  
Software Program Statistics

SOFTWARE PROGRAM STATISTICS		
PROGRAM: _____	NETWORK VERSION: Y N # USERS: _____	
RELEASE VERSION: _____	SERIAL #: _____	# OF ORIGINAL DISKETTES: _____
SUPPORT PHONE: _____	SUPPORT ID #: _____	SUPPORT REP: _____
TYPE: DATABASE ACCOUNTING SPREADSHEET WORD PROCESSOR GRAPHICS DESKTOP PUBLISHING CAD/CAM TELECOM OTHER		
PROGRAM IS LOCATED ON SERVER: _____ OR LOCAL COMPUTER: _____		
DISK DRIVE: _____ IN DIRECTORY OF: _____		
INSTALLED ON: ____/____/____	LAST UPGRADE: ____/____/____	LAST UPGRADE: ____/____/____
FILES ASSOCIATED WITH PROGRAM	DRIVE & DIRECTORY	PURPOSE OF FILE
1) _____	_____	_____
2) _____	_____	_____
3) _____	_____	_____
4) _____	_____	_____
5) _____	_____	_____
6) _____	_____	_____
7) _____	_____	_____
8) _____	_____	_____
9) _____	_____	_____
PROGRAMMER'S PHONE: _____ CONSULTANT'S PHONE: _____		
SOFTWARE NOTES		

For every application, list its directory path, file names, associated files, a brief description, and backups.

with the network identification number of each. This flow-chart helps you plan for disaster control. Determine in advance where you can move a LAN server if trouble occurs.

## Security and Access Control

When installing a LAN, physical and data security should be a primary consideration. Many companies unknowingly throw traditional procedures out the window after computers are installed. You wouldn't let a company clerk walk into the personnel office and begin looking through the file cabinets. Yet, on a LAN, employees often have this access level.

To implement a basic security plan, first educate and warn all employees on computer data's importance to the business operations. At least twice a year check that users are practicing good habits and are not sharing passwords or taking data or programs home without permission.

Identify and segregate crucial or proprietary data. Each business has its own priorities. For some, only accounting files are confidential. For others, engineering drawings, research, or reports are confidential. Sensitive data should be in separate directories that can be accessed only by employees whose jobs require it.

Passwords provide your first line of security. If an application has a password system built into it, always use it in addition to the network operating system's security scheme.

Don't use the same passwords for software applications as you do for logging onto the LAN or getting e-mail. Some

employees might resist; you should insist. Passwords should be five characters or more that create a nonsensical word such as TRGYHJ.

Passwords should be changed at least every six months and should be issued by management only. Users have a tendency to create passwords from things they like or are familiar with, such as a spouse's name. These types of passwords are easily and quickly guessed. Some forethought in picking passwords can go a long way when you consider that a disgruntled employee is more likely than an outsider to breach security.

A form for user passwords can be a single sheet for every LAN and each application. The form should list the application name, directory or path, user's name, function, password, and last date of issue. A six-month review of these forms helps to regularly assess your company's security needs. Unless a single person is responsible for system administration, keep these password forms separate from the rest of the manual.

Security on a LAN also includes giving users selective privileges to read, write, modify, create, or delete files and directories. Access control can be the most complicated—but necessary—aspect of LAN security. If possible, apply passwords to entire subdirectories.

Consider the physical security of your hardware. Like any other office equipment, LAN hardware is usually out in the open. Plan for potential theft by sticking to a good backup procedure. Also, check your insurance policy.

## Backup and Recovery

Network data must be protected with a comprehensive and regular backup schedule. When a desktop system fails, the results are unfortunate. When a LAN fails, the results can be disastrous. Your backup procedures should encompass data on network and local disks.

Backup procedures can be as simple as copying data to floppy disks or as comprehensive as using an off-site storage of tape cartridges. You also need a plan for recovering and restoring data.

Invest in a tape backup unit. To ease the tedious process of backup, most backup software allows for automatic timed and unattended backups. Set it up, go home, the unit performs the backup overnight, and in the morning, you change the tape. Verify backups to assure their accuracy. Make more than one copy, and keep a set of tapes off-site. Even having multiple generations of a backup shouldn't lull you into a false sense of security.

The first form for backup information can be a single sheet for each disk being backed up. You may want to maintain separate forms for each partition. Enumerate how often backups are made (days, weeks, or months), the type of backup (full, partial, or incremental), the person responsible for backing up the system, and where copies are stored. On the rest of the form, list the directories and applications stored on the disk that's being backed up.

A common backup schedule might include a full or incremental backup once a week, rotating four tapes, a different one for each week, with one always kept off-site. If you are making incremental backups, then be sure to include a full and complete backup at least once a quarter.

A second form for backup identifies the local hard disks that contain important data. Requiring users to back up their own hard disks is risky; basically they won't do it. The best strategy is to invest in a utility such as Fresh Technology's (Gilbert, AZ) MapAssist that enables you to access

(and therefore back up) local hard drives. Backing up to floppies is tedious. A smarter way is to back up to a hard disk or to tape.

---

### **Accumulated Knowledge**

A network must be maintained to operate properly and minimize problems. Employees come and go, user rights

increase or decrease, and software is added and deleted. Without documentation, it's impossible to retain the employees' accumulated network knowledge. Assembling a manual and implementing policies helps address administration issues and prepares you to handle problems. By being able to discuss problems with consultants or tech support, you can greatly reduce or avoid costly downtime and service visits. ■





# Training Network Users

## In this report:

Types of Training .....	2
Training Organizations .....	2
Audio Cassette Training .....	2
Training Tools and Consultants .....	3
Video Cassettes .....	3
Disk-Based Tutorials ....	4

## Datapro Summary

Proper network training is essential for the welfare of an organization. Ineffective training poses a threat to the security of a LAN, and becomes a constant burden to those responsible for running the system. There are a variety of friendly and effective options—such as professional training services, videos, and diskette-based training—that can help users achieve an appropriate level of operating proficiency.

It's the busiest time of day on the network. You're trying to track traffic and bandwidth utilization as you contemplate whether to subdivide the network and add another file server. The phone rings, interrupting your calculations. On the other end, a user asks "Can you tell me how to copy a disk, again?"

Does this sound familiar? It does if you are a network administrator supporting dozens or even hundreds of people. Proper training is the solution. To bring users to a respectable level of network and software proficiency, you can use personalized training services and products on videotape, audio cassette, diskette, and in print.

While providing a training program for every network user may seem burdensome, it is essential. Remember: you will not get much network management done spending the bulk of every business day answering the same questions over and over.

In addition, a poorly trained user poses a threat to the safety and security of the LAN. For example, if the network is not thoroughly protected, an unskilled user can accidentally wipe out batch files, command

files, and configuration files and severely drain the technical resources of the company.

Working in a large governmental agency, I confront the need for suitable training daily. While a few power users are at the top of the knowledge pyramid, most users operate on a fairly rudimentary level. Many hesitate to experiment with their computers, and even more are afraid of damaging or permanently erasing valuable data. They timidly approach the operating system and insist on being set up in easy-to-use menuing programs. Fortunately, there are several training methods that can help these timid users gain confidence on the network and free you to concentrate on other tasks.

## In-House Centers

One of the most ambitious approaches you can take is to establish an in-house training center. It could operate as a service of the management information systems (MIS) office, or it could stand as a separate entity altogether. If you delegate the training function to your staff, make sure you assign (or hire) enough people to handle the extra workload. When you and your staff are managing the network, you won't have time to teach.

If you go the other route and set up a separate training department, be certain the instructors maintain close ties with the network administrators. The training lab

This Datapro report is a reprint of "Bringing Users Up to Speed" by Joe Lazarro, pp. 73-78, from *LAN Technology*, Volume 7, Number 1, January 1991. Copyright © 1991 by M&T Publishing, Inc. Reprinted with permission.

should work hand-in-hand with the MIS department. However, be sure to define each group's role in detail to avoid rivalry due to overlapping responsibilities. The training center should perform training only, and leave network management up to the network managers.

Training instructors should sponsor in-house classes on hardware and software, offer courses in field offices, and perhaps call outside consultants and speakers to conduct seminars on various topics. The size of the training center staff is also important, as you will want to minimize costs wherever possible. If you select your teachers carefully, a few people can go a long way.

Don't forget that the teaching staff will require guidance in developing lesson plans and finding source materials. If you need help assembling a training department, turn to Logical Operations Inc. of Rochester, New York. The group provides courses nationwide for trainers or those interested in setting up a corporate training facility. The staff will assist you in customizing courses from the ground up.

Intellisance Corp. of San Jose, CA, also sells textbooks and accompanying sample file disks to help network administrators plan and implement a classroom training program. The \$950 kits of books and disks come bundled in packs of 10.

---

## Types of Training

Once you've decided to train your users, you must determine what form of instruction to adopt. Will you opt for individual mentoring? Classroom instruction? Self-paced methods?

One-on-one training consumes the most worker hours of any arrangement, for you are expending one skilled trainer per employee. This kind of interaction is the most effective form of teaching. Both student and instructor have plenty of time for interaction. The instructor can evaluate the user's needs and tailor the lessons accordingly.

If this method proves too expensive, you can limit costs by training users in a class. Classes should not have more than 20 or 30 students. With more people in a class, the quality and effectiveness of your training may suffer. Some students are more aggressive than others, and the shy ones, who are too timid to ask questions, will fall through the cracks.

You can limit the number of classroom hours by having students prepare before the lessons begin. Assign homework in the form of videocassette courses, printed workbooks, audio cassette materials, or a combination of all the above. You want to pump as much information as possible into your students before they walk into class. This will cut down on rudimentary questions that will bog down the sharper people. Classes should cover the DOS or Macintosh operating system and software applications. The syllabus ought to cover whatever network resources the employees must access, such as electronic mail, word processors, database managers, spreadsheet functions, terminal emulation, and desktop publishing packages.

While you must introduce network survival basics like logging on and transmitting and retrieving files, your emphasis should be on deepening the users' knowledge of their own desktop applications. After all, you are not training users to become members of your technical staff.

---

## Training Organizations

If you lack skilled teachers, look to professional training organizations that come to your company and teach your users. One such organization is Hands On Learning Corp., based in Burlington, MA. In addition to providing on-site training, the company sells a line of instructional videocassettes.

An alternative to educating network users internally is sending them to an outside training outfit. Getting people in a classroom setting frees them to concentrate on network training, isolated from the day-to-day pressures of the job, such as ringing telephones or angry supervisors.

One source for outside instruction is Catapult, formerly Egghead University, based in Bellevue, WA. Teachers conduct training courses for most off-the-shelf applications software on DOS and Macintosh platforms such as dBase, WordPerfect, and Lotus 1-2-3. Courses start at \$195 each and last a full day. Available in most major cities, classes usually consist of 8 to 12 students. Catapult also offers on-site training.

Off-site education has its drawbacks. For instance, the users' training setting will probably not match their exact workstation environments. The computer commands users key in at the training site might be very different from those on the job. Setting up an exact copy of the students' hard disk and network environments is nearly impossible at the training center. As a result, users must figure out how to bridge the gap between what they were taught and how to apply it on their computers.

---

## Audio Cassette Training

A third alternative is self-teaching via audio cassette, videocassette, or disk-based methods. A self-paced method is the right ticket for professionals who don't have the time to attend regularly scheduled classes.

Audio cassette is old, yet it is a reliable and effective technology. Although audio tapes can be used without instructors, cassette tutorials are an inexpensive means of reinforcing class training, especially when the user is away from the job for a brief period.

Audio cassettes are not for everyone, especially for those who are more visually inclined. If you force these users to solely listen to tapes, they may wind up being bored and distracted. However, audio tapes can be played next to the workstation, while the user follows along operating the computer. Audio tutorials can be played in the car or on a Walkman, so the user can remember what was learned at the desktop. The typical tape is fairly low priced, about \$200. Hence, establishing a lending library of these tapes can be extremely cost-effective.

The leading audio cassette media marketer is FlipTrack Learning Systems, of Glen Ellyn, IL, which offers a full family of training courses for the Macintosh and IBM PCs and compatibles. The Flip-Track cassettes are more than recorded books on cassette tape. As you proceed through a course at your computer, you are invited at intervals to turn the tape over to hear in-depth explanations of a given topic. When you are finished with the flip side, you simply rewind the tape to your recent starting point and pick up where you left off. If you do not require the extra help, you continue listening.

For the Macintosh, the FlipTrack courses include: "How to Use the Macintosh Plus and SE," "How to Use the Macintosh II," "How to Use Microsoft Word," "How

## Training Tools and Consultants

**Logical Operations Inc.**  
595 Blossom Road  
Rochester, NY 14610  
(716) 482-7700

Provides courseware and consulting for trainers and LAN administrators. Instructor's starter package and manuals for 12 students is \$900; includes 800 number and upgrades to instructor package.

**Hands On Learning Corp.**  
27 Cambridge Street  
Burlington, MA 01803  
(617) 272-0088

Offers job-site training; costs up to \$12,000.

**Catapult (formerly Egghead University)**  
10900 N.E. 4th Street  
Suite 1350  
Bellevue, WA 98004  
(206) 646-6767

Conducts computer training on Catapult premises and at customer's location, with prices starting at \$195 per course.

**FlipTrack Learning Systems**  
999 Main Street  
Glen Ellyn, IL 60137  
(800) 222-FLIP

Sells audio training tapes for users and LAN administrators; under \$200.

**Intelligence Corp.**  
1885 Lundy Avenue  
San Jose, CA 95131  
(408) 432-0430

Produces video courses for personal computer software; \$295-\$495 per course. Also sells disk software at \$179.95 for single-user packages; \$750 for multiuser versions.

**The Information Factory**  
208 Charter Oaks Circle  
Los Gatos, CA 95030  
(408) 374-1235

Offers videocassettes on basic telecommunications and networking concepts for \$450 each.

**Anderson Soft-Teach**  
2680 N. First Street  
San Jose, CA 95134  
(408) 434-0100

Provides line of videos for DOS applications at \$275 and \$595.

**LEARN P.C.**  
5101 Highway 55  
Minneapolis, MN 55422-5134  
(800) 532-7672

Provides line of video learning tapes for DOS applications for \$295-\$2,900.

**Personal Training Systems**

828 S. Bascom Avenue,  
Suite 100  
San Jose, CA 95128  
(408) 559-8635

Sells audio-and-disk training courses for Macintosh for \$79.95 each.

**Folio Corp.**  
2155 N. Freedom Boulevard  
Suite 150  
Provo, UT 84604  
(800) 543-6546

Sells software tools at \$495 for single-user versions and \$695 for LAN packages. FolioVIEWS can be used to create a custom training package. Company refers you to publishers selling InfoBase.

**Micro Video**  
attn: Travis Huddleson  
91 Fifth Avenue, 6th Floor  
New York, NY 10003  
(212) 255-3108

Offers video instruction at \$595-\$995.

to Use Microsoft Excel," "How to Use Microsoft Works," "How to Use PageMaker," and "How to Use QuarkXPress."

The IBM tutorials feature DOS courses such as "How to Use DOS on a Hard Disk" and "Using MS-DOS 4." Flip-Track also offers courses in WordPerfect, Microsoft Word, WordStar, Display-Write, Multimate, Lotus 1-2-3, Excel, Microsoft Works, Quattro Pro, Symphony, Harvard Graphics, Ventura Publisher, dBase, Paradox, and R:base.

By the time you read this, FlipTrack should have hit the market with a new audio cassette series on local area networks for network managers and would be network managers. The first course, "Planning a Local Area Network," will include a workbook that leads the students step-by-step through the network planning process. After completing the workbook, they will have essential details concerning what type of network is best suited to their organization. The \$189 Flip-Track tape discusses types of wiring schemes, topologies, network interface cards, servers, electronic mail, shared resources, and how to determine whether a LAN is right for the company.

The second and more advanced course is titled "Managing a Local Area Network." This \$195 tutorial thoroughly describes the typical duties of a network administrator. The course covers network operating systems, server and remote personal computer hardware, security, upgrade planning, trouble-shooting, and backup methods,

as well as manager and user training. The audio cassette also describes Novell Inc.'s Netware, 3Com Corp.'s 3+Open, and Banyan Systems Inc.'s VINES network operating systems.

### Video Cassettes

The hottest training tool on the scene is the videocassette. The videocassette is an excellent medium for providing training to a wide audience, and is equally effective in a small group or large classroom setting. It can also be implemented in one-on-one teaching situations or played by the user privately. The videos can be used in a computer lab so students can practice what they learn immediately.

Training companies produce an array of videos for various learning needs. If you need to gird users with the fundamentals of telecommunications or networking, turn to the videos offered by The Information Factory, Los Gatos, CA. Information Factory's tapes are well suited for workers who need the basic hardware and software concepts behind data transmission. The Information Factory distributes three video tapes: "Introduction to Telecommunications," "Introduction to Local Area Networks," and "Introduction to Digital Communications." These courses are not software specific, but provide a solid foundation on these three key building block technologies. Students can watch the videos before attending application training classes.

Intelligence sells courses on hardware for IBM PCs and compatibles, the DOS operating system, and numerous software applications.

Anderson Soft-Teach of San Jose, CA, sells a complete line of video cassettes for the DOS world. The tapes are packaged with workbooks and sample file disks, making it easy to follow along and learn on your PC. The guide tracks the video with step-by-step instructions and also gives computing tips and techniques. It can also serve as a handy reference. Both workbook and training disk give the users the opportunity to perform actual tasks on the computer with their newfound skills.

The current video library includes a beginners' guide to personal computing, and introductions to local area networks, MS DOS, Lotus 1-2-3, WordPerfect, dBase, Microsoft Windows, Microsoft Word, and DisplayWrite. Two-course sets cost \$595; the one-tape LAN course costs \$275. The company will soon roll out several titles in Open Caption format, with subtitles of the spoken portion, giving the deaf full access to the material.

LEARN P.C. of Minneapolis, MN, has unveiled a line of video training packages for both IBM PC and compatible and Macintosh platforms. PC-based courses include WordPerfect, Lotus 1-2-3, dBase, Computer Fundamentals, Unix Programming, C Programming, Microsoft Word, Enable, DisplayWrite, Microsoft Windows, OS/2 and Paradox. For Macintosh users, the company sells courses on Hypercard, Excel, Pagemaker, and Microsoft Word.

LEARN P.C.'s videos come with workbooks and practice disks. Prices range from \$295 for single cassettes to \$2,900 for multivolume packages. All training videos are closed captioned for the hearing impaired.

Micro Video of New York produces tapes that display on a television monitor what you'd see on a computer screen. Courses include Lotus 1-2-3, WordPerfect, Paradox, Microsoft Excel, dBase, MS-DOS, Harvard Graphics, PC Primer, and R:base. Tape prices run from \$595 to \$995, and course length runs from two to five hours, depending on the complexity of the application.

---

## Disk-Based Tutorials

The newest media for computer educators is the disk-based tutorial. A disk trainer simulates an actual applications program, allowing the user to operate the software in an environment that looks and feels like the real thing and is saturated with help. This type of training assumes that the novice has some computer skills. A new user may not be comfortable with this technique and feel more at home with class training, either group or one-on-one.

If disk training is appropriate for your situation, Personal Training Systems of San Jose, CA, delivers Macintosh courses that combine disk and audio cassette technologies. Students listen to an instructor on tape as they sit at their Macintoshes and run lesson files on a disk. Courses include basic Macintosh operation, Microsoft Word, Persuasion, Microsoft Excel, Hypercard, PageMaker, Microsoft Works, Adobe Illustrator 88, FileMaker II, FileMaker Pro, Aldus Free-Hand, and QuarkXPress. While most courses are available in four levels, you can purchase the disk/tape package one module at a time. Programs start

at \$79.95 each and include a command summary card. Each lesson lasts approximately 90 minutes.

For the DOS arena, Intelligence markets disk-based training courses on WordPerfect, Microsoft Word, DisplayWrite, Lotus 1-2-3, and dBase. Unlike other programs, Intelligence tutorials are memory resident, allowing you to flip screens between the actual software application and the training session. The learning material contains both text and graphics and features valuable interactive question and answer sessions. A single-user version is priced at \$179.95, while the tutorials operable on LANs start at \$750.

If your users need more detail than the Intelligence tutorials provide, you can create your own training software using the InfoBase toolkit from Folio Corp. of Provo, Utah. The FolioVIEWS InfoBase software environment provides you with a searchable, Hypertext-like text retrieval system, capable of displaying screen after screen of explanations. For example, you can set up a word processing training program, and if users are confused about a particular command description, they can search for any target word or phrase within the explanatory text to obtain more detail. Users can print any selection, and copy part or all of any InfoBase disk to a hard or floppy disk. They can also switch back and forth from their application to the InfoBase.

Published Folio InfoBases on existing applications are sold by independent companies, but you can call Folio directly for sources. InfoBases are available on WordPerfect, WordPerfect Office, WordPerfect Executive, PlanPerfect, DataPerfect, DrawPerfect, LAN Times, LANDEX, Novell Netware Buyers Guide, McGraw-Hill's Datapro Master Index, Novell Compatible Products And Service Directory, Novell Netware, Open Access III, PC Glossary, WordPerfect Compatible Products and Service Directory, and other software.

---

## More Training Options

As computers continue to proliferate in the workplace, the need for classroom training and personalized instruction will only expand. Vendors will be cranking out products integrating windows-oriented operating environments and graphics to make their software and hardware easy to use. Yet no matter how friendly computers strive to become, there will always be a need for a competent and friendly teacher at the user's elbow.

The tutorial methods discussed here are likely to increase in popularity and improve in quality. However, even more fascinating training techniques are on the drawing board. You can expect to see several disk-based tutorials, incorporating more exciting interactive text and graphics. CD-ROM technology will become commonplace, enabling users to carry training material previously loaded on more than five floppies on a single compact disk.

As a network manager, you already spend a small fortune on system hardware and software. Be willing to add the extra capital for training to your overburdened budget. Organizations that train their workers effectively will be able to compete more strategically within their markets. ■

---

# LAN Training Sources

---

## In this report:

What the Vendors Offer .....	2
Distributors and Resellers ..	2
Independent Trainers.....	3
Which Is Best? .....	3
Who to Send.....	4

## Datapro Summary

To keep pace with the fast-moving networking industry, formal training is invaluable for LAN managers, administrators, and users. Fortunately, LAN training is available from a wide variety of sources, including systems integrators, vendors, resellers and distributors, and authorized training centers. Those seeking formal training would do well to select a source that offers plenty of "hands-on" experience.

Intelligent, articulate people, not wholly unacquainted with technology, ruefully admit that they have never successfully programmed their VCRs to record a show at a later time. Ignorance is not bliss for network managers and administrators either. Not only must they keep their LANs up and running, they must install and remove users, change user privileges, load applications, and answer the plethora of questions that users have about the network.

Where do you get the training to handle this multiplicity of jobs? Training is available from several sources, including vendors and their authorized training centers, independent training centers, integrators and resellers, videotapes, and books.

In principle, everyone recognizes the value of training. Technical training allows people to use equipment more successfully. Imagine how much more satisfactory your VCR would seem if you could program it. Training provides benefits for users, network administrators, and vendors. Users are happier with equipment that they can use successfully—they look better to their bosses. Administrators are happier because

they are in better control of the network, which makes both the users and their bosses happier. And vendors are happier because their products look better, which increases the chances for further sales. But what is the source for training network administrators and end users?

---

## Myriad Training Sources

Hardware and software vendors, resellers, and independent training organizations offer training programs on how to install, manage, and use network products. It is also possible to receive training at some two- and four-year colleges. Other means of training are available as well, such as computer-aided tutorials and training videotapes. Programmed instruction, popular in the 1960s and 1970s, is a methodology that reduces information to small steps. Students read material, immediately answer a question about the material, then check the correctness of the response. This learn-respond-verify technique reinforces the learning process. It is also deadly dull.

Computerized tutorials, programmed instruction, and training videotapes all lack an important ingredient in the learning process. Diana Berry, information systems manager for Harding Lawson Associates, an environmental services company in Novato, CA, says the best way to learn is through hands-on activity. "Classwork doesn't cover what you're going to run into," Berry says. "It tells you how to install

---

This Datapro report is a reprint of "A Trainee's First Steps" by Lindsey Vereen, pp. 131, 133, 135, 137-138, 140, from *LAN The Local Area Network Magazine*, Volume 6, Number 10, October 1991. Copyright © 1991 by Miller Freeman Publications, Inc. Reprinted with permission.

a user. But it doesn't cover what you do if something doesn't work right." You can attribute the same drawback to computer-aided tutorials and programmed instruction methodologies. Nothing can take the place of roll-up-your-sleeves training on the hardware itself.

Hardware experience means more than sitting at the keyboard of a computer. Doug Wilmsmeyer, chief operating officer of Wave Technologies, a training company in St. Louis, gives the reason. "No matter what anyone says, administrators themselves are usually responsible for assembling the hardware."

### What the Vendors Offer

The network operating system vendors have well-established training programs for systems administrators, support personnel, and end users. Vendors also require distributors, resellers, and integrators to take training to gain certification.

Novell (Provo, UT), with the largest share of the LAN market, also has the most highly evolved training program. Novell Education Centers (NECs) located around the country offer advanced training, reseller training, internal training, and instructor training. However, Novell does not train end users at these centers.

End-user training takes place at Novell Authorized Education Centers (NAECs). Novell has established the NAEC program for third-party companies to train NetWare users. Participants include distributors, resellers, and independent training centers at more than 400 locations. Novell instructors develop "beta" versions of new courses at the NECs, according to Jerry Christiansen, Novell's NAEC program manager. When the "courseware" is developed, Novell distributes it to the NAECs. NAEC instructors earn certification through the Certified NetWare Instructor (CNI) program, which includes "train the trainer" classes at NECs. Novell administers its Certified NetWare Engineer (CNE) program through the NAECs. This training requires a 14-day commitment and comprises six classes.

Banyan Systems (Westboro, MA) provides training programs at five sites in North America. The Banyan training programs lead to two levels of certification, the Certified Banyan Specialist (CBS) and the Certified Banyan Engineer (CBE). As with the CNE program, CBS and CBE certification requires that candidates pass a test to demonstrate their proficiency in course material.

The CBS program is a stepping stone to the premier CBE program. CBEs must be recertified each year. Twice a year, Banyan publishes a list of requirements for recertification. Banyan also supports independent training centers, similar to Novell's NAEC program. About 70 Banyan-certified education centers are located around the country.

Microsoft's training arm, called Microsoft University, is part of Microsoft's consulting services. Microsoft University offers LAN Manager training in Redmond, WA, where the company is based, and at nine training centers in the United States and Canada. Microsoft trains end users primarily, but it also has a reseller authorization program.

Microsoft resellers, or *network specialists* are required to take 80 hours of training. After 40 hours, they are provisionally authorized to sell LAN Manager. Within six months, they must take an additional 40 hours, according to Steve Kanzler, national marketing manager for Microsoft's network business unit.

In addition, Microsoft has authorized outside training organizations to provide network training much as Novell

has. Microsoft has established nearly 46 Network Training Centers (NTCs) and Advanced Network Training Centers (ANTCs). The NTCs teach basic classes only; the ANTCs teach basic and advanced courses. Instructors are required to attend Microsoft's "train the trainer" courses. They also have to take the course they will be teaching.

Microsoft does not yet have the kind of certification program that Novell and Banyan offer, but according to Kanzler, the company is in the process of establishing one.

### Distributors and Resellers

An important link in network training, distributors, resellers, and integrators also provide training services to end users. They can offer more diversity in their training programs than vendors. Whereas the manufacturers only offer training on their specific products, companies in the distribution channel can offer a breadth of courses. Integrators, in particular, are intimately familiar with your network design and therefore can tailor courses to your company's needs. The tradeoff is that distributors and integrators generally train only on the products that they are authorized to sell, which limits your training options.

Resellers and distributors may offer greater flexibility in training options. For example, Westcon, an Eastchester, NY distributor and NAEC, offers "advanced" Novell courses in LANalyzer and NetWare SNA connectivity; SynOptics (Santa Clara, CA) network management, configuration, and installation courses; and Eicon Technology (New York) WAN courses, in addition to the standard fare of Novell authorized education courses.

A reseller or distributor often offers alternatives to the formal CNE training. For example, at Westcon, one option is for students to take the classes at night over a period of time, much like night school. Another accommodates "people who through life experience are pretty much CNEs already," says Bruce Hanson, Westcon's vice president of training and education. They attend an accelerated five-day program in preparation for the certification test.

Using your network integrator as your personal trainer has one major advantage—they have intimate knowledge of your network configuration. This knowledge allows them to give customized training. They tailor the classes to the exact configuration of the installed system.

David Berman, executive vice president of Data Systems Marketing, an integrator in Upper Marlboro, MD, points out that "the more information the trainer has about the network, the more useful the training [will be]. We have also designed and installed the network. We have a better understanding of what the company is going to do with it. If we have sold them bridges and routers, we try to customize the class."

Some resellers and integrators are also authorized training centers for some of the products that they resell and install. International Micronet Systems (San Francisco) is authorized to sell Novell, Microsoft, and Banyan products, but only offers training on Banyan networks, according to vice president Elliott Chuang. Like other resellers, International Micronet tailors training to the environment, says Chuang. The company also offers the certified Banyan engineer program.

Resellers offer more diversity than vendors by providing training for network administrators on equipment and software from various several sources. However, keep in mind that training is usually tied to the products that they sell.

## Independent Trainers

Independent training companies and consultants are also sources for network training. Offerings can range from a few courses to an extensive program. Independent trainers can provide a broader range of training than the resellers, and they have the advantage of being less tied to products than either resellers or vendors. Their advantage is their independence; their revenues do not depend on selling a particular product line or service. However, they might not be certified by the manufacturer and might have less access to vendor resources.

Independent trainers often offer courses and coursework that you can't gain through training run by manufacturers or resellers. For example, Wave Technologies offers classes for NetWare, LAN Server, LAN Manager, and VINES, although the bulk of its training services are on NetWare. Although not an NAEC, Wave offers "the full spectrum of courses that will allow students to become Certified NetWare Engineers," says Wave's Wilmsmeyer.

Hands-on training is an important part of Wave training. "Since students are taking on more hardware and software responsibilities," says Wilmsmeyer, "all students go through a troubleshooting methodology. We teach them a strategy for buying spares. We teach them how to isolate and replicate problems. The instructor will create problems in class, such as physically cut a cable and show the errors that will occur."

Independent training companies "don't have to sell propaganda," says Wilmsmeyer. They don't have to show just a single vendor's solutions; they can show alternatives as well. Wave's courses focus more on business solutions than on the products of a single vendor, according to Wilmsmeyer.

The advantage of going through a class that is vendor-independent is that "you may hear things that [the vendor] won't tell you," says Michael Colby, vice president of conferences and expositions at Digital Consulting (Andover, MA). Digital Consulting runs conferences and seminars and offers introductory and advanced level NetWare courses for system administrators. The four-day classes are given by independent consultants three times a year in major cities around the country.

Because the focus of training is often on a specific LAN, internetworking issues are sometimes given short shrift. Interoperability training is often theoretical rather than hands-on, particularly when a vendor offers it. Novell's Christiansen says that Novell offers "a data-communication course that discusses internetworking issues. It's a lecture course." Because an internetwork is built of so many manufacturers' products, independent training companies may have an advantage here. Wave offers "training on how to integrate internetworks," says Wilmsmeyer. "We try to get students to look at internetworks, bridges, and routers. We are not restricted by a distributor license."

Some, but not all, independent training centers are vendor-certified. New Horizon (Santa Ana, CA), which is a Novell NAEC and a Microsoft ANTC, offers NetWare, LAN Manager, LANtastic, and Macintosh networking classes, according to account executive Elise Abrego. An independent training company "can train anyone—a reseller or end user," she says. "We cater to everybody. We train people on what they want."

Since instructors at independent training companies aren't necessarily qualified by network vendors, the chances are greater for varying quality than with reseller or vendor training.

You should carefully check the reputation of the training company. Ask for former student references, then check them out. Contact them and ask pointed questions, such as, "Did you learn enough to be effective at your job?" Find out how much hands-on experience you get. More than half of the class should be hands-on. "You're going for the experience, not the education," says Wilmsmeyer. Ask what guarantees the training company offers. Find out what your options are after you finish the course. For example, can you go for certification? Can you go get support after you complete the course?

Check the qualifications of the instructor, such as vendor certification. Find out if the instructor is a part-time or contract employee. An instructor who is a consultant may have other interests, such as selling you consulting services.

## Other Training Sources

If your training budget is lean or if you can't spare time away from the office to learn, one possibility for training on a shoestring is through colleges and universities, an increasing number of which are offering network training.

Colleges approach education differently. Classes tend to extend over a longer period of time. For example, completing a NetWare class may take six weeks, two nights a week. It has the advantage of being inexpensive, and you can take the training after normal business hours.

Lastly, under certain circumstances video training programs or other programmed-instruction resources may serve your purposes. Although they offer no hardware interaction and are often tedious, they are useful if you can't get away for training.

## Which Is Best?

Vendor and vendor-authorized training programs offer the most up-to-date information on a particular product, but they tend to focus on their own products and solutions. Resellers offer more variety and cross-vendor training. They also offer the advantage of being able to tailor their training to your exact configuration. But they, too, focus on products they sell. Independent training companies usually have fewer ties to vendors and can offer a more independent perspective, but they may not have the certification that you will find from vendors and resellers. And they may not get information from vendors as quickly as authorized resellers. Colleges offer (usually inexpensive) training over a longer period of time, which can be good or bad.

Whatever source you choose, hands-on training is vital for learning. The program should offer at least 50 percent hands-on work. The ratio of students to computers should be as close to 1-to-1 as possible. The class size should be small, from four to 10 people.

No matter where you go, you should find a program that matches your specific needs as closely as possible. It should be tailored to your network configuration, it should give you the information you need to do your job, and it should offer the certification that you need.

---

## Who to Send

Training programs offered by vendors, resellers, and independent trainers cost about \$300 per day. Then factor in travel expenses—your employees' expenses if they're trained offsite—or the trainers' expenses if your workers will be trained onsite. Training classes usually last from three to five days, although programs leading to certification take longer.

A common strategy is to send one person to be trained with the expectation that that person will train others within the company. A better strategy is to send several employees through network training. The functional minimum for training is "a departmental [LAN] manager who has been entrusted to put in a LAN and a backup for that person," Wilmsmeyer says. This way, if the primary administrator is sick or busy, a backup person can still manage the network.

---

## Where to Do It?

No matter whether you receive training from a vendor, a reseller, or an independent trainer, you have the option of training at your facility, the trainer's, or sometimes at a rented facility. Trainers differ in their opinions as to which is best. It depends on the people being trained, says Berman of Data Systems Marketing. Some people being trained away from their own facility act like they are on vacation. On the other hand, some people being trained at their facility are inhibited by the presence of their supervisors in front of whom they don't want to ask foolish questions.

Trainers acknowledge that training away from a user's facility has the advantage of freeing students from interruptions such as unwanted phone calls and the day-to-day press of work. Westcon's Hanson cautions that you can't do training on a live, functioning network. This consideration may limit the ability to do training at the customer site.

Training is less expensive at the customer site, says Wilmsmeyer. However, students get interrupted at their own site, and they lose productivity in the training program. Wilmsmeyer offers one advantage of having a homogeneous class at the customer site: "You can discuss specific sensitive problems at the customer site," which you can't do at a training facility in a class made up of people from several companies.

Students often learn better away from their normal workplace. They have the opportunity to interact with people from other companies. However, if the problems can be overcome, inhouse training has the advantage of being less expensive, and it can take place around people's work schedule. For example, training can take place during the evening or on weekends. It also offers the opportunity to train the maximum number of people.

Regardless of the training source, technical training usually does not involve testing in the way you experienced it in school. It's hands-on; it's immediately useable. Most network training programs offer a way for you to measure what you have learned, but this information does not find its way back to your supervisor's desk. The certification programs differ in that they do involve testing, but are worthwhile because being certified enables you to claim high salaries on the open market. ■



---

# LAN Management Issues

---

## Datapro Summary

Local area networks have come a long way in the past few years. Traditionally a mechanism for sharing printers and other peripheral devices, LANs have evolved into an integral part of a company's day-to-day operations and long-term business strategies. With LANs taking on a more critical role, network management presents a morass of challenges and difficulties. LAN development and management now requires the attention of top-level corporate management.

---

## LAN Challenges

LAN environments run the gamut from stand-alone departmental networks to geographically dispersed LANs linked together via an Ethernet backbone or a wide area token ring. These wide area networks are often hooked back into a remote data center. But for all the benefits of using these higher forms of LANs instead of mainframes—increased processing efficiencies, the distribution of information to all levels in the organization and reduced long-term costs—CIOs complain that they present the same problems that mainframes caused 10 years ago but that they lack equivalent management mechanisms. Moreover, delegating responsibility for the LANs may require a restructuring of the IS organization and procedures. Delivering corporate applications and data in a distributed environment demands corporate control and support to establish standards for hardware and software, operating procedures, application development and end-user services.

“Managing LANs, in essence, is like managing a large computer system that just

happens to be in lots of places,” said Jay Hamann, vice president of MIS at Schreiber Foods Inc., a cheese processor in Green Bay, Wis. “It’s important that we understand [that LAN environments] are like a mainframe and provide the necessary management structure around them.”

That isn’t to say that CIOs long for a return to a mainframe-controlled IS world. They are simply being realistic about the challenges facing them in the LAN environment. Matthew Cain, a senior research analyst in desktop computing services at the META Group in Westport, Conn., said that “as IS tries to corral the LANs around the organization, they bring to [the process] a centralized perspective, which essentially is a mainframe perspective. . . . They are looking for the same [management controls] that they have in the mainframe environment, which I think is totally valid.”

## Problem Management

CIOs lament that the dearth of comprehensive tools for implementing, monitoring and troubleshooting LANs is at the root of most LAN management problems. “I have a bundle of concerns, and any one of them is bad enough in itself,” said Marvin Ehlers, vice president of information systems at MidCon Corp., a wholly owned subsidiary of Occidental Petroleum based in Lombard, Ill., that operates an interstate natural gas pipeline. “LANs are still an adolescent form of technology.”

---

This Datapro report is a reprint of “The Lay of the LANs—Micro Evolution” by Leslie Goff, pp. 72-82, from *CIO*, Volume 4, Number 8, May 1991. Copyright © 1991 by CIO Publishing, Inc. Reprinted with permission.

Troubleshooting on a geographically distributed network becomes a cat-and-mouse game between the support person and the hundreds of bridges, routers and other devices on the network. Monitoring usage for chargeback purposes is nearly impossible. Network security and data backup are ad hoc efforts. Some users are waiting in the wings for tools to address these problems, making do the best they can in the meantime, while others simply cannot wait.

For Stephen Yates, the CIO at Brown & Root Inc., a Houston-based engineering and construction firm whose credits include designing and building the first NutraSweet plant as well as the largest oil refinery in Kuwait, troubleshooting the 65 LANs his company operates around the world is a top concern. Both of the company's mission-critical applications—computer-aided design and the corporate financial system—run on the LANs, which are tied into an Ethernet backbone. "If a LAN goes down, we're out of business," he said. "When we have network problems, I get a lot of complaints."

Yates said that one hour of downtime is considered major. It usually takes 10 to 20 minutes to identify and work around most network outages. Currently, Brown & Root is using concentrators from Synoptics Communications Inc. of Santa Clara, Calif., to troubleshoot the LANs because the concentrators run over existing phone lines, Yates said.

"[The concentrators] have some pretty good smarts about saying, 'This device is the problem'" on a LAN, he added, but the problems are not so clear-cut when they involve a router—a network traffic cop—or a gateway into the wide area network. "All of those boxes are made by different companies with different technologies. If one takes a hiccup, it's hard to tell where it is."

The sheer number of vendors in the LAN management market is a testament to the difficulties of troubleshooting large networks. MidCon's Ehlers said he had identified between 25 and 50 vendors whose products are used on the company's LANs and are the potential sources of problems.

"It's difficult to find the source of the problem and fix it in the time needed today," he said. "We're kind of our own worst enemy because we're trying to help users and hook them up to the networks, but the more complicated the network is, the more difficult it is to manage."

## Resource Management

Calculating usage chargeback is another mainframe ability CIOs say they would like to see on LANs. On a LAN, the server acts essentially as a small mainframe. The difference is that mainframe tools can monitor how quickly resources such as disk space and memory are being used up. Such metrics are not yet available for LAN servers, and "if you can't measure it, you can't manage it," Ehlers said. When the resources run out, however, users clamor for more.

"The problem is, how do we pay for this?" Yates asked. His answer: User departments rent their equipment from Brown & Root's IS organization. "That way, if they want to keep growing, we just put in bigger boxes at bigger [fees]."

## LAN Support

Until better LAN tools are available. CIOs can overcome the inefficiencies of the LAN environment by fostering a dedicated LAN support group comprising individuals from across the organization. "With traditional IS support

## CIO Distributed Computing Survey

LAN Management Issues	Agree (%)	Disagree (%)
The departments using LANs do not make full use of their capability	82	13
I am looking for new ways to leverage the power of our LANs	77	19
The role of the LAN manager is becoming more pivotal to the strategic direction of computing in our company	73	25
Troubleshooting LANs off our WANs is difficult	63	30
Creating effective, easily used interfaces between LANs and corporatwide systems is difficult	59	35
Setting and administrating responsibility for LANs is difficult	57	40
The most senior LAN manager reports to me	54	41
It is difficult to establish and enforce satisfactory standards	52	45
Information kept on departmental LAN servers is difficult to access, redundant or inaccurate	44	53
LANs are too slow for the WAN technology I wish to use	33	62
LANs created by users are incompatible with corporate standards	30	65
LANs frequently crash	27	70
I am working toward transferring control of LANs from end users to the IS department	24	71
I am concerned about finding an optimal role for the CIO in LAN management	22	75
The departments using LANs are putting too much of a computing burden on the servers	20	75
LANs are too slow for the WAN technology we now use	15	84

you needed depth of knowledge," said the META Group's Cain. "For LANs you need breadth of knowledge, . . . and you rarely find that in organizations today." Current LAN support strategies are often piecemeal efforts with LAN responsibility divided among various support groups. For instance, at Burroughs Wellcome, a pharmaceuticals outfit based in Research Triangle Park, N.C., LAN responsibility is split up among hardware maintenance, software support, application development and long-term network development. Each of these specialized groups reports to central IS, according to Jonathan Petersen, network services administrator.

But it can be difficult to "strategically plan for LANs when the responsibility is spread so thinly throughout the organization," Cain noted. He expects a move toward interdisciplinary LAN support teams made up of cross-trained personnel pulled together from such areas as mainframe application development, microcomputer support,

data communications and end-user business units. (Cross-training involves putting individuals with various skills into one group where they are trained in each other's disciplines.) Petersen said Burroughs Wellcome is evaluating cross-training personnel and restructuring its LAN support under one umbrella. "I think a dedicated LAN support group would be better," he said. "The more we feel we're part of a team, the more we'll play together."

IVI Travel, a national corporate travel agency based in Northbrook, Ill., is cross-training office managers and microcomputer support personnel to support future LAN development. The agency recently began tying its 120 offices into its corporate minicomputer so travel agents can access a central database of hotel availability and client rates.

Barry Rogers, IVI's vice president information resources, said that most of IVI's offices have LANs that link via gateways into one of the airline reservation systems (Sabre, Apollo or PARS); many of the larger offices have such links into more than one system. A LAN administrator—usually the IVI client service manager—is trained by the airlines to configure and troubleshoot the LAN and serves as the first line of defense for users. Within the IS organization, a Micro Systems department handles LAN application development and problems that are too big for the LAN administrator.

"We have a person in [Micro Systems] who is receiving very technical training in LAN management" and who will act as a go-between for the LAN administrator and IS, Rogers said. "I anticipate we'll have to expand that function as we expand our installations."

As the company links the individual office LANs into the corporate mini, Rogers' staff and the LAN administrators will be able to coordinate and integrate their activities. But in most companies end users are spending 10 to 15 percent of their time on routine departmental LAN management and administration because most companies lack "the commitment to train people and bring them up through the organization," according to Todd Dages, director of communications research and consulting at the Yankee Group Inc., a high-tech market research firm in Boston. For example, a network manager in a large manufacturing concern with international LANs noted that the company "does not have time to do much cross-training." Although cross-training makes logical sense, if not managed properly it can ignite rivalries within the IS organization. Cain said that over time, however, resisters will "see that [cross-training] will extend their careers and make them more valuable to the organization."

Petersen of Burroughs Wellcome said the degree of specialization of its myriad support groups has resulted in past turf battles. Reorganizing into a dedicated LAN support group, he believes, "should make them feel like they are on their own turf—not someone else's."

Like any other technology, LANs will continue to evolve in sophistication and capacity. CIOs' LAN savvy will evolve in a parallel fashion, but in the meantime, networks will fail, troubleshooting will be tricky and support organizations will struggle to put the right people in the right areas. "It takes a fair amount of learning if you're going to use [information technology] for anything useful," observed Schreiber Foods' Hamann. "The same mentality goes into saying, 'I'm not going to use my mainframe anymore; I'm going to use my LAN.' Well it may be a little easier, but really you still have many of the same problems."

## The LAN Toolkit

### While Individual Products Do Exist, Information Executives Long for the Integrated Package That Will Do It All

CIOs complain that most of their LAN management concerns stem from the paucity of comprehensive tools to troubleshoot, monitor, configure and measure LAN systems.

Tools that will address key areas of LAN management are coming, however. Todd Dages, director of communications research and consulting for the Yankee Group Inc., a high-tech market-research firm in Boston, said users have a "hierarchy of needs" that includes problem management, configuration management, performance management, security and accounting (measuring usage). Vendors will meet these needs as customers advance through each stage of LAN management.

"Users are asking loudest for products at the lowest level of the hierarchy," Dages said. "Once they get problem management under control, they'll start worrying about configuration management" and on up through the hierarchy. "And that is the way vendors develop the products: They build what users are going to buy."

Ultimately, these individual pieces of the puzzle will be integrated into the major LAN operating systems, where

they will act as comprehensive intelligent hubs. For some users, however, the needs are too urgent to wait.

Dale Terrell, executive vice president of information systems at Security Pacific Automation Co. in Los Angeles, the data processing subsidiary of Security Pacific Corp., oversees some 200 LANs in five states supporting customer service, sales and marketing, bank branches, software development and an executive information system.

"We had to design and build a few remote network management tools ourselves because the ones on the market are not comprehensive enough or are too expensive," Terrell said.

With their custom-built tools, Terrell's staff can monitor LAN performance 24 hours a day from one facility. The company also developed a system to centrally control and distribute software releases.

Security Pacific's business needs were such that it was critical to develop tools in-house rather than to wait for the release of effective commercial products. "We can't put much on the back burner waiting for tools," Terrell explained. "There is always a business case strong enough to move ahead."



# Establishing a LAN Dossier

## In this report:

System Services and Applications.....	4
System Operations Guides .....	4
Logs for Management Effectiveness.....	4
Vendor, Supplier, and Purchase Information.....	5
User Directory .....	5

## Datapro Summary

A LAN dossier will help a network manager completely document the network. Compiling the dossier is time-consuming, but once completed, this data will provide many benefits. It will help LAN managers monitor performance, prepare budgeting information, locate and resolve problems, and provide information to help the decision-making process.

As a network manager, one of your vital goals is to document your network completely, in order to improve LAN management and your accountability for the network. You and your staff should be able to find important technical data for repairs, updates, and trouble-shooting—and to explain your actions to management. In effect, you need to compile a dossier on each piece of the network.

Proper LAN hardware and software documentation goes beyond having a list of your users and workstations. Almost any network with more than a dozen machines is constructed of a multitude of products from a variety of vendors and suppliers. Multiple protocols, different network operating systems, peer-to-peer LANs, centralized servers, various physical topologies, organizational differences, and the enormous variety of applications make creating a dossier a daunting task.

## Documentation Goals

The data should be collected in light of the strategic goals and tactical operations of your organization. Often, senior management is more concerned with the unseen components of a network. They may be

aware of the size of the investment in the system, how many workstations are installed, or how many cities are connected. But they may be more interested in other aspects of the system that are less visible than the hardware and software. To top management the important questions may be: How does the network further the goals of the organization? How secure are the organization's information assets? Is the network worth the expense? With so much at stake, they also may want to know if you've been a good steward of the system. Documentation should be available that supports your decisions and actions, justifies your requests, and affirms your conclusions.

Your LAN documentation should help you assemble and prepare budgeting and forecasting data. Having records of software licenses, configuration data, the number of users supported, and quantities and types of machines installed will be valuable information when the time comes to calculate maintenance, support, and upgrade costs. You can use this data to compare the current installation with your overall plan. With it, you can determine proper staffing levels, establish operating budgets, and formulate productivity measurements for your system personnel.

Network documentation serves as a critical reference resource to your network operations, technical, and support staffs. Well-designed network documentation acts as a foundation for a change management

This Datapro report is a reprint of "Establishing a LAN Dossier" by Glenn Pritchett, pp. 39, 40, 42-44, 46, and 48, from *LAN Technology*, November 1991. Copyright © 1991 by M&T Publishing, Inc. Reprinted with permission.

## Your Mission Statement: The Network Charter

As time goes by, people change jobs and new employees enter the organization, memories tend to fade, and months or even years later questions often arise asking why things were done the way they were and why various alternatives weren't considered.

A network charter or mission statement explains the purpose and scope of the network and why management approved it. It can also discuss what is planned for the network.

### System Overview

Within the network charter there should be a section for senior management and users to get a capsule overview of the LAN. This summary can explain what capabilities your system offers and answer many questions before they have a chance to become issues. In many cases, people only see three or four workstations in their department and are aware of a server or equipment room somewhere in their facility. But they may not know about the dozens of other workstations in other parts of the company, the communications links with other sites, and the different applications involved. This information helps individuals appreciate more fully the nature and capabilities of the LAN. It can

help them to plan for network-related services and requirements.

Internal memos, architectural and design issues, and other documents should be included in the network charter if they provide justifications for design decisions and network policies and procedures. Flowcharts, maps, and diagrams can be included to provide a capsule view of your system in a physical sense. The network charter should have a summary of the physical installation and the functionality offered by the system.

In sum, the system overview portion of the network charter should include:

- design philosophy
- technologies employed and alternatives that were considered
- number of users and workstations
- number of servers
- locations of servers
- services provided on the system
- functions offered users
- standard workstation configurations
- server configurations
- network diagram or map

### Department Structure

The network charter should clarify the roles and responsibilities of both the user community and the network

support team. It should clearly state who has responsibilities for such items as assigning and approving logins, authorizing purchases of equipment and supplies, reporting technical problems, and approving new system functions or services. It could even include job descriptions for your support staff to explain what "all those techies" are doing all the time.

### Policies and Procedures

System-related policies should be publicized to all users and management. Policies may pertain to individuals or classes of users, department-level LANs, or enterprise-wide matters. In some cases, the policies you establish may have legal ramifications and should be reviewed by your corporate legal staff, particularly when involving personnel matters or employee rights to privacy. Policy information can include:

- software licensing and copyright policy
- E-mail privacy policy
- disaster recovery
- information security policy
- data storage allotments and charges
- service and support charge-backs
- theft or abuse of information or equipment
- remote access policy
- maintenance schedules and operational hours

Procedures developed for your network should be included to inform various groups of the ground rules for the system's operations. Procedures developed in most organizations include:

- how logins and passwords are assigned
- how authorizations of rights and permissions are made or altered
- how to get system support or access to support staff during special projects outside office hours
- how to get additional help with problems
- how to get new programs or functions installed on the network
- how to get additional workstations or specialized equipment

### Service-Level Agreements

If your network supports several mission-critical applications service-level agreements may be developed between your support organization and the user group or department being serviced. They should specify in writing what priority is assigned to a particular department or function when multiple departments require simultaneous technical support or troubleshooting.

Service-level agreements further help to clarify the roles and responsibilities each group assumes, the hours that the system will be available either staffed or unstaffed, and how much support of what nature can be expected by the parties involved.

### Summary of Resources

Providing your users with a summary of additional resources available to them will greatly relieve your support staff of repetitive and time-consuming questions. Additional resources may take many forms but can include such items as:

system, giving you a better handle on system changes, upgrades, and other enhancements. Technical documentation improves overall network integrity because problems get researched faster and resolved more rapidly. User confidence in the network and your support staff rises because users suffer fewer and shorter periods of downtime. If in a major disaster, large portions of your system are completely destroyed or damaged, your documentation will be the key element in reconstructing or repairing your network.

Like any good reference material, your LAN documentation can be an excellent training tool, providing a first-level introduction to the network. As a guide for users, it can indicate the nature and extent of LAN services. It can explain how to reach your support staff as well as the procedures involved to obtain server logins or gain access to different applications. When appropriate, such documentation can provide users with a look at your service, support, and backup logs, letting them see that you and your staff are performing your duties in a conscientious and professional manner.

- computer-based training (CBT) materials
- training manuals
- magazine and information service subscriptions
- electronic and written tutorials
- schedules of training programs and classes offered in-house or from third-party organizations
- help-seek services
- resource libraries
- consultant directories
- user groups and bulletin boards

#### Forms

Forms are the easiest way to handle user requests. LAN-related forms described in your network charter can include:

- service and support request forms
- requests to add, change, or delete user logins and profiles
- equipment purchase order forms
- request for program modifications
- error or anomaly report

#### Wiring and Cabling Guide

Wiring and cabling guides are a critical part of your system documentation for use by your technical staff. Precise records of cable construction, locations, and naming conventions are essential in problem resolution and make for improved efficiency in operations. Items to include are:

- network cable schematics
- custom cable configurations for printers, modems, and other equipment
- cable suppliers, part numbers, and specs

- floor plans and building wiring diagrams indicating node locations
- naming conventions for all aspects of the LAN
- server and rack equipment wiring diagrams
- hub and wiring closet diagrams

#### Technical Documentation

For every piece of equipment installed on your network you should collect such information as serial numbers, memory configurations, system switch settings, network interface settings, installed boards, display type, and processor type. You should also note system-level software, such as device drivers, network protocols, memory managers, and operating system utilities particular to specific workstations or equipment. You should generate regular printed reports that your support personnel can carry to a user's office or site, in case a functioning network workstation is unavailable near a problem machine. Be sure to document each piece of equipment, including:

- workstations (by type and vendor)
- file servers
- print servers
- other servers (such as database and electronic mail servers)
- printers
- bridges
- gateways
- routers
- uninterruptible power supply units
- other service equipment (such as hubs and concentrators)
- peripherals

- modems
- scanners
- facsimile equipment

Recently, a number of software products have been introduced to assist with collecting this information from equipment that is already installed. Diagnostic products and memory management utilities, such as Quarterdeck Office Systems' Manifest, Aston-Tate's Control Room, and Touchstone Software Corp.'s CheckIt, give you a printout of a workstation's configuration, containing information about the system's memory, installed peripherals, configuration files, AUTOEXEC files, and memory management method. Products such as CheckIt also perform diagnostic tests and performance evaluations of the workstation that are useful. These products are adequate for networks small enough that the network support staff can go to each workstation and run the product to obtain the report. On larger networks, however, it may be impossible to physically access every machine, and the amount of information provided by these programs regarding network drivers, versions of shells, and other related items, may be insufficient.

Two products of special note for network environments are Brightwork Development Inc.'s LAN Automatic Inventory (LAI) and Horizons Technology Inc.'s LAN Auditor. In addition to collecting workstation hardware configuration information, these products capture information regarding network drivers, interface card, and shells.

Furthermore, they do this in real time on the network every time a workstation logs in. These products will greatly assist in the production of a comprehensive configuration database of workstations on your network.

#### Server Software Configurations

Network servers require documentation significantly different from workstations because of the large amounts of disk storage they provide, the varieties of programs they contain, and their numerous network functions. Changes to one portion of a server's operating parameters can greatly affect other aspects of the system's use.

Any documentation on servers should include information in at least the following categories:

- server installation parameters or defaults
- server AUTOEXEC and other start-up procedures and scripts
- directory structures
- system and user login scripts
- user access rights
- batch files
- network and application menus
- applications (including dependent files and directories, installation notes, update and revision history, serial numbers, and number of licenses)
- system environment variables
- default security assignments
- internetwork addresses

Having on-line information available only to you and your staff is not adequate. You may be quite confident in your system's integrity and know that everything is in order. However, unless you operate alone, with complete autonomy, there are others in your organization who must be aware of your activities and need to know how the system is performing, how the system is structured, and what its capabilities are. It may be your direct supervisor, it may be an executive in the finance department, it may be other business or line managers who are part of your user base. In many cases their need to know is as important as your

own, since their ability to perform their jobs may be closely tied to the system's ability to service their needs.

In today's business environment, few business managers or executives have the time, patience, or understanding to follow your attempts to explain security, for example, while viewing screen after screen of your network's system administration utility. Nor will they be impressed with your ability if every encounter involves rummaging through the system's batch and script files, or viewing directory trees and control files, to answer what they believe

to be a simple question. Your documentation should include a system overview that explains the system's capabilities and answers basic questions (see the sidebar "Your Mission Statement: The Network Charter").

## Documentation Components

The best time to document the system is while it is still in the planning and development stages. Design and analysis documents, notes from user interviews, memos, proposals, requests for proposals, and other information belong in your archive. Collect hardware and software serial numbers, purchase dates, costs, and license and warranty information while the products are being installed to save time later. Plan from the outset to maintain a library of duplicate or master copies of all software manuals, hardware manuals, installation guides, and other forms of documentation provided by manufacturers and vendors.

Instruct your installers or staff to keep notes of their activities and note the sequence of events during the installation, default parameters or settings, and other product requirements. Identify problems that occur during installations and note the solutions devised to correct them.

No single tool exists to help you assemble and produce all of this information for your documentation. File management programs and utilities, such as XTree Pro from XTree Co. and PC Tools from Central Point Software, will be helpful in obtaining printouts of your directory trees and structure. Text editors and utilities provided with menu programs and your network operating system may help you print system scripts, batch files, menu structures, and other information. For NetWare networks, a number of utilities, such as Fresh technology Group's Fresh Utilities and Cheyenne Software's Utilities for NetWare have reporting capabilities for documenting such items as user logins, system parameters, printer queues, and user profile information.

## System Services and Applications

In a smaller network environment, system services and applications can be included in the server software documentation. In larger environments, however, additional documentation may be required to adequately cover custom applications, gateway and bridge services, E-mail servers, or other components of a network. Many network managers develop custom routines to distribute network files, perform system maintenance and housekeeping procedures, and execute a variety of other tasks, often by combining capabilities of several existing applications or utilities. Typical candidates for documentation include:

- printer queues and servers
- support, housekeeping, and maintenance
- logical configurations, naming conventions, and organization
- remote access servers
- automated task servers and processing routines
- async servers
- telephone and access numbers
- custom or site-specific applications (such as accounting, telemarketing, or expert systems, or other complex applications that have unique requirements)
- custom gateways

- E-mail servers and routing routines
- database servers
- distributed processing routines

## System Operations Guides

Giving comprehensive documentation to your support staff is beneficial as you train new people and cross-train your staff to handle a variety of situations. Although most products installed on your network have their own documentation, producing summaries of routine procedures (such as system backups and restorations and system start-up and shutdown) lets you incorporate important items that reflect your particular system and method of operations. Furthermore, this documentation reduces the need to search through the voluminous manuals and references when a problem or failure occurs. Information can be included about methods and procedures for troubleshooting various system components, and who should be notified in the event of a problem, and how such notifications should occur. Other relevant items include:

- hours of operations and availability to users
- daily, weekly, monthly, quarterly, and annual operational procedures or processing tasks
- printer-related operations (such as how to change paper, toner, and ribbons)
- how to re-order supplies and forms
- backup schedules and procedures
- file recovery and restoration procedures
- file and data archival procedures
- tape, diskette, and other media specifications for re-order
- database and application maintenance
- help-desk services and problem determination procedures
- power-up and power-down procedures
- troubleshooting procedures
- system accounting and billing procedures
- security audits and reviews
- performance monitoring
- log purging and archiving

## Logs for Management Effectiveness

Maintaining logs of various system activities improves your ability to monitor system and staff performance and serves to increase your management effectiveness. Which programs, operations, or procedures should be logged depends largely on your particular organization and its concerns. In a high-security environment, access to the network and applications may require constant and detailed logging of every user's activity from login to logout. In other environments, you may want to monitor application usage to determine how often various programs are used and who is using them. In every installation, you should maintain logs of system backup operations, unattended processing procedures, and service and maintenance activities.



System performance information should be logged and periodically reviewed. Capturing data on disk storage availability and utilization, server performance and utilization measurements, and network activity and traffic utilization gives you an overview of vital network performance and capacity measurements. This information can be used to perform meaningful analysis and trend studies.

Several commercial software products are available to assist with the generation, collection, and analysis of system logs. Software metering products let you track application usage for virtually any application on your network. Metering products, such as Brightwork's SiteLock and Saber Software Corp.'s Saber Meter, give you reports and graphic representations of system usage and tell you what time of day various applications are used most heavily, who is using the application and how often, and how much time users are working within various applications. Both of these metering products have an added benefit of actually restricting access to applications according to the number of licenses owned and assisting you in determining the correct number of licenses you should purchase of a particular application program.

Help desk support programs, such as Brightwork's LAN Support Center, keep track of troubleshooting sessions, run remote diagnostics, note equipment repairs, and put the information on work ticket. Typical logs should include:

- system backups and restorations
- remote LAN accesses (who and when)
- gateway usage
- disk storage capacity and utilization
- application usage
- service and maintenance calls and results
- help desk and support activities
- automated processing routines
- unauthorized access attempts

---

### Vendor, Supplier, and Purchase Information

Maintaining adequate records about your vendors and suppliers and the products or services you have purchased for installation on your network is often viewed solely as an administrative function and seldom discussed in terms of system documentation. However, maintaining such records, in an electronic database when possible, will greatly facilitate the production of management reports regarding costs and expense trends, expense analysis, budget planning, and software licensing, and will provide answers to many other important questions.

You should create file folders containing information for every product installed on your network. Be sure to

include copies of software licenses, product warranties, keyboard templates, printouts of READ ME files, installation notes, technical support contacts and phone numbers, and upgrade announcements.

Copies of purchase orders and invoices should be filed by vendor. Even if you are not responsible for purchasing, having ready access to this form of documentation is important. With this information you can locate proofs of purchase to obtain product updates or support offerings from manufacturers. And, in the event your company is subject to an audit of software-license compliance from an investigative authority, you will have supporting documentation at hand.

---

### User Directory

Finally, the user community itself needs to be documented. With an on-line or hard-copy written record of user logins and profiles and group and access rights assignments, you can better audit these areas and have a convenient method to identify individuals on the network. On some networks, login identities are used exclusively, sometimes making it impossible to know if "JSMITH" is John Smith in accounting or Jan Smith in marketing. In some organizations, it is often difficult to determine correct titles and departmental affiliations when setting up group assignments and E-mail distribution lists.

Written user profiles give department managers and senior management the opportunity to review and authorize assignments that you have made. It helps management determine whether or not individuals in their departments are receiving proper access to system functions and services. A user directory should include:

- user name
- E-mail address
- department
- physical location
- phone number

Assembling and producing documentation for your network is an important but time-consuming task. The quality of your system documentation can significantly impact your network's overall reliability and ease of use. Good LAN documentation helps you identify and resolve problems quickly and efficiently. And the quality of information you give management helps everyone concerned make informed decisions and accurate plans. There is no one-size-fits-all system that can answer all of your documentation needs. However, by taking a global look at the network and effectively applying the tools at your disposal, you will surely improve the management and performance of your LAN. ■



# Network Analysis: Ten Steps to Fine-Tuning Network Performance

## In this report:

Cable Diagrams.....	2
Naming Nodes and Layers.....	2
Study Traffic Patterns.....	3
Run Benchmarks.....	4
Optimize Your System.....	4

## Datapro Summary

Optimizing network performance involves the use of network analysis, benchmarking, and troubleshooting tools. This report offers a systematic, 10-step approach to network analysis and optimization. These guidelines are designed to give LAN professionals a more structured approach to network analysis.

Today, analyzing a network is a necessary job for any network professional charged with improving performance and solving complex problems. Yet, if you're new to networks or to the discipline of network analysis itself, learning to perform such an analysis will be a challenging task. One approach to mastering it is breaking the task into 10 steps—each of which will be discussed in this report—that provide a rationale for maintaining accurate records and providing directions for benchmarking, optimizing and troubleshooting a network.

As an essential part of these steps, network administrators should begin using network analysis tools as early as possible to familiarize themselves with the network environment. Protocol analyzers, for example, allow network administrators to view the contents of packets transmitted on the network wire.

Using high-performance network interface cards and sophisticated analysis software, many analyzers can store several megabytes worth of captured LAN traffic to historical files for later review. Other typical analyzer features include triggering and

recording functions on specific events, and the ability to generate network traffic for test purposes.

While the cost of a protocol analyzer can extend into the tens of thousands of dollars, that cost is justified for anyone who must support mid-sized and larger networks. To use distributed processing technology is to accept the need for test equipment as a means to ensure the stability of sophisticated topologies and protocols. This approach will help deliver robust, error-free networked communications. (For more information on who sells protocol-analysis tools, see the vendor list at the end of this report.)

## Why Analyze?

Network analysis tools, benchmarking, optimization and troubleshooting are interdependent tasks. The foundation for optimization and troubleshooting a network is to benchmark its performance. In order to improve performance, you must benchmark current performance. Because it is impossible to benchmark a broken network, benchmarking must be part of routine maintenance.

Through benchmarking, you will learn the working characteristics of your network. These working characteristics, contrasted with those of a malfunctioning system, will help you find problems much faster. If you wait until you have a problem before using your analyzer, your task will be much more difficult. In short, an analyzer is a tool to be used often and consistently. It is

This Datapro report is a reprint of "Network Analysis: 10 Steps to Fine-Tuning Network Performance" by Bill Alderson, pp. 90-94 from *Network Computing*, May 1991. Copyright © 1991 by CMP Publications, Inc. Reprinted with permission.

a prerequisite to performing the following steps. Although the optimization techniques presented can be applied without an analyzer, it is more difficult to benchmark performance increases without an analyzer and troubleshooting is nearly impossible.

## Cable Diagrams

Most routine networking problems are related to the physical cable, an example of which is shown in Figure 1. Damage is often caused by daily movement as cables are crushed behind desks and crammed in wiring closets, as well as by users moving their computers.

Step one is fairly obvious but easy to forget: keep on top of these problems by documenting your network's physical cable segments and rings, making sure to note the location of any multistation access units (MAUs), hubs and repeaters. This will help you relate traffic and error statistics to a specific physical cable and location.

Specifically, make use of copies of building floor plans and carefully mark cable layouts and office locations on them. The more detail you use in your diagrams, the better. In the future, these plans will help you to avoid climbing around the ceiling to see where a cable goes. And be sure to review your diagrams regularly. Though many organizations document their networks initially, it's a rare one that updates it when a new segment or station is added.

## Diagram the Flow

The second step is to document the logical architecture of your internetwork by identifying routed areas and bridged segments. Doing this may be difficult, especially if the network combines multiple routed protocols and bridges. However, the ability to view the internet logic in the context of actual traffic pattern levels will point to changes that may improve performance. That's because visually

defining the architecture simplifies the process of improving the logic. Otherwise, the complexity will either confuse you or you'll lose track of the reasons for your earlier decisions and conclusions—or both.

Each internet protocol must be separately documented, even if you use multi-protocol routers. That's because most sites have host systems that route individual protocols, with special filtering for each protocol at different points. Individually documenting each protocol provides you with yet one more level of clarity.

## Naming Nodes and Layers

The third step requires that you name each physical node address with an alias in your analyzer—something just about all analyzers allow you to do. Use a naming convention that lets you identify the physical location of the station, the department, and user name of each node. This makes it easy to identify traffic by nodes and physical segments. You might also include each user's telephone extension so you can call them when they perform an inappropriate activity, such as backing up a hard disk over the network during peak working hours.

Routing systems use the internetwork address for routing decisions; packets are forwarded with a router's own physical address, rather than the sending node's physical address. In order to see a packet's real source and destination node, you must display a packet's internet address. By using aliases that combine physical and internet addresses, you can see who is communicating with and through whom—that is, you can identify the nodes and routers. At the network layer, this is one of the most powerful capabilities an analyzer has to offer.

The fourth step, then, is to name each network layer (internetwork) address on your system in your analyzer's alias names table. Not all analyzers support this capability, but it is very useful in learning and proving the logic of the internet packet flow.

Figure 1.  
Floor Plan Cable Layout

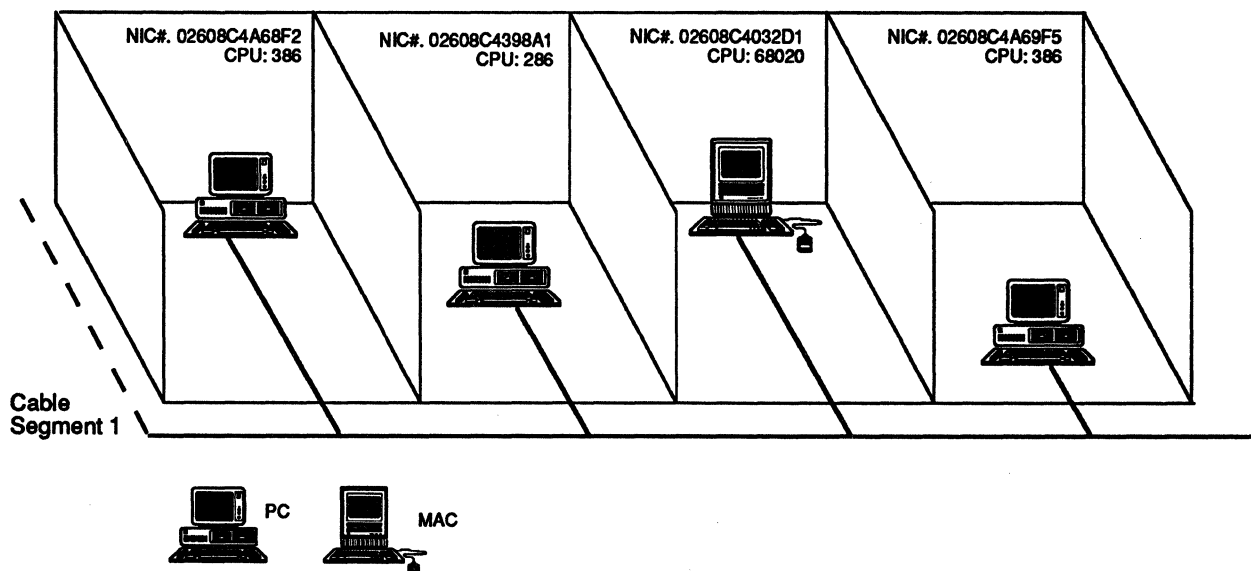
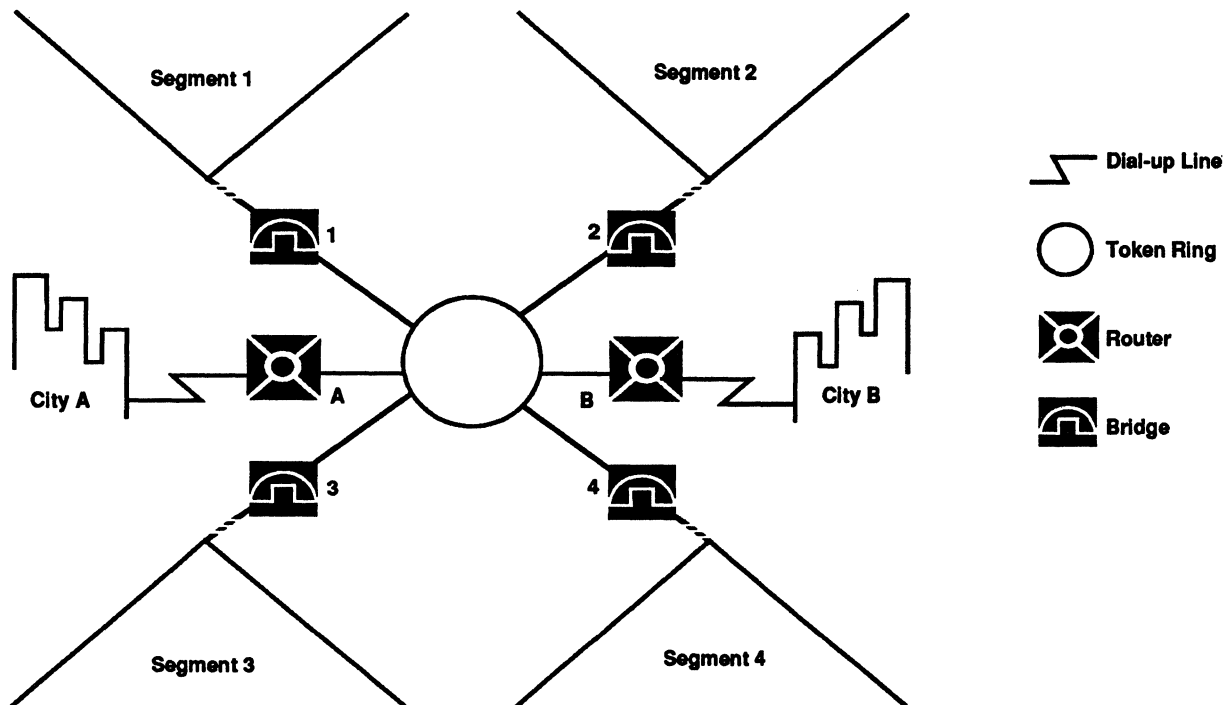


Figure 2.  
Physical Architecture



Naming the network layer can be difficult, because you need information about your internet protocols that may be difficult to obtain. You may also need help from whoever assigns the internet numbers for your company. Once this step is completed, however, you can view traffic from the internet perspective.

### Study Traffic Patterns

With steps one to four completed, you can now move on to studying the traffic patterns through the routers and bridges on your internet. New "connectivity" boxes are added to the network backbone regularly, and you need to stay on top of these additions and the changes they cause in terms of network traffic. Installers of such devices may not know that the amount of new traffic these boxes generate can affect the stability and performance of the original network. There are two specific cases to track. Figure 2 provides a diagram demonstrating traffic patterns.

You will need to track and analyze the traffic generated through any routers on your network. Follow packets from one node across a router. Record the path that packets take both to and from their intended destination. A packet may hop through several routers, which can be difficult to trace. However, you can move the analyzer to different areas to continue tracing the logic. If a packet leaves your physical plant, you can still trace the possible logic by studying your internetwork design documentation and making some educated guesses.

Once you've successfully followed one station's traffic, do the same for typical stations and servers at various points on your internet. Following the traffic will help you verify the internet logic diagrams you created in step two.

Next, configure your analyzer to filter between routers so that you can determine the amount of transient traffic

on that segment. You may find that your system is configured to route so much transient traffic through a segment that the network is saturated for stations on that segment. Servers may be routing more traffic than is prudent, which can degrade performance. Watch your router and server routing statistics to determine if that's the case. Usually these statistics include packets that could not be forwarded and were destroyed.

You can also use your analyzer to filter a particular network number or to include "hop counts" greater than one, which will show internet-bound or incoming traffic. (A hop count is a counter in a packet that is incremented every time a packet passes through a bridging or routing device.)

Once your router traffic is mapped, you will need to move to the bridges on your network. Because bridges use a station's own physical address across the entire network, a good naming convention, as suggested in step three above, will help you identify traffic patterns throughout your network. Filtering transient bridge traffic is difficult, and requires multiple traffic monitors and sophisticated use of an analyzer's alias table.

To find the optimal home segments for stations at a local site, use a spreadsheet to compare data from multiple traffic monitors. By naming every device, you can use an analyzer to watch packet flow statistics in each bridged area and follow the logical flow. You may discover a high level of transient traffic through a bridged area, so you will need to adjust bridge locations accordingly.

By using a thoughtful strategic naming convention and by sorting traffic by alphanumeric alias names, you can watch traffic flowing throughout any bridged environment on the internet. For example you might begin each name in a given bridged area with a different character, such as A-Z or AA-ZZ, which makes it a good deal easier to track traffic.

## Review Logins

The sixth step requires that you study the different types of logon sequences and the operational patterns of the applications being executed. Learning these will familiarize you with the logic of these sequences. By saving analyzer trace files of such activities to disk, you can study and compare traces when problems arise.

It is very important to familiarize yourself with all of your systems before you experience a problem, so make sure to maintain an ongoing and up-to-date collection of trace files. By comparing working traces to malfunctioning traces, you can concentrate on material problems without being distracted by normal sequences. For example, if you observe a NetWare login, you will see the workstation check the server version number from eight to 14 times within a few seconds, sometimes less than 8 milliseconds apart. If you waited until you experienced a problem to see this, you would probably call Novell about a server-workstation version mismatch.

## Run Benchmarks

The seventh step—benchmarking performance on the network wire—is easy. Observe a command going from a station to a server and watch it come back, then compare the time-stamps on the packets. Performing a precise response time measurement is easier than you might think. Every analyzer has a time-stamp feature that allows comparison of one packet-time to another.

Some analyzers use a relative time feature which can mark the command packet and calculate the time between spanning frames. This relative time-stamping can be done for various applications, workstations and servers in a quiescent or loaded traffic environment. In particular, internetwork response time can be quantified to calculate the effect of a highly loaded system on an application or to help determine the bandwidth requirements of a given WAN component.

## Review Configurations

Study your application, workstation, server and internetwork configuration parameters to identify potential variables that may affect performance. Observe sequences, such as the way applications search for files, use workstation caching, and negotiate block sizes. The type and configuration of workstation and server interface cards and drivers will greatly affect performance.

In addition, the location of a file server on an internet can affect the performance of stations logging into the server because of internetwork routing delays.

At times, rethinking your internetwork architecture will improve performance. For example, a programmer in a software development company was compiling code on three NetWare servers that all company programmers shared with the general computing population. But the network went down whenever multiple program compilations saturated the servers and network cable. A six-hour compile would fail in the middle, requiring system restart.

Steps to analyze this particular network's problems included documenting the physical and internet logic, naming the physical and internet addresses on the analyzer, proving internet logic, making productive changes in the routing logic, recording statistical utilization data, and examining various logons and application execution.

It turned out that the compiler program was looking through all the office productivity software search paths for every file it needed. Although the compiler did in fact have its own search path for the files it needed, it was not going directly to the files using the application path.

Why not? The answer was in the NetWare manual, which explained that workstation shell-configuration defaults specified file server search drives first, an application's path statements second, and the workstation's path commands third.

## Optimize Your System

This configuration hit the network with 15 times the necessary packets, which saturated the network; and with 15 times the necessary number of file access requests, which saturated the server. Simply changing a one-line parameter in the SHELL.CFG file of each workstation specified the application's path first, which reduced compile times by nearly half. Optimization can be this simple.

All of the steps detailed above lead to step nine—implementing an optimization scheme based on all of the information you've gathered. Choose a sample workstation and change the parameters you identify as affecting performance for your network one at a time. After changing a parameter, exercise all aspects of system operation and then benchmark performance by timing packets. If the best configuration works properly from all accessing nodes and overall performance is improved, you've managed to optimize your system.

In order to optimize the system in the software development company mentioned earlier, we experimented with the workstation shell-configuration parameters one by one, watching the search sequence on an analyzer, and measuring the response time from the first file-access command until the file was actually accessed.

In one case, each search was reduced by over 32 milliseconds, which translated into an aggregate 50% reduction in total compile time.

## Document, Document, Document

Finally, step 10: document—in detail, regularly and frequently. Organizations large and small suffer from a lack of documentation procedures. We have yet to find a system that would not benefit from better documentation; not just for its physical cabling, but for all aspects of its configuration. Consequently, technicians must repeatedly and needlessly figure out the overall structure of the network in order to solve each trouble call.

Plan for the time and resources to document the system and it will pay big dividends. Otherwise you run the risk of having employees who know the network and who will move to other departments or companies, taking with them all of the accumulated knowledge locked inside their brains with them.

Those are the steps; repeat them often. And make sure to become proficient in using the professional tools available to you. Not only will you get more from your system, it will prepare you for solving the most complex problems in minutes.

---

**Product Notes**

For additional information contact the following vendors:

**Bytex Corp.**

Southborough Office Park  
120 Turnpike Rd.  
Southborough, MA 01772-1886  
(508) 480-0840

**Hewlett-Packard**

HP4972A Division  
5070 Centennial Blvd.  
Colorado Springs, CO 80919  
(719) 531-4000

**Microtest**

3519 E. Shea Blvd., Suite 134  
Phoenix, AZ 85028  
(602) 971-6464

**Network General Corp.**

4200 Bohannon Dr.  
Menlo Park, CA 94025  
(415) 688-2700

**Novell**

LANalyzer Products Division  
2180 Fortune Dr.  
San Jose, CA 95131  
(800) 243-8526  
(408) 473-8333

**Spider Systems**

12 New England Executive Park  
Burlington, MA 01803  
(617) 270-3510

**Wandel & Goltermann**

1030 Swabia Ct.  
P.O. Box 13585  
Research Triangle Park, NC 27709  
(919) 941-5730 ■





---

# Fault-Tolerant LANs

---

## In this report:

Fault-Tolerant Topologies .....	2
Selecting the Right Equipment .....	3
Hubs and Connections .....	4
Network Management Issues .....	4
Remember the Basics .....	5

## Datapro Summary

Taking measures to implement fault-tolerant systems into company networks greatly reduces overall costs in the long term. Three key aspects of a network—design, installation, and equipment—must be addressed by fault-tolerance steps in order to maintain a network. Simple things, like misconnected wires, are easy to overlook and cause big problems for networks. By following certain guidelines, it is possible to construct a fault-tolerant network that is compatible with your requirements and budget.

Once seen as a novelty, a convenience, or an efficient way to share resources, LANs have become an integral part of the computing infrastructure of most businesses. Yet, unlike other services that are essential to corporate life—telephones, heat, electricity, coffee machines—networks are often viewed with suspicion, because they are seen as being less reliable than they should be. It is no longer enough for networks to work; they must work as needed. The solution is to make the network fault-tolerant.

Fault tolerance in a LAN is not an all-or-nothing proposition. The amount you build into a network depends on the importance of the operations the network supports. However, you can enhance the fault tolerance of any network without significant added expense by careful planning and installation. For example, you can deploy existing bridges in a way that adds redundancy between links. You can also avoid placing cables next to large electric motors.

Beyond such basics, you can take steps to ensure that your network will be available when you need it. These steps need not

dramatically increase network costs. Indeed, if the cost of being without critical applications or data is considered, adding fault tolerance may dramatically *reduce* costs overall.

The three key aspects of network fault tolerance are design, equipment, and installation. If any one of these is overlooked, the network will not be fault tolerant (e.g., a network may have redundant power supplies, but they won't help much if a backhoe digs up a cable and takes out a link).

## Fault Tolerance by Design

Often, you can make a network more fault tolerant merely by reconfiguring the existing components. However, mixing networks, hubs, media, and protocols is as much an art as a science. Do you need a totally redundant network, or can you get away with only backing up some key links? Do you need fault tolerance on every desktop? Although the permutations may seem limitless, there are guidelines you can follow in designing a fault-tolerant network, with an eye to costs and your needs.

The network's fault-tolerance requirements depend on the importance, number, and distribution of its applications and workstations. These factors determine the network's size and topology.

First, you have to define the scope of the network. The fault-tolerance plan for a workgroup network will be different from

---

This Datapro report is a reprint of "Perpetual Networks" by David Fowler, pp. 205-212, from *Byte*, Volume 16, Number 8, August 1991. Copyright © 1991 by McGraw-Hill, Inc. Reprinted with permission.

one for a facility network that spans an entire building or campus. Chances are, the larger the network, the more valuable it is going to be and the greater the need for facility-wide fault tolerance.

Once you define the scope of the network, you need to identify the mission-critical applications. Sometimes it's not the number of computers that determines the need for fault tolerance but the importance of the applications. If all the traders in a securities company need access to market feeds, then the links to those market feeds must never fail. On the other hand, it may not be a severe problem for any workstation (or even workgroup) to go down if other workstations are available to access key links. It's the link to the mission-critical application that requires fault tolerance.

The obverse of the mission-critical application is the mission-critical user. The premise of enterprise computing is that key users, such as the company president, can monitor most of a company's critical activities from their desktops. From a key user's point of view, all applications may be mission-critical. So, regardless of the fault-tolerant strategy of the network as a whole, according to this person, the whole network may need to be fault tolerant.

Critical applications and key users notwithstanding, it is size and topology that drive most networking decisions. As networks grow, their topologies change to serve more people at a lower cost per user. Rather than have one large network serving everyone, most companies install many small networks dedicated to individual workgroups and then tie these networks together. An important benefit of this natural evolution is that many small networks are generally more reliable as a group than one large network. If a small network fails, the other networks keep working (unless they rely on the failed network as a common link or backbone).

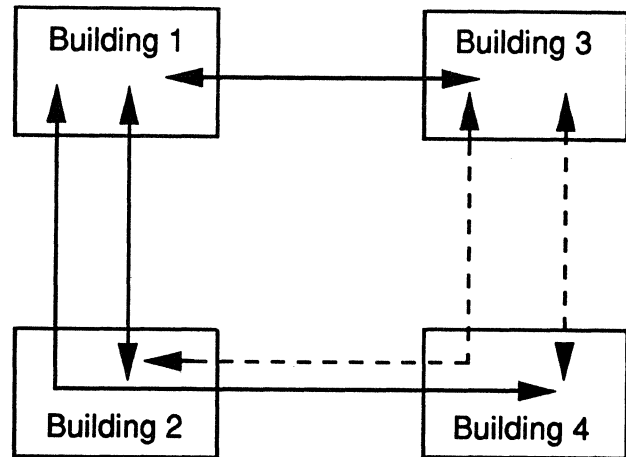
A basic principle of fault tolerance is that topological changes yield the greatest reliability for the money since they require the smallest incremental investment per user. Because companies usually change their network topologies to accommodate growth, they may as well select topologies that are inherently more fault tolerant than others. Once those topologies are in place, a company can go to the next level of fault-tolerance considerations: deciding which links to make fault tolerant, usually by making them redundant. The greater the need for the link (because of either the number of workstations, the importance of the application, or the importance of the user), the greater the need for redundancy on that link.

## Fault-Tolerant Topologies

A topology achieves fault tolerance to the extent that it decreases the number of *hops* (a connection between two networks) between redundant links and workstations. A network may be fault tolerant even though no workstations are attached to redundant links, provided that the end links are connected to redundant intermediate and backbone links. In the past, networks were interconnected on a large scale, and workgroups of 40 users were common. Today workgroups of eight users are typical. The difference is that those eight users can reach a much larger network that is more fault tolerant than the workgroup LAN.

While traditionally thought of as loops or buses, most workgroup LANs being built today are actually star configurations, with each workstation in the star tied directly to a single LAN module in an intelligent hub. Hubs (also called concentrators) provide a central facility to interconnect dozens of network links. You can connect LAN modules

Figure 1.  
Backbone Redundancy



Providing increased fault tolerance is often a matter of re-deploying network assets or making modest investments in equipment. Here, the backbone between the four buildings can be made completely redundant by adding two links (dotted lines).

inside a hub to provide a path between the workgroups and tie the hubs together on a backbone to interconnect everyone. To enhance reliability, you make backbone links between hubs redundant and create multiple links to tie workgroups to a hub.

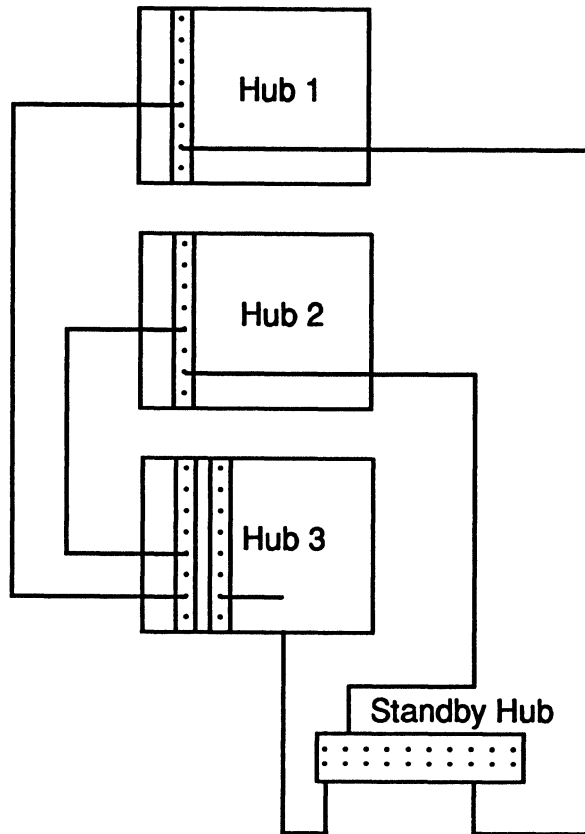
The prime candidate for redundancy is the network backbone because it is the link that ties your subnetworks together. As the network grows larger, backbone fault tolerance contributes progressively less toward the total network costs, because larger networks have more links available for use as backup. Figure 1 shows how you can make a backbone connecting four buildings redundant by adding two links.

You can accomplish backbone redundancy at two levels: cable redundancy and hub redundancy. In the former, a second physical wire links each of the hubs. While this protects the network if one of the wires is broken, it doesn't help if one of the hubs fails. To ensure against hub failure, you can add a third, smaller backup hub, as shown in Figure 2. If one of the three main hubs fails, the other two can still communicate.

Figure 3 shows how a hub/backbone-style topology is inherently fault tolerant and how it provides a natural platform for additional levels of fault tolerance. Some of the inherent fault tolerance results from using hubs to divide workstations into groups. Even without redundancy, the failure of a link to a workstation would not affect the ability of the workstations on the other hubs to communicate.

You achieve further fault tolerance to the extent that you isolate the workstations at each hub from each other. If all workstations share the same LAN module, as in hub 3 in Figure 3, that card is a single point of failure for those workstations. If workstations are connected to different modules, as in hubs 1 and 2, a failure in one card won't affect the workstations connected to the other card. In hub 1, half the workstations are connected to one module, and half are connected to the other; a module failure will affect only half the workstations. In hub 2, each workstation is assigned its own module so that one link failure will affect only one workstation.

Figure 2.  
Total Backbone Fault Tolerance



Making a backbone truly fault tolerant requires redundant hubs as well as redundant cabling. The standby hub ensures that the link between any two primary hubs will not be lost if the third hub goes down.

By using hubs with large numbers of slots for LAN modules, you can increase the number of workstations with direct access to the redundant backbone and decrease the number of workstations on nonredundant links. Some hubs even provide the ability to set up redundant links between the modules in the hub.

Having a network with a high level of fault tolerance may not help much if the link to a critical user or application goes down. In those situations, you may need a redundant link directly to critical applications or users (as shown in Figure 4) to achieve *to-the-desk* fault tolerance. Here a transceiver splits the link coming from the application's host computer, and each link goes to a different module in the hub or to a different hub.

### Selecting the Right Equipment

Network components are critical to fault tolerance for two reasons. First, if equipment did not fail, there would be no need for fault tolerance. Second, the most fault-tolerant network topologies cannot be built unless the equipment being used supports them.

Nearly every network component can contribute to the overall goal, either by being fault tolerant itself or by delivering features that make the network fault tolerant in a substantial way. The most vulnerable components in any computer system are those with moving parts and those

that generate a significant amount of heat. Such components present the best opportunity to directly build in fault tolerance.

In major network components, such as hubs, power supplies are appropriate points for redundancy. You can enhance the fault tolerance of the power supply by monitoring its operation and reporting any irregularities to the management system. Such monitoring can include hot-spot detection and fan diagnostics, which anticipate trouble before it happens.

Redundancy is of limited value unless there is a clean *switchover* of control between units after a fault has occurred. The switchover should be accomplished, either automatically or under management control, without disrupting the network.

Other network components that contain redundancy features may include modules with built-in switch-over logic between ports so that redundant backbones can run off the same module. However, the ability to install redundant hardware takes the network only so far. Being able to automatically switch between the primary and backup units without service interruption makes a network component truly fault tolerant.

Other features can also increase a network component's reliability and fault tolerance. *Hot swapping* is a perfect example. To remove and install standard modules, you must power down the system (e.g., a hub) to prevent the change in the resistance load from causing a voltage spike and harming the circuitry. But slots and modules can be designed to keep voltage swings within safe limits when swapping modules in or out.

Another feature is the ability to isolate a card from any port, either by setting a switch on the module or by sending a software command from the network manager. This feature is helpful if you need to isolate the network running off a hub from a backbone during testing.

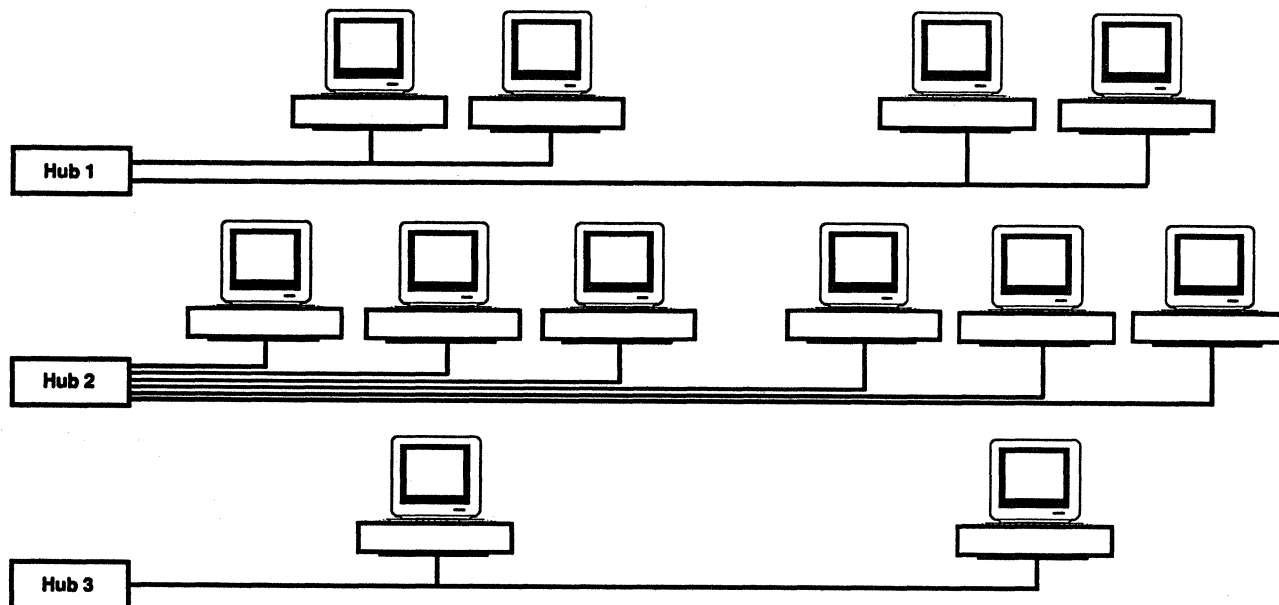
At another level, module isolation can also be important. In Token Ring networks, a failed network module in a workgroup can bring down the LAN for the entire group. With an *auto-wrap* feature installed in the module for link failure and card failure, the card becomes a passive part of the link if the module fails or loses power. Thus, other network users can send data to each other through the module as if it were still functioning.

Another instance where module isolation comes into play is in spanning-tree software used in bridges that tie two networks together. This software enables bridges to passively adjust the network traffic load running over parallel links, based on transmission loads or link outages. If two networks are connected by three bridges and one of the three fails, the other two will take over without operator intervention.

At the most basic level of component fault tolerance are reliability enhancement features—those that reduce the likelihood of a problem occurring in the first place. One such feature is placing female plug connectors in the hub slot rather than in the modules. This way, if a pin is bent during installation, the module must be repaired or replaced, not the hub.

Components need not be completely fail-safe to deliver features that are critical to network fault tolerance. The key system that defines network topology is the hub, and not all hubs are created equally fault tolerant. Simply by having more slots, a hub can support more links to workgroups. More links mean greater redundancy, and more connections to the hub decrease the impact of any single link failure.

Figure 3.  
Fault-Tolerant Topology



A hub/backbone topology isolates parts of the network, ensuring that the failure of one module will not affect all the rest. The higher the ratio of modules to workstations, the greater the fault tolerance of the network as a whole.

### Hubs and Connections

Much more critical than the number of slots is the way the hub creates connections between all the links coming into it, including the backbone, and how those connections are managed during network operation. The basic function of a hub is to offer a channel between several links. Networks attached to the hub can talk to each other and to the backbone over these channels. The greater the number of channels available in the hub, the greater the variety of possible network topologies, and the greater the opportunity for fault tolerance. If a hub has three channels, one might be used as a standby for the other two (a so-called  $n + 1$  redundancy strategy).

The number of channels inside the hub is important; so is the way those channels are implemented—whether they're hard-wired or use bridges. Bridging a link/channel or a backbone/channel connection lets you use a spanning tree to reroute around failed channels, links, or bridges. Figure 5 shows how a hub with three channels uses a spanning tree to achieve multilevel redundancy over four bridges. In this example, either one of the bridges on channel 2 or channel 3 can fail without interrupting traffic to anyone. A failure in channel 2 or channel 3 would not affect workstations on the remaining channels.

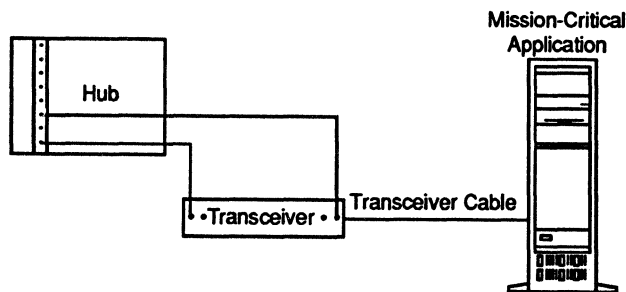
Another aspect of topology is the device used to assign channels to links. *Software topology switching* lets you define channels and reroute traffic using software commands. This is more productive (and inherently more reliable) than shutting off the hub and rewiring the chassis. A prerequisite for software topology switching is that the hub be able to define channel protocols on the fly. For example, implementing an  $n + 1$  strategy on three channels—one Token Ring, one Ethernet, and one standby—is impossible if the standby channel cannot be defined dynamically as either Token Ring or Ethernet.

### Network Management Issues

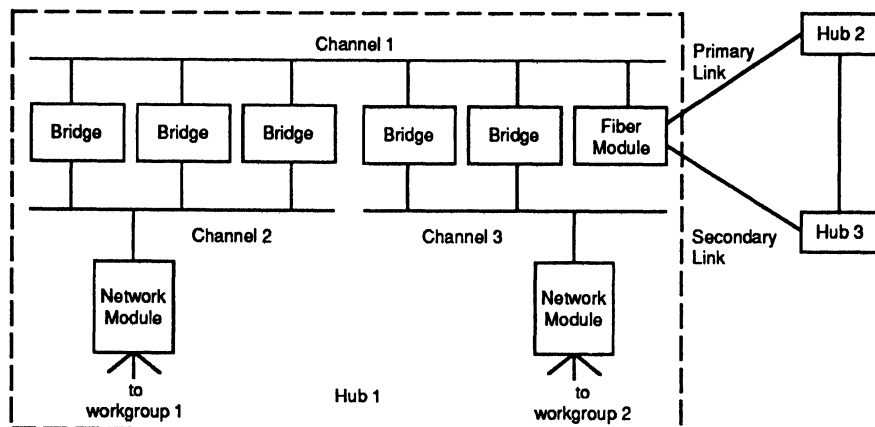
Network management is another component that helps define the fault tolerance of a network. Can the management system initiate recovery after a fault, or does it simply flash a red light on a screen? What amount of detail is reported by the top level of the network management hierarchy? Can an operator running NetView identify a user on a specific port of a specific Ethernet module at a specific hub and change that person's access privileges?

Operators of fault-tolerant networks need to know what is going on throughout the network and to be able to take direct action from the management console. Industry-standard management protocols, such as the Simple Network Management Protocol, provide a way for devices from different vendors to send status information over a

Figure 4.  
To-The-Desk Fault Tolerance



Often it is the connection to a critical user or application that must be made fault tolerant. Here, the machine supporting a mission-critical application has two links to the hub. You could provide even greater fault tolerance by making the links to two different hubs.



*Figure 5. Multichannel Bridging Within a Hub*

*The use of bridges within a multichannel hub provides fault tolerance and, using a spanning tree, congestion control. No module will be isolated if any bridge fails, and the failure of any channel will not affect the workstations on the other channel.*

network and to receive operational instructions. The protocol specifies both mandatory and undefined frames within a management data packet. Vendors are able to utilize these undefined frames to add significant functionality to the basic protocol. Typical SNMP status information consists of the number of devices, the number of errors, the incidence of device failure, and the number of packets transmitted.

More robust status-reporting capabilities can provide information about the following:

- ports (e.g., the on/off status, the security privileges, and the identity of the users);
- modules (e.g., the networks they belong to);
- backup links (e.g., the links they back up);
- bridges (e.g., the networks they bridge and the configurations they support); and
- networks (e.g., the active and backup configurations).

How management systems capture and use status information varies from vendor to vendor. To support fault tolerance, a system should make the information as clear as possible to the operator. Some systems support a point-and-click option that lets you select devices on a network map to display their status information; some support a window hierarchy that lets you click through successive levels of networks, subnetworks, and devices. For example, if you click on the picture of a hub, the system displays the front panel of the hub, complete with status indicators. You can then click on a module to display its ports, with a list of which users are connected to which ports.

A management system should allow for preprogrammed and proactive responses to network situations. A

preprogrammed response is necessary when problems occur and the network manager is not available. It can take actions that can be planned in advance, such as switching to a backup channel when a primary channel fails. Proactive responses are measures an operator can take to make the network run better—before a problem occurs (e.g., fine-tuning spanning-tree parameters to reallocate traffic passing through selected bridges).

The highest level of management control is to be able not only to operate the existing network but also to create the network by bringing up new topologies, device configurations, and equipment from the management station. For example, in response to network commands, intelligent bridges and network modules within the hubs can establish or cancel links, define channel protocols and configurations, and allocate ports among workgroup subnetworks. The fault-tolerance benefits of such network-creating capabilities are that they allow the network manager to define solutions to problems that may not have been entirely foreseen when the network was installed.

### Remember the Basics

Ironically, the primary issue surrounding fault tolerance is the one most likely to get you into trouble: installation. In fact, the most common source of network failures is the wiring. Wires are attached to the wrong connectors or installed over fluorescent lights. Other problems occur because planners forget to measure the equipment closet before ordering the equipment. Or they remember to measure the equipment closet but neglect to order enough cable extensions.

Many of these issues, no doubt, will fade as the presence of networks in the business environment becomes more familiar. As networks become more reliable and more commonplace, we may even start taking them for granted. ■



---

# The LAN Diagnostic Process

## In this report:

Diagnostic Devices .....	3
Uses for Diagnostic Devices .....	5
Planning, Training, and Resolution .....	7

## Datapro Summary

Problem diagnosis on local area networks (LANs) varies from trivial to extremely difficult. The difficulty is proportional to the variety of protocols on the LAN and the complexity of the network. LANs are evolving via better software, more powerful servers, and internetworking devices that extend the physical network. Bridges and routers regulate traffic flow, improve performance by subdividing a large LAN into smaller segments, or combine separate LANs into subnets of a larger LAN. Additionally, LANs are being interconnected using a combination of gateways and wide area networks (WANs), across a geography ranging from intrabuilding to worldwide.

All of these variations can produce a network topology where information originates anywhere and travels anyplace; and in a healthy, well-designed LAN, it does not impair its performance. Like a diplomatic courier traveling Europe, it must cross many borders, speak many languages, and maintain many protocols. Such diversity requires better and more sophisticated diagnostic tools for fine-tuning a healthy LAN and solving problems on an ailing LAN. Whatever the topology, simple or complex, an understanding of the diagnostic process helps keep any LAN operating and functioning at its best.

---

## Background

The diagnostic process involves recognizing that problems occur; preparing to deal with them; and, when they do occur, finding and solving them. In short, it demands planning, training, and resolution. The diagnostic process in an organization should evolve with the growth of the LAN. The difficulties encountered in dealing with problems are normally exponentially proportional to the complexity of the network.

This section provides three sample networks with increasingly complex topology, examples of situations in which a protocol analyzer might be useful, and an overview of several types of LAN diagnostic devices.

—By Arnold S. Cleff

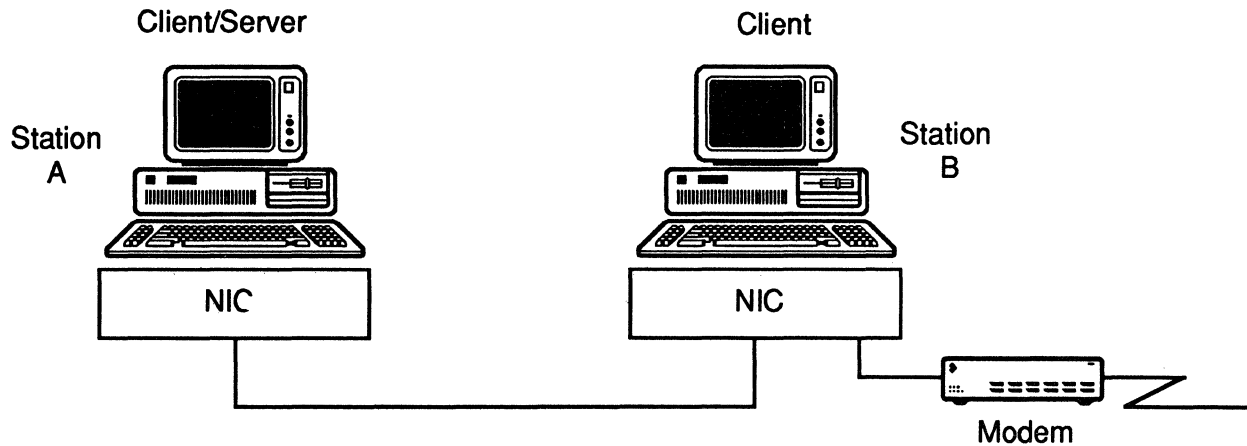
## Sample Networks

### Two-Station Network

The most fundamental LAN consists of two computer-controlled devices connected by a high speed (M bps), privately owned communication path. (See Figure 1.) The computer-controlled devices can be workstations, personal computers, or any host computer. The communication path (twisted-pair wire, coaxial cable, or fiber optic cable) and its interface to the computer via a network interface card (NIC) may be proprietary in design, such as Ethernet or Arcnet, or it may conform to the IEEE 802.X standards. The LAN's network operating system (NOS) software may be purchased (Novell, Banyan, Microsoft, etc.) or it may be user written.

A server station provides network services, while a client station is a user of network services; in certain instances, mainly

Figure 1.  
Two-Station LAN



A fundamental LAN consists of two computer-controlled devices linked by a private communications path.

for economic reasons, a station can perform both roles. It would be difficult to justify major expenses for diagnostic devices for only two stations; perhaps a volt-ohmmeter would do in the beginning.

#### Multistation Network

Adding more stations, particularly if they are in the same physical area, provides only minor complexity. (See Figure 2.) These added elements are likely to be more PCs, servers, and/or specialized workstations, and perhaps a bridge or router to segment the LAN to decrease traffic flow and provide performance benefits. The bridge or router passes only the traffic specifically destined to a station on a different network segment. At this stage there is still only one LAN and ample room for growth. Problems on this LAN can be diagnosed with minimal effort, but the need for sophisticated diagnostic devices is beginning to emerge.

#### Interconnected Network

As the need to add more users increases, the LAN in Figure 2 could be extended into subnets. Later, when it becomes advantageous to share information, the LAN can be interconnected with one or more additional LANs. (See Figure 3.) Each subnet is an independent LAN, complete with its own server, and its stations can send information through the bridge or router to stations on the other subnet. Repeaters boost signals on lengthy cable sections. Gateways translate from one protocol to another.

Installing devices such as a repeater (which operates at OSI Layer 1), a bridge (which deals with OSI Layer 2 frames), a router (which deals with OSI Layer 3 packets), or a gateway (which operates at OSI Layer 4 and beyond), or using a wide area network (WAN) to transport data between LANs, is a function of designing the network to satisfy the organization's needs. This complexity calls for considerably more sophisticated diagnostic devices and the necessary user skills in handling them.

#### Wireless LANs

The LANs discussed in the previous sections imply the use of hardwire connections. A relatively new technique, the wireless LAN, is also available for connecting stations within a LAN or interconnecting separate LANs. The basic

wireless transmission techniques use infrared for line-of-sight transmission, with a nominal upper limit of 150 to 500 feet; or radio waves for non-line-of-sight transmission, for LANs within a building or from building to building.

Some wireless devices are LAN operating system-dependent, and are suited only for a specific implementation such as Novell NetWare or Banyan VINES. Others are transparent to the operating system. Wireless LANs present an additional expense and add another degree of complexity in maintaining the network. At present they are in limited use, but the market for wireless LANs is expected to grow.

A wireless LAN can be used in any situation—a two-station network, a multistation network, or to interconnect LANs—where it is beneficial to replace wired connections with wireless connections.

#### LANs, WANs, and MANs

When two LANs are physically adjacent, they can be interconnected directly using bridges and routers. When LANs are separated by larger distances, various media are available to make the interconnection. These media include dial-up telephone lines; leased telephone lines high-speed circuits such as T1, T3 or ISDN; or microwave. Choosing the right one requires special knowledge and usually the help of an expert.

The media that provide remote LAN interconnections typically fall into one of the following categories: wide area network (WAN), metropolitan area network (MAN), or private network. Various combinations are frequently used. The decision on which to use can be complex and is often a factor of the hours of use per day, traffic rate, economics, and availability of equipment.

The equipment required to diagnose problems related to WANs is different than that required for LANs. If a problem originates from within a WAN, the WAN service provider is responsible for fixing it. However, the user often must first prove that the WAN equipment is the problem source. Users should discuss the methods and responsibilities for problem diagnosis when contracting for these services.



**Table 1. Uses Versus Device Type**

Diagnostic Device Type	Use 1 Defining Network Norms	Use 2 Preventive Diagnostics	Use 3 Problem Solving—Nonintrusive Monitoring	Use 4 Problem Solving—Intrusive Testing	Use 5 Postrepair—Monitoring
Protocol Analyzer	X	X	X	X	X
Software-Only Protocol Analyzer	X	X	X		X
Network Management System	X	X	X	X	X
Time Domain Reflectometer	X	X		X	
Optical Reflectometer	X	X		X	
Volt-Ohmmeter	X			X	
Cable Tracer	X			X	
Oscilloscope	X	X	X	X	X

### Diagnostic Examples

#### Light-Duty Diagnostics

Assume for a moment that Station A in Figure 1 frequently performs CAD-related, CPU-intensive functions and that—with a large disk capacity—it also doubles as a file server. This situation creates the possibility that Station B becomes sluggish while waiting for files from A. If B in turn were tasked with a transfer, via modem, of large CAD files stored in A to a node external to this LAN, B may not be able to sustain the maximum modem transfer rate. If A and B communicate via a newly installed, home-grown LAN operating system, it is not implausible that some subtle bugs remain, such as extraneous or repetitive handshaking, that further degrade performance. By copying and decoding the LAN traffic, a protocol analyzer could reveal the abnormalities and permit the necessary corrections to improve performance.

#### Medium-Duty Diagnostics

The network illustrated in Figure 2 introduces the added complexity of LAN segments and myriad possible problems. Problems can include multiple stations inadvertently configured to the same logical address by human or physical failure (each logical address must be unique); traffic congestion at an overworked bridge, causing lost messages as the bridge buffer overflows, followed by even more traffic congestion as unacknowledged messages are repeated by the sending station; routers erroneously programmed for nonexistent stations; or a subtle mismatch in protocol between different vendor devices. Again, through the use of diagnostic devices, these conditions will become visible.

#### Heavy-Duty Diagnostics

Should a LAN manager need to troubleshoot jointly with a colleague in another interconnected network, they must have the right tools. Figure 3 illustrates the following scenario. A department manager in Seattle uses an interconnected corporate LAN to prepare a key report based on information stored in the IBM mainframe (SNA gateway) in Los Angeles and the Digital Equipment minicomputer (Ethernet) in New York. The manager then forwards the executive summary to the vice president in Chicago at an IBM PC (connected to a token-ring LAN) and also sends

selected details to subordinates in Seattle. Imagine the correct steps in problem solving in that scenario when troubles arise between different locations. In this case, LAN managers at each LAN, coordinating their actions via telephone, must use diagnostic devices such as protocol analyzers to track real and/or test messages to attempt to isolate the problem source(s).

### Diagnostic Devices

For this report, a LAN diagnostic device is defined as consisting of a hardware/software combination, hardware only, or software only, used for nonintrusive monitoring or intrusive testing of a LAN to measure and analyze electrical, optical, or logical parameters. Examples of diagnostic devices are cable tracers, volt-ohmmeters, time domain reflectometers (TDRs), optical reflectometers, oscilloscopes, and protocol analyzers. Diagnostic devices, particularly protocol analyzers, can come with an array of software packages specifically designed to expand their capabilities. Increasingly, network management hardware and software plays a large part in diagnosing problems on the LAN.

#### Protocol Analyzers

These are among the most versatile devices available for diagnosing problems on a LAN. While generally more expensive than other devices (typically \$9,000 to \$20,000+), they can extensively analyze problems of logical origin. Some are also equipped to deal with TDR measurements.

Fundamentally, a protocol analyzer captures frames of information from the network and selectively examines the frame content. Based on the OSI seven-layer model, the protocol analyzer has access to all layers within the frame, from the physical interface (Layer 2) to the application (Layer 7). The frame's content depends on the network protocol, the frame's source of origin, and any malfunctions encountered along the network. Furthermore, not all protocols are modeled exactly on the OSI standards. To keep manufacturing costs competitive, vendors design protocol analyzers to deal with the subset of all possible protocols and LAN types that they choose for their products.

Software-only protocol analyzers are a less expensive way to analyze the LAN, usually costing less than \$1,000. While they have fewer features than hardware-based protocol analyzers, they can extensively analyze a large variety

**Table 2. Diagnostic Planning Matrix**

Planning Element	Category				
	Geography	Inventory	Documentation	Personnel	LAN Problems
Plant Layout	X				
LAN Interconnects	X				
Equipment List		X			
Spare Parts List		X			
Diagnostic Equipment List		X			
Diagnostic Equipment Reference Manuals			X		
LAN Software Reference Manuals			X		
Application Software Reference Manuals			X		
LAN Specifications			X		
Users				X	
Internal Maintenance Personnel				X	
External Maintenance Personnel				X	
Active Trouble Tickets					X
History Log					X

of protocols and should be given more than casual consideration. Software analyzers permit network monitoring for traffic rate and active nodes, filtering for parameter and pattern detection, and other types of analysis.

### Network Management Systems

**Network Management Software:** These products are playing a larger role in LAN diagnostics. Network Management requires a special protocol to communicate information between the central control station and the controlled devices in the network. By default, the Simple Network Management Protocol (SNMP) has become the dominant network management protocol. Many LAN devices now have a built-in capability to provide essential diagnostic information and communicate it back to a central SNMP management station. From the SNMP station, the LAN administrator is able to track in real time the changing status of the network and to initiate recovery commands to SNMP-controlled devices.

In addition to the SNMP protocol, various proprietary protocols are employed in many network management systems to communicate with servers, bridges, routers, and other LAN devices. These network management software packages are available from a variety of vendors and function in operating systems such as DOS, UNIX, and OS/2. They provide a range of capabilities from alarms for change detection to full network management diagnostics and restoral. Feature sets can include maintenance of a database of all devices on the LAN, display of a map of the network with various equipment states and performance statistics, and remote control and switching of LAN devices.

These products operate in different modes. Some send and receive information in-band (like any other device in

the LAN); some operate out-of-band (by means of a communication path separate from the LAN); and some have both capabilities. They range in price from a few hundred dollars to over \$10,000.

**Network Management Hardware:** These products come in two forms—standalone devices for controlling various LAN components or built-in components of a bridge, router, or hub. Some network management hardware automatically detects faults such as failed hubs, adapter cards, or cables and initiates a switchover to remove the defective path from the network. They can function either in standalone mode or in combination with network management software. When used with the SNMP protocol, the device sends a signal to the SNMP station so further corrective action can be taken.

Other types of network management hardware are installed remotely in the system at strategic locations to monitor various parameters and then transmit these statistics to the central SNMP station.

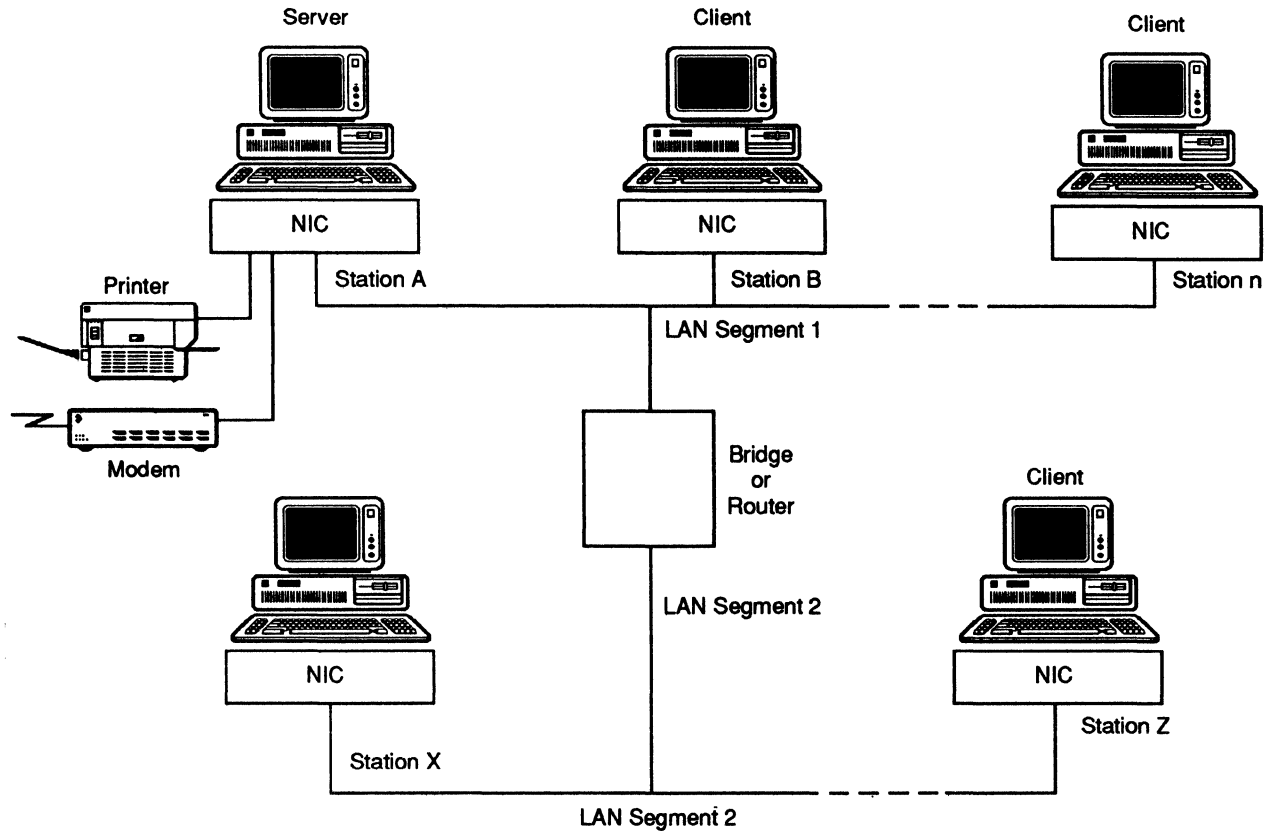
### Time Domain Reflectometers (TDR)

TDRs are standalone devices used to locate cable faults to within a few feet (or inches, for the expensive ones). Normally in the \$1,000 to \$5,000 range, they provide a picture (graph) and are therefore useful for diagnosing a deteriorating condition to provide an early warning before a severe failure occurs.

### Optical Reflectometers

These are used to test for breaks in the fiber optic cable path. They are relatively expensive (\$10,000 to \$20,000), but they are certainly superior to using a flashlight and a guess. For a long run of fiber backbone, they may be a necessity to ensure a quality path.

Figure 2.  
Multistation LAN



A multistation LAN of minor complexity.

**Volt-Ohmmeters**

These are relatively inexpensive devices (\$50 to \$100) used to measure copper conductivity when searching for breaks or poor connections. They also provide a rough means to measure voltage levels and voltage losses. Volt-ohmmeters are a lower cost substitute for TDRs; however, they cannot pinpoint the location of a break or poor connection from a distance, nor can they provide a graphical picture for analysis or for documenting a current state for trend analysis. They can also be used to validate proper voltage and grounding conditions at the wall outlet.

**Cable Tracers**

These are devices used in pairs to continuously inject a signal into the cable at a fixed point with the transmitting device and then track the cable's path with the sensing device through ceilings, behind walls, in ducts, etc. A limitation is that during tracking, the sensing device must be positioned within a foot or less of the cable to detect and indicate the transmitted signal's presence.

**Oscilloscopes**

Oscilloscopes are often used to view the analog waveform of network transmissions. Typically priced from \$3,000 to \$6,000, they are helpful in searching for abnormalities or hard faults. Unlike protocol analyzers, they cannot show

information content. The waveform is useful for diagnosing an electrical transmission problem and for recording a picture for trend analysis.

**Uses for Diagnostic Devices**

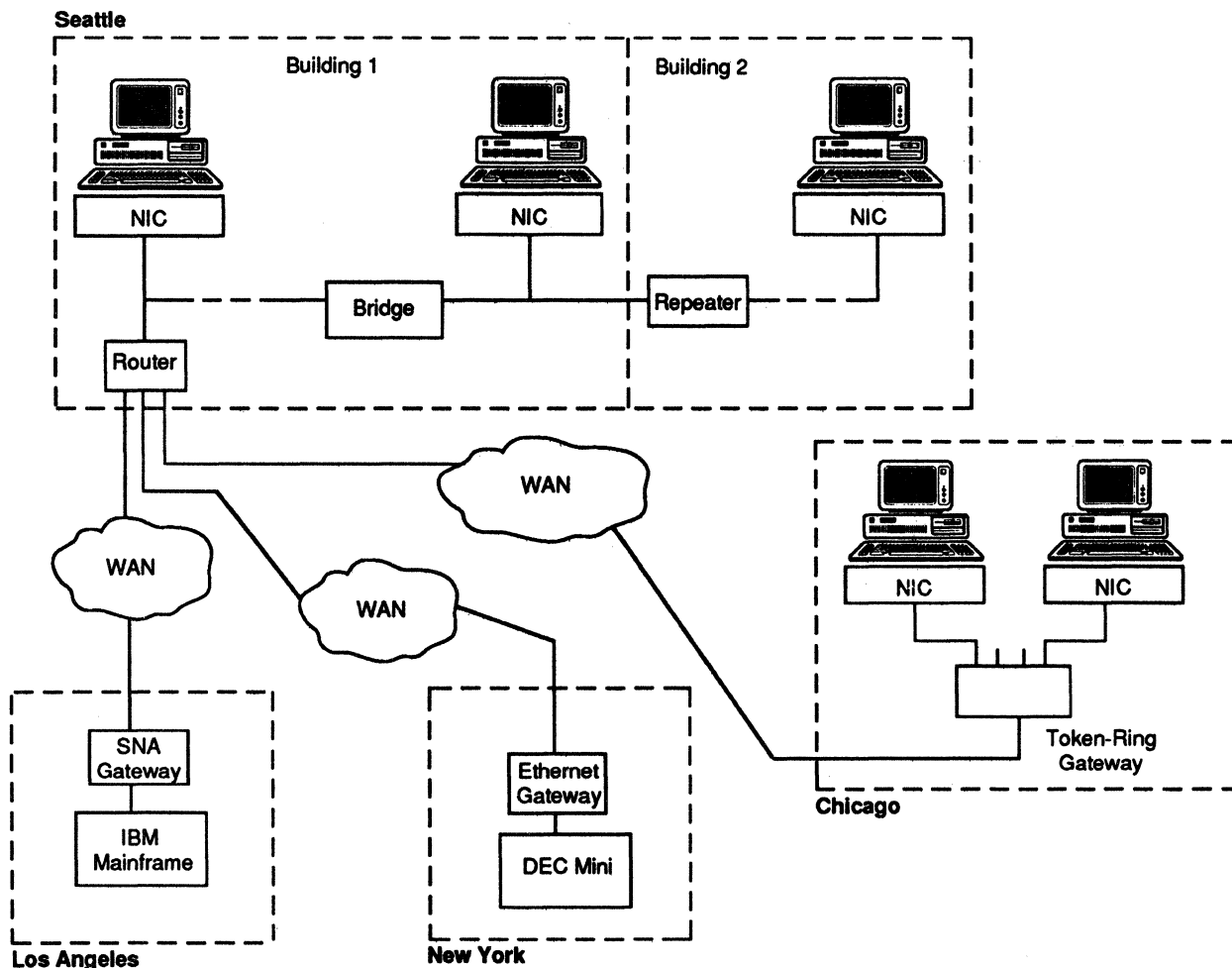
LAN diagnostic devices are useful not only for remedial work, but before, as well, to measure and calibrate LAN vitality. Table 1 illustrates a suggested spectrum of uses versus device type. This spectrum is a guide—the LAN manager should tailor a diagnostic program to best suit the network and its needs, both preventive and restoral.

**Defining Network Norms**

When a network is first installed and tuned, its performance parameters should be measured and recorded to establish the baseline for normal operation and growth. Then, periodically, these parameters should be re-measured to observe any trend variations. The need for establishing baseline performance increases with network complexity, and its importance intensifies for realtime networks serving commercial or industrial applications and for networks providing high-priority demand services. Nonetheless, a network supporting personnel developing software, preparing documents, or transmitting files of information needs reliability too.

The networks illustrated in Figures 1 and 2 normally do not justify the expense of sophisticated diagnostic devices. Once properly configured, with reasonable care exercised

Figure 3.  
Interconnected LANs



Linking widely dispersed LANs greatly increases network complexity.

not to abuse cabling and connectors, their simplicity is likely to keep them out of serious trouble. However, even for simplistic networks, extra tuning may be required to establish satisfactory norms. This could occur at the outset and again later as more software packages, more stations, and/or more shared peripheral devices are added. Such tuning is commonly accomplished by altering network logical parameters and carefully selecting application packages and the appropriate network operating system and options. All of these functions can be performed without the use of diagnostic devices.

As the number of devices in a network substantially increases, tuning the network requires more sophisticated information. Review your intended network with your LAN equipment vendor and LAN diagnostic vendor for suggestions, information, and guidance. It is particularly useful to conduct the review before finalizing your topology.

Some fundamental measurements are:

- average network load
- peak network load
- frequency of errors and retransmissions

- time and frequency of reconfigurations

Each type of LAN—token-ring, token-bus, and carrier sense—has its own particular set of physical and logical characteristics. Properly recognizing LAN-specific parameters, such as cable impedance, power loss between stations, maximum number of stations, and minimum and maximum signaling time between stations requires careful study and understanding to satisfactorily maintain the LAN. These specifications exist in the IEEE 802.X standards and vendor-proprietary specifications.

#### Preventive Diagnostics

Once network norms are established, periodic monitoring for value changes reveals trends resulting from changes in the network topology or uses and habits of users or software packages. These changes can either improve performance or reduce it. More memory, faster or larger server disks, more efficient client or server network software, and changes in some tuning parameters will tend to improve performance. More users at a given time, adding too many subnets through bridges or routers, and new CPU- or disk-intensive client application software will tend to reduce

performance. LAN managers should calibrate their networks at least every three months and immediately during major changes. For a static network, the three-month calibration can be extended to six months or even yearly.

As LAN managers develop a network performance history, any permanent degradation can be recorded and the history will help justify timely and appropriate network upgrades to maintain the customary performance level. If the budget cannot handle it, at least you have been forewarned, and you should consider a contingency plan for temporarily removing lower priority items should performance significantly degrade.

Preventive diagnostics, at a minimum, can reveal the following:

- intermittent problems (sometimes)
- cable weakness, poor connections, electrical interference over copper wires and cables
- logical loading resulting from changes in the NOS or application software
- loading from new traffic coming from other subnets or LANs

#### **Problem Solving: Nonintrusive Monitoring**

This form of monitoring adds no load to the network, nor does it take the network out of service. The information is analyzed in realtime, and any unusual conditions are captured for further analysis. Once the norms are established, each appropriate norm should be monitored, permanently documented, and compared for change.

Nonintrusive monitoring can be limited to observing problems appearing in frames on the network segment in which the protocol analyzer is located. Problems appearing only on the other side of a bridge or router will not be available for monitoring, unless a remote monitoring device that can send information back to the central protocol analyzer is positioned in the problem segment.

#### **More Problem Solving: Intrusive Testing**

Intrusive testing means injecting information into the network to determine whether it arrives at its intended destination, as well as observing network response such as the number of errors, error rate, and error patterns. In a wide area network, this usually means disrupting service. For a LAN, intrusive testing can be performed without significantly affecting the network. It is a recourse often used when nonintrusive testing is insufficient. It can be used to send test messages to devices located on any LAN segment or subnet.

#### **After the Fix: Postrepair Monitoring**

Once the problem is resolved, the network should be observed with a combination of nonintrusive and intrusive monitoring to be sure that the problem is indeed solved and that no other problems exist that were masked by the one(s) just fixed.

---

## **Planning, Training, and Resolution**

### **Planning for Diagnostics**

Preparing to diagnose problems means establishing a formal procedure. This ensures that testing is fully performed and records are kept orderly and complete for historical review when needed. See Table 2 for a suggested planning matrix.

### **Initial Planning**

The ideal time to plan for LAN problem diagnosis is during network design. By establishing a formal goal of adequate preparation for LAN problem diagnosis, and supporting that goal with appropriate budget and personnel at the start, a company ensures satisfactory LAN maintenance. It also recognizes the need to retain that capability as the LAN grows in scope and diversity. To prepare adequately, the group responsible for LAN maintenance will have a LAN administrator/manager (real or de facto) who understands LANs in general and is trained in all the individual components of the company's LAN.

Additionally, there will be a formal plan for dealing with problems, an appropriate set of diagnostic tools available, prior practice in their use, and—if necessary—arrangements for calling upon outside experts for special advice and possibly on-site help. The LAN diagnostics plan must be sufficiently versatile to deal with hard failures, intermittent failures, and subtle problems that may not surface for months.

### **Planning Never Ends**

LAN managers must continue understanding their network's scope and complexity and be prepared as the LAN industry continues its rapid advance. Improved by new vendor products, LANs offer one of the best combinations of personal and corporate-wide functionality to almost every aspect of an organization. Extending a LAN through an entire department is quite feasible and, in turn, offers the opportunity to connect LANs from different departments. For larger companies, the next logical step is to extend the LAN from one location to the next, often across wide area networks. Implicit in this growth is the interconnection of LANs of different transmission media and a multiplicity of operating systems and protocols.

### **Building Diagrams**

Whether dealing with a small or large network, record the physical location and identity of every element in the network, including cable type, length and route from one end to the other, and stations (both server and client). Also include information on such devices as bridges, routers, repeaters, amplifiers, cable connectors, and extenders.

### **Equipment**

Keep a record of the details for each station and cross-reference it with the building diagrams. Include manufacturer, model, additional boards, memory, disk, etc. Record the LAN operating system, version, options, and any other information pertinent to its identification. Record the contact name and phone number for each piece of equipment.

### **Placement of Network Elements**

To the extent possible, keep cables untangled and accessible and avoid locating them near potential hazards, such as electrical noise (for media other than fiber), water leaks, or in the path of equipment and carts.

### **Interconnected Networks**

Interconnected networks potentially introduce logical problems. The trend to use de facto standards, such as TCP/IP, helps greatly to minimize problems. Nonetheless, differences, subtle or otherwise, can cause problems when interconnecting two or more networks. Managers should plan as much as possible to either interconnect networks from a common vendor, to help assure that the products

interoperate, or to obtain assurances that the interconnection will work. Take advantage of your vendor's experience and help.

### Degraded Operation

In some cases it is possible to continue LAN operation when problems occur by running the network in a degraded mode. This provides time to locate and fix the problem without shutting down the network completely.

### Redundancy

Eventually, a LAN will malfunction. If your organization cannot tolerate losing service for the entire LAN or for segments, then the cost of redundant equipment must be evaluated. Redundancy is not only expensive, it is complex to implement because it must not permit a single failure of any component (including a database, which adds more complexity to a distributed system) or related suite of components that would crash the system. In many instances, the cost of good diagnostic devices, spare parts, and well-trained personnel is a more cost-effective substitute for redundancy. Also, depending on the network topology, partial redundancy of key elements may prove to be both cost effective and satisfactory in most cases.

### Training for Problem Diagnosis

The amount of training necessary depends on existing knowledge and skills combined with network complexity and tolerance for downtime. Training should not be limited to maintenance personnel; users need to understand their role, and management must recognize its capabilities, value to the organization, and degree of complexity by endorsing the LAN budget.

### Maintenance Personnel

The LAN manager is the primary (perhaps only) service person. LAN managers and all others servicing the LAN should understand its topology, each LAN element, and how to troubleshoot each component—both hardware and software—individually and as a subset of the network. One or more complete sets of manuals for all hardware and LAN software should be kept, safely, as reference. Formal training classes should be scheduled initially and then subsequently as changes in equipment, operating systems, or personnel indicate. These courses normally cost in the range of \$1,000 per person.

LAN managers would benefit by training other key personnel and establishing and maintaining a set of maintenance procedures tailored to the company's network. The right time for training is before the LAN is installed. Afterward, maintenance personnel can get comfortable with equipment during installation and start-up, and grow with it before the first emergency occurs.

For larger, more complex LANs, training should also include courses provided by LAN diagnostic equipment vendors. The cost of these courses varies with the type and amount of equipment purchased.

### Users

LAN users need training not only in its use but also in how to recognize problems, record symptoms, and report them to the LAN manager. Users should be trained to keep a log of all new software and their version numbers to correlate any unusual affects on the LAN caused by the software.

These effects may not show up for some time, and a permanent log is useful for accurately tracking trends, especially as they may interrelate to effects caused by actions and additions of other LAN users.

### Management

Management would benefit from a formal overview course on LANs. At a minimum, the LAN manager should prepare a half-day to one-day session to overview LANs in general and to deal with the specifics of the company LAN. This course is most appropriate for the next level of management and should be tailored to the backgrounds of those attending.

The LAN market is now experiencing technological growth, however, making it difficult for those not directly involved to fully appreciate the improvements, potential user benefits, and where LAN connectivity (via public and private networks) is headed. Upper management is likely to appreciate information on market activities as they relate to LANs and LAN diagnostics, occurring in T1/T3, metropolitan area networks (MANs), ISDN basic rate and primary rate interfaces, and the longer term Broadband Integrated Services Digital Network (BISDN). These topics can be covered more briefly in a half-hour presentation, supported by visuals.

### Diagnosing Problems

Diagnosing problems is a skill best structured through formal preparation. The following paragraphs suggest one approach. Naturally, each organization will develop and extend a process best suited to its own needs. In the following sections, a hard failure represents a physical failing; a logical problem represents a programming condition or an anomaly within the network; and an intermittent problem represents a situation caused by either.

### History Log

Maintain a log of all network-related changes and activity. This includes hardware, software, periodic maintenance, cable inspection or relocation, vendor demonstrations of software or test equipment, etc. Whenever problems are reported or suspected, include them in the log.

### Hard Failure

Hard failures, once located, are fixed by replacing the failed component with an equivalent. Depending on its location, the network topology, and the failed component itself, a hard failure may take down the entire network or only a subset. Swapping components, while unsophisticated, is the most common form of locating the problem. Bypassing sections of the LAN in an attempt to isolate a hard failure's effects is another obvious procedure.

These classic procedures work better for smaller, relatively simple LANs. For more complex LANs—recognizing that each LAN topology (bus, ring, or star) has its own peculiarities for troubleshooting—the LAN manager must have access to the diagnostic equipment best suited to the specific topology. Protocol analyzers can deal with different protocols, such as TCP/IP, OSI, DECnet, etc., and with the topologies of token-ring, token-bus, Ethernet, and Arcnet. Protocol analyzers are more expensive than other diagnostic tools and less likely to be available for small networks. For a hard failure, a simple voltmeter or TDR may suffice.

### Logical Problems

Logical problems can be obvious, such as two devices on the LAN with the same logical address; or, they can be extremely subtle. For example, operating systems marketed by different vendors and installed on the same or interconnected LAN can have information encoded differently in the same physical fields. Logical problems can be physically induced by broken cables or poor connections and be observed as traffic overload, high frame error rates, frequent Ethernet collisions, frequent loss of the token on token-ring and token-bus, or constant broadcasting by bridges and routers to establish topology.

Accurate problem analysis takes training and skill to understand the LAN's protocols and devices. Protocol analyzers are the diagnostic tools that can provide the information, but the user must be skillful in their use.

### Intermittent Problems

An intermittent problem can be caused by a broken cable or poor connection that, while impaired, still functions, causing errors at a random rate. Alternatively, an intermittent problem can be caused by excessive retries when heavy traffic across a bridge causes it to lose one or more frames in a file transfer, but only at a certain time of the day or week during other forms of high activity. Intermittent problems, by their nature, are often extremely difficult to resolve and they never seem to show up during observation. The capability of a protocol analyzer to capture data for playback or to inject test information into the network often helps in tracking and resolving these problems.

### Nonintrusive Approach

Unlike high speed wide area networks, such as T1, where inserting or removing equipment is undesirable because of the propagated perturbation, LANs are more tolerant of being observed. In some topologies, such as token-ring, protocol analyzers can be inserted into the ring (at a hub) with essentially no disturbance. The real intent of nonintrusive monitoring is to analyze the information flow without affecting it, to provide a true picture of LAN performance.

In some cases, dedicated diagnostic equipment can be permanently installed at central and/or remote locations and then used to observe conditions in the natural state. Analyzing this information can result in identifying problems and, ultimately, their source.

For a more complex LAN topology, particularly one spanning multiple LANs, remote devices combined with a central monitoring device help track problem sources. Some protocol analyzers provide this type of monitoring. For less complex topologies, a single device will be satisfactory.

### Intrusive Approach

The intrusive approach is necessary to search for the operational presence of stations on the LAN. By using a diagnostic device to broadcast for a station to identify itself, the technician can quickly determine the status of the network and selected stations. If the station in question does not reply, then stations adjacent to it can also be queried to confirm information routing and path performance, yielding checks on logical performance and physical state. Further intrusive testing can narrow down and isolate abnormal conditions. This form of intrusive testing allows the LAN to function at its current capability without severe stressing.

In the case of a totally collapsed LAN path, a TDR provides another intrusive method to detect and locate physical conditions. A more severe step is to disconnect individual LAN sections and use a volt-ohmmeter to check continuity and/or voltage levels segment by segment.

### After Repairs

Once repairs are made, technicians must continue monitoring to determine that the LAN has been returned to normal.

### Preventive Actions

The key steps in preventive maintenance have been discussed throughout this report. They are:

- proper planning;
- training;
- trend analysis through periodic diagnostics;
- keeping equipment in shape;
- having adequate spare parts available and in working order; and
- keeping an accurate log of changes of any significant nature.

Tracking changes on other interconnected LANs is considerably more difficult, but not impossible, since LANs only interconnect by agreement. Periodic discussions and other formal communications among LAN managers will help.

### Recommended for Problem Solving

#### Protocol Analyzers

The protocol analyzer has evolved steadily, along with communication networks, as a heavy-duty, all-purpose tool for solving problems at any OSI level in the network. Several companies have developed highly sophisticated LAN versions of protocol analyzers to handle almost every variation of a LAN operating system. Among the leaders in LAN protocol analyzers, Network General provides an extremely varied product line, based on the popular Sniffer. Other leaders include Novell and Spider Systems.

To diagnose LAN problems in depth, it is important that diagnostic devices deal with different LAN operating systems and their individual protocols. Effective use of a protocol analyzer requires extensive knowledge of the LAN's protocols, equipment, and the features and capabilities of the analyzer itself.

Protocol analyzers have several operating modes. In the capture mode, the analyzer stores an image of information frames, which pass the analyzer at a rate of thousands of frames per second. The stored image is separated into fields according to the frame protocol, displayed, and then the field content is analyzed (to a limited extent) and translated from ones and zeros into an English phrase. Thus, the analyzer determines exactly what information is being transmitted on a field-by-field basis.

Analyzers have other modes for continuous monitoring, filtering and trapping predefined data patterns, keying on a specific protocol layer, inserting messages into the LAN, and more.

The protocol analyzer is a portable tool. It serves as an independent service device and can also be used in combination with other diagnostic tools. Choosing a protocol analyzer requires an examination of available funds, LAN

complexity, personnel training and capabilities, future expansion projections, personal opinion, and vendor confidence. Given its relatively high cost and its potential diagnostic value, the purchase decision should only be made after careful evaluation, preferably with live demonstrations by vendor representatives.

### Network Management Systems

Network management systems do not perform diagnostics in the strictest sense; they function in the broader categories of fault detection and alarming, performance monitoring, and configuration control. When used in combination with diagnostic devices (such as protocol analyzers), they provide a powerful toolset to maximize LAN performance levels.

The network management system receives information from remote detection devices located at strategic nodes in the network. When the performance at any of these nodes degrades, it is reported to the network management center. Thus, personnel at the management center can determine the need for direct action, often before LAN users are aware of the problem. Some forms of network management permit reconfiguring the troubled node from the management center with backup equipment as a temporary solution until the specific cause is diagnosed and repaired.

Network management implementations range from relatively inexpensive integrated options within a LAN operating system to a suite of rack-mounted hardware at central and remote sites functioning over a separate network and costing hundreds of thousands of dollars. A network management system can be designed exclusively for managing the LAN and consists of one or many LAN operating systems, with single or multiple media types. The LAN portion of the network management system can also be a subset of a larger system for WAN network devices. The system can be located at the LAN site. For a large interconnected network, the network management system may have a network of its own to manage.

Large-scale network management systems require a trained staff to operate them. Even the simplest system requires training of LAN personnel on both the network management system and LAN diagnostic equipment to facilitate timely LAN maintenance.

Network management systems can view the entire network at a glance and zoom in to individual sections for a more detailed examination of performance conditions. This helps determine whether problems are widespread or localized both before and during a diagnostic/repair session. The same process can be applied after repairs have been done to determine whether the network has been restored to a normal state or additional problems have arisen.

---

This report was developed exclusively for Datapro by Arnold S. Cleff. Mr. Cleff is Manager of Project Engineering for Liberty Technologies. He has developed software and hardware for test equipment, process control, energy management systems, data communications, and network management systems. He was formerly Director of Systems Engineering for Digilog and Manager of Systems Software for Leeds and Northrup.

## Diagnosing Connectivity Devices

### Bridges

A bridge sends information to specific devices on the other side of the bridge. Defects in bridges can cause duplicate packets to be sent, resulting in extra traffic. In addition, a bridge may not send information that should be sent, or send information that should not be sent. Diagnosing bridge problems can be simple, although there is opportunity for subtlety as well. The most straightforward diagnostic method is to determine the destination addresses of information arriving at the bridge and information that has already passed through. Several diagnostic devices provide this capability, including protocol analyzers and network management systems.

The bridge must learn the addresses of devices on either side of the bridge to determine which traffic is to pass through. Therefore, a bridge may appear to be defective when, in fact, misleading information originates in another LAN device and confuses the bridge. The straightforward functionality of the bridge should not be underestimated when performing diagnostic procedures.

### Routers

Routers determine the best path for LAN packets to traverse. The functionality of a router is more complex than a bridge and therefore more likely to cause difficulty in problem diagnosis.

Determining whether packets are transmitted correctly can be done in a manner similar to that of a bridge. However, determination of the best path, as opposed to no path, is more difficult. Considering the dynamics of the LAN, this determination requires special knowledge of LAN performance and may be difficult to determine without a means to make this determination, such as a network evaluation software package. Therefore, accurate evaluation of the performance of a router is difficult.

Routers operate at OSI level three, the Network level. They are programmed in accordance with the LAN operating system protocol and direct received packets to the most efficient network path. Diagnosing problems at this level requires a device that can read and act on path addressing, such as a protocol analyzer, as well as the skill to understand the protocols and routing algorithms.

---

## Conclusion

LAN problem diagnosis, for most installations, remains manageable. For complex LAN installations, proper preparation and use of the latest and most sophisticated diagnostic devices may be the best (and only) solution to maintaining adequate LAN performance and uptime. ■



# Managing LAN-Based Services

## In this report:

Managing the Applications.....	2
The Windows Connection .....	2
Dealing With the Data .....	3
Smart Storage .....	3
Managing the Configuration.....	4

## Datapro Summary

As LAN technology matures, so evolves the role of the LAN in corporate settings. As organizations depend more intensely on networks, managers must learn to integrate central services, such as distributed applications, facsimile equipment, and database services, more efficiently. There are many training, network management, and data integrity tools available to help managers control these centralized services.

The typical PC LAN of a few years ago consisted of a file-and-printer server connected to a group of workstations. That picture is rapidly changing. LANs today can provide centralized communications, fax, and database services and support a wider range of distributed applications. The LAN is becoming the MIS data center of the 1990s.

The increasingly complex task of managing LAN-based services often falls to ad hoc administrators who acquire their skills on the fly, learning to deal with new demands as they arise. As enterprise-wide LANs materialize, that approach won't remain viable. It's critical that organizations protect themselves from uncontrolled organic growth.

I will offer some advice, drawn from my own experience, concerning common pitfalls and the methods and tools that can help solve them. My examples are DOS/Windows/NetWare-oriented, since that's my specialty, but the principles apply equally to other kind of networks.

This Datapro report is a reprint of "Control Central" by Jeffrey Sloman, pp. 175-180, from *Byte*, Vol. 16, No. 3, March 1991. Copyright © 1991 by McGraw-Hill, Inc. Reprinted with permission.

## Integrating Central Services

As your needs demand more services, they must be provided in ways that do not compromise existing ones. Unfortunately, when you add a new service, it may interact with the LAN in unforeseen ways. Take a look, for example, at adding an asynchronous communications server to an existing LAN.

Fresh Technology's Modem Assist is one solution to the problem of how to share a pool of modems on a network. It works in conjunction with a smart communications adapter, such as Arnet's Multiport/8 serial card. In principle, the process is straightforward: You just add a dedicated modem server to the LAN. But, as is typical of centralized LAN services, you have to make changes at the workstations, too. Modem Assist requires an INT 14 driver on the workstation, which redirects interrupt 14 calls across the network (INT 14 is the BIOS communications hook).

Most communications programs don't use INT 14; they write directly to the hardware. So, to use the modem pool, you have to acquire and install a program that communicates with an INT 14 interface. Hopefully, your existing communications program is available in a version that supports INT 14. DynaComm, Reflection, Procomm Plus Network Version, Crosstalk Mk.4, and CoSession LAN are programs that support INT 14.

If your communications program does not support INT 14, you will have to use the

14-oriented communications program that comes with the modem-pooling software (in the case of Modem Assist, it's MODEM.EXE) or acquire a third-party package that "speaks" INT 14. Either way, you are in for more software installation, and possibly more training, than you bargained for.

That's just the beginning, though. The INT 14 driver—either a TSR program or an installable device driver—may require too much memory or interact nastily with other TSRs that are in use at the workstation. If you rely on DOS-based multitaskers, such as Desqview and Windows, things can become even stickier. At present, these environments aren't always able to use DOS communications drivers reliably. A solution that saves money, but takes away Desqview's or Windows' ability to download files in the background, really isn't a solution at all.

The term *central service* belies the true complexity of the issue. While resources may be centralized either in the file server or in a workstation dedicated to providing a service, the infrastructure that grants access to central resources is distributed throughout the network. Managing that infrastructure can consume far more time and effort than managing the central services themselves.

---

## Power to the People

Integrating central services involves more than just hardware and software; it also involves integrating people, and that can be even trickier. In the modem server example, the justification for centralization probably includes reducing the number of telephone lines that are dedicated to modems. That plan assumes that phone lines will be shared on a contention basis—*contention* is the operative word. It's much easier to add features than to take them away. For instance, if you are familiar with a directly attached modem, you're likely to be upset the first time you receive a message that no modems are available.

Part of the transition, therefore, is dealing with people's expectations. The successful implementation of any new LAN feature depends on it. To manage expectations effectively, you must eliminate surprises. Thorough testing and documentation of any new service before it is made public is an absolute necessity.

Being trained how to use new services is important. Although it often happens, skipping the training step is a big mistake. No matter how smoothly the service is integrated, how reliable the hardware is, or how straightforward the software is, you must modify your behavior to some degree. Without training, resistance to new additions can be great.

Some organizations implement a simple policy: "no training, no service." In other words, you must demonstrate proficiency with a new service before you're allowed to use it. This is a good policy to implement if it's practical for you.

---

## Managing the Applications

Installing and configuring applications software is another integration task that can cause endless grief. Software vendors have only just begun to address the LAN market. Programs usually operate in the LAN environment—in some cases, they are aware of the LAN; in others, they are not—and may even provide some information on network installation.

But if you're concerned with central management, security, and data integrity, the situation is far from ideal. As

applications become more complex, so does their administration. Unfortunately, the lack of standard ways in which to design, install, and configure LAN-based applications makes working with each one a new adventure.

What files belong where? It's an obvious question, yet one that even LAN-oriented applications often don't adequately answer. Obviously, the executable file should go in a shared location. That might not involve just one file, however. There may be one or more supplementary overlays, the existence of which isn't always documented. Discovering all the executable components can itself be a trial-and-error affair.

Applications rely on one or more configuration files, the names and purposes of which, again, may not be documented. Generally, these should be distributed to private directories so that the shared program can adapt to individual preferences. But some configuration data may need to be public, too.

In the case of the modem server, for example, you don't need to maintain a copy of the list of installed modems and their associated settings at each workstation. Sorting out which configuration data should be private and which should be public can be a vexing task. Some programs are quite secretive about where and how they store and search for configuration data.

Still, it's worth the trouble to ferret these things out. When you add a new high-speed modem to the network, it's easier to update a shared configuration file than to distribute a new configuration file and verify that everyone receives it and installs it correctly.

Software that provides for device independence, such as Microsoft Windows and many CAD programs, adds another layer of complexity. These programs rely on drivers that adapt the software to particular hardware configurations. If you need to be able to log in at more than one physical location, a tricky coordination problem can result: trying to preserve each user's identity, as well as each machine's identity.

One solution involves a menu front end to the log-in procedure that prompts you to specify aspects of your hardware configuration: for example, VGA versus Super VGA. This is a poor technique, though, since a wrong choice can crash your system. It's better to record aspects of the machine configuration in DOS environment variables and then use batch files to select appropriate initialization files. In the DOS LAN environment, creative use of batch files is a requirement for effective administration.

---

## The Windows Connection

Windows has been both a blessing and a curse in network administration. Windows' consistent environment can greatly simplify training and technical support. The powerful memory management features of Windows 3.0 allow you to bypass limitations imposed by DOS (if you have the proper hardware). The available applications enlarge the scope and power of desktop computing. However, although Windows 3.0 provides more network support than any previous version, its focus is on simplifying access to network resources, not on expanding the ability to manage Windows and its applications on the LAN.

Ordinarily, a LAN administrator relies heavily on DOS-based menu systems that advertise available LAN services, and on batch files that launch and control the programs that provide those services. Under Windows 2.11, the usual practice was to provide menus that would invoke

Windows applications. But with Windows 3.0's new ability to multitask DOS sessions and the availability of more (and more powerful) Windows 3.0 applications, many users want to work exclusively within Windows.

Enter a new class of utility designed to facilitate integration. Windows Workstation from Automated Design Systems gives back the tools Windows 3.0 takes away. With a graphical menu system, a batch language, and a vastly improved print manager, Windows Workstation makes Windows 3.0 more manageable on a network than off. It's available for NetWare and Microsoft's LAN Manager; the latter version ships with LAN Manager 2.0.

The menu system provided with Windows Workstation is, of course, a Windows 3.0 application. That makes it familiar to Windows users. To specify options on a menu, you fill in a simple form, which in turn generates a script, or, if necessary, you write your own script in Automated Design Systems' MultiSet script language. MultiSet looks a lot like DOS batch language, but it adds Windows- and NetWare-specific features. The variable %LOGIN\_NAME, for example, returns your log-in name.

Even better, you can use MultiSet to insert user-specific entries into the Windows 3.0 configuration file, WIN.INI. For example, PageMaker uses an entry that looks like this:

```
[PageMaker]
Defaults=c:\pm\pm.cnf
```

You would probably want PageMaker to fetch its settings from the network rather than from the local hard disk. That way, some problems could be resolved over the telephone by running PageMaker at your workstation using those settings. Hard-coding a network path in WIN.INI doesn't separate the machine's configuration from individual preferences.

You can use MultiSet to create a PageMaker-launching menu entry that can tweak WIN.INI so that PageMaker will find PM.CNF in a public directory whose name is qualified by the %LOGIN\_NAME variable. Of course, everyone may not need a PM.CNF file. It might make sense to identify groups of users who share similar configurations—say, Art and Production departments—and use the script language to direct individuals to the appropriate configuration file by group.

## Dealing With the Data

Backing up the file server remains a vital responsibility. Until recently, this has been a thankless, labor-intensive job. Worse yet, even the most vigilant regime—daily backups—does not always protect critical data. Files can change several times a day, and each version may represent hours of work. If the system fails, the files you want most—the “working set” of files in active use—will be the very ones that yesterday's backup can't restore.

In addition, the usual method for ensuring backup integrity entails tape rotation. If you use multiple tapes in rotation—one a day, usually—you can maintain a history of versions and guard against the loss of a tape (or damage to a tape). But how do you keep track of the tapes? Maintaining a catalog and documenting what's stored on each tape require a lot of work. Moreover, a simple rotation scheme doesn't account for the need to archive particular files (i.e., storing special “frozen” versions of files on a separate archive tape).

What you really need are tools that can automate the backup chore. Several vendors provide automated data

management systems. ARCserve, from Cheyenne Software, runs as a value-added process or NetWare loadable module in a NetWare file server. Full or incremental backups can be performed interactively, or you can use NetWare's queuing services to schedule and dispatch unattended backups. The VAP/NLM implementation makes server backup a quick process: Files move directly from the server to the attached storage device, generating no network traffic. Jobs can be queued on a onetime-only basis or scheduled to repeat at regular intervals.

The VAP or NLM running in the server can also communicate with a TSR program running on a DOS workstation. That eliminates the need for a two-stage backup procedure: first from workstation to server, and then from server to secondary storage. Instead, workstation backups can be scheduled the same way server backups are. At the appointed hour, the VAP or NLM makes contact with the TSR in the workstation and moves the specified set of files through the network and straight onto tape.

You can also schedule unattended workstation backups or perform attended backups, so there's no need for locally attached tape drives. A single backup device can serve the whole LAN—although, as with the modem pool, it's on a first-come, first-served basis.

ARCserve also does a good job of logging each tape session. The administrator can see the whole log; users see only their own backup jobs. The beauty of ARCserve is that it overcomes human inertia. Once repetitive workstation and server backups have been scheduled, everything's automatic—almost. Someone still has to pop in a new tape every day, if you're using a tape rotation system.

Storage Dimension (San Jose, CA) makes innovative use of ARCserve. Bundled with the company's LaserStor erasable optical disk drive, ARCserve works the same as it does when it's used in conjunction tape storage. In this case, however, Storage Dimensions' drive provides fast, random access to the archived data. This combination protects against not only data loss, but also drive failure.

The LaserStor, being a direct-access device, can be used as an emergency replacement for the server's hard disk drive. It's slower than a fast hard disk drive, so it's not a full-time replacement for your ESDI or SCSI drive. But the ability to restore the network by switching to a 1-gigabyte optical backup device is an intriguing form of fault tolerance.

## Smart Storage

By combining features such as those in ARCserve with a rule-based expert system, Palindrome's The Network Archivist represents a forward-looking approach to automatic data management. Palindrome has structured TNA's functions into four groups, which are described below.

- *Backup.* TNA's backup function relies on “checkpoints” that resemble traditional incremental backups, but with a twist. For each file, you can specify that it be written to tape always, never, or—crucially—only when changed.
- *Archiving.* An archive, called a “save” by Palindrome, is a permanent copy of a stable file. TNA defaults to six weeks without change as an indicator of stability, although you can tweak this parameter on a per-file basis. TNA will make sure that the file is written to at least three different tapes in the tape rotation before considering it protected.

- *Restoration.* The ease of recovering a file's previous versions is one of TNA's most powerful features. You find the directory on a graphical tree, select the file from a list, press Enter, and choose the version you want. The checkpoint history is kept in an on-line database and tracked across tapes.

A catalogue of this type is essential if a backup system is to be useful for anything short of disaster recovery. TNA can also restore entire volumes. When it does, it uses its database to avoid restoring files that were intentionally deleted prior to the last checkpoint.

- *File-system maintenance.* TNA's most compelling feature is the automatic migration of unused files to tape. This "pruning" ability works by monitoring file access dates. Using rules defined by the LAN administrator, TNA determines when to move files from primary storage (i.e., the server's disk) to secondary storage. This can occur automatically, or it can require TNA prompt for confirmation.

A "phantom file" can also be left in place. A phantom file has zero length and carries the name of the migrated file. A TSR is loaded on the workstation. When you try to access such a file, the TSR pops up, explains that the file has migrated to secondary storage, and advises you to ask the LAN administrator to restore it.

The future of products like TNA is bright. One interesting prospect entails the use of a three-tiered storage system consisting of a primary magnetic disk, a secondary optical disk, and a tertiary tape drive. In this scenario, static files migrate from primary to secondary storage, and, if untouched for some additional period of time, they migrate from secondary to tertiary storage.

Ideally, you would be able to access files in secondary or tertiary storage as easily as you now can access files in primary storage. The only difference would be an occasional message announcing a delay while the system activates an archived file.

---

## Managing the Configuration

Traditional methods of backup and archiving don't address the need to preserve the work that has been done to configure the network operating system itself. Although any good backup program will save the network's system files (or the "bindery," in NetWare lingo) to tape, products like Cheyenne's NetBack save the logical configuration of the network in a form that allows for its reuse.

NetBack interprets the data that is stored in the bindery files and stores it as a description in what Cheyenne calls a "vault." The information is now more useful than a literal copy of the NetWare bindery. While it can be used to restore a server, it can also be used to add a new server to a network, endowing the network with the same configuration of users, groups, print services, log-in scripts, and security.

Programs like NetBack represent a welcome trend. Vendors of network operating systems have, understandably, concentrated on the operating systems themselves. Auxiliary network administration tools are typically weak. This has exacerbated the tendency to ignore key maintenance tasks.

Several vendors offer integrated utility packages designed to assist in network administration. Products like Fresh Utilities for NetWare from Fresh Technology Group and Cheyenne Utilities for NetWare from Cheyenne Software provide a number of programs aimed at managing the logical and physical configuration of the network and its servers.

One of the principal tools for managing the network is documentation. A record of each user's access rights, log-in script, group membership, and other pertinent data, along with a hardware configuration inventory, can save countless worker-hours in troubleshooting and disaster recovery.

As a manual process, compiling such documentation can be an overwhelming task. It's another example of something that ought to be done but is put off for lack of time. Once again, automation is the solution. Both of the products mentioned above produce extensive network configuration reports drawn from direct examination of the live network.

As the quantity of data grows and the number of services multiplies, it becomes harder and harder to manage a LAN manually. Products such as Palindrome's TNA herald an era of automated LAN management based on expert systems and AI techniques. The transparent operation of future network management systems will free you to work smarter and to focus on the what, rather than the how, of LAN administration.

The PC world in general is on the verge of maturing into what the mainframe world has become. Demands for better management, fault tolerance, and security, driven by a trend toward downsizing from mainframes and minicomputers, will help to fuel the development of tools and methodologies for managing tomorrow's LANs. ■

# Server Backup

## In this report:

Reliability .....	2
File System and Net OS Support.....	3
Tape Cataloging.....	4
Backing Up Local Drives .....	4
Server- or Workstation-Based Backup? .....	4
Backup Costs.....	4
Backup Procedures .....	4
Backup Cycles .....	6
Scheduling Backups .....	6
Four Software Approaches to Backup .....	6
Hardware and Software Recommendations .....	8
Choose What Fits Your Needs .....	8

## Datapro Summary

If LANs are not properly equipped with a backup system, the results could be devastating to success of an organization. According to recent studies, if a company experiences LAN downtime resulting in data loss for a period of ten days or more, it most likely will go out of business. The backup issues and procedures facing LAN administrators are often confusing. By following guidelines to select and implement backup systems, LAN managers can be assured that their company's data will remain intact.

With the advances in local area networking technology over the last few years, it would seem that proper backup should be as simple as painting by numbers—attach a tape drive to a file server or workstation, load some software, and start the backup. Unfortunately, it's far from being that easy. Many LAN administrators find out too late that their backup systems or procedures are inadequate or unreliable. The resulting data loss can be catastrophic. Creating a solid backup system—comprising hardware, software, and clear procedures—is more of a fine art than first meets the eye.

The value of the data stored on file servers is usually far greater than the value of all the LAN hardware and software components combined. Recent studies show that most companies that experience a data disaster lasting 10 days or more either are acquired by another company or file for bankruptcy within a year.

The major reason for catastrophic data loss is a lack of understanding of the requirements for preventing that loss. This is compounded by the fact that some of the

available backup hardware and software are unreliable. The backup issues facing the LAN administrator are not always clear-cut. Many tape drive vendors would like us to concentrate on a few specific issues, such as backup time, capacity, and mean time between failures (MTBF). However, it takes a broad comprehension of the issues—from backup media options and methods to restoration procedures—to select and implement a solid backup system. Let's take a look at these issues and how to address them.

## Backup Options

The first thing most LAN administrators consider when choosing a backup system is media. The most common media for backing up file server disks is magnetic tape. Tape offers relatively high capacities (currently up to 2.2 Gbytes per cartridge without data compression) and low cost.

Other options include optical drives and high-capacity removable magnetic disks. Optical disks have capacities of 300 to 600 Mbytes per side (only one side can be used at a time), while high-capacity floppies have capacities of 10 to 40 Mbytes per disk. Both media have the advantage of fast and easy file restoration, requiring only a DOS COPY command.

This Datapro report is a reprint of "The Fine Art of Server Backup" by Patrick H. Corrigan, pp. 26-42, from *LAN Technology*, Volume 7, Number 8, August 1991. Copyright © 1991 by M&T Publishing, Inc. Reprinted with permission.

However, the relatively low capacities of high-capacity floppies make them unsuitable for regular backup in most circumstances. As for optical media, the cost per megabyte is approximately 30 to 40 times that of high-capacity tape. This significantly higher cost makes optical-based backup with proper media rotation (how many sets of backup media you maintain and how often you rotate the backups) cost-prohibitive for many companies. On the other hand, the long shelf life of optical disks, estimated at 10 to 25 years, makes them well suited for file archiving.

There are three primary tape drive types used for personal computer and LAN server backup: quarter-inch cartridge (QIC), digital audio tape (DAT), and 8-millimeter (mm) cassette.

QIC drives, which have been around since the early 1980s, use the DC-600- and DC-2000-type data cartridges designed by 3M and have a capacity of between 40 and 500 Mbytes. QIC drives record data in a serial, serpentine fashion, writing multiple, narrow tracks back and forth along the length of the tape. Although there is a QIC standards committee, there is little standardization of the higher-capacity QIC drives.

QIC is mature and stable, but it is relatively slow when compared to other alternatives. In addition, given the fact that file server disk capacities are often 600 Mbytes or more, the capacities of QIC tapes make unattended, full-server backup difficult.

Backup systems that use 8mm tape are based on the D8 cartridge format developed by Sony Corp. Exabyte Corp. has licensed the technology from Sony and is the only manufacturer of 8mm data drives at this time.

These drives use a method of recording called helical scan in which a rotating read-write head records data in short, diagonal tracks on the tape. The current top capacity of 8mm drives is 2.2 Gbytes, which is sufficient for performing full backups on large-capacity file servers. Exabyte has also begun shipping a 5-Gbyte drive to OEMs.

DAT is 4mm wide and comes in a cartridge similar to the 8mm cartridge, but it is much smaller. There are two competing DAT formats: DataDAT and DDS (digital data storage). DataDAT is an update-in-place format that treats the tape as a random-access device similar to a disk drive. It has few supporters. DDS is a backup format promoted by Sony and Hewlett-Packard Co., and it has been adopted by a large number of tape drive manufacturers and integrators. DDS drives are manufactured by a number of vendors, including Archive Corp., WangDAT Inc., WangTek Inc., and HP. DDS is emerging as the DAT format of choice.

Like 8mm, DAT uses helical scan recording. DDS DAT currently provides 1.3 Gbytes of capacity with 60-meter tapes and 2.0 Gbytes with the newer 90-meter tapes. The DDS specification includes a high degree of error detection and correction. In addition, DDS supports a high-speed file find mode that lets users find a file on tape in an average of 20 seconds.

DAT, especially the DDS format, provides the capacity for performing full backups on large-capacity file servers. In addition, DAT's ability to quickly find and restore files, its multi-vendor support, and the imminent arrival of DAT changers make it an attractive choice for LAN backup.

There are two primary classes of tape system vendors: manufacturers and tape drive integrators. Manufacturers such as Archive, WangTek, Exabyte, and WangDAT build the drive mechanisms. Tape drive integrators such as Mountain Network Solutions Inc. (formerly Mountain

Computer), Maynard Electronics, Tecmar, and Palindrome Corp. combine tape drives with host adapters, cables, software, and value-added services.

Although it is possible to buy drives, controllers, and software separately, you will probably receive better support when you buy a package from a tape drive integrator. Backup systems are the wrong components to save money on.

---

## Capacity and Backup/Restore Times

With file server disk storage constantly increasing, backup system capacity is a major issue. Today, 1-Gbyte hard disks are common, and it is not unusual to find file servers with multi-gigabyte disk capacities. A LAN server should be backed up when all users are off the system; in most cases, this means at night. Ideally, a backup system should have a large enough capacity to back up an entire file server, which allows for unattended backup. (See the sidebar "What to Look for in a Tape System" for guidelines for selecting a tape backup system.)

Fortunately, tape capacities have kept up with server capacities. For example, many 8mm drives from Exabyte have a 2.2-Gbyte capacity, and some can hold 5 Gbytes. DAT drives have a 2-Gbyte capacity with the newer 90-meter tapes. In addition, using data compression can increase the capacity of backup systems. Although compression ratios vary for different file types (and can go as high as 98 percent), on average, compression can double backup media capacity. Compression schemes, however, usually lower backup system speed.

Changers for DAT tape drives that hold five to 12 tape cartridges are going to hit the market soon. Advanced Digital Information Corp. (ADIC) has been shipping prototype DAT changers to customers since May. ARDAT is shipping changers and development kits to backup software developers.

With small local hard disks, excessive backup or restore time is usually more an inconvenience than a major problem. However, with high-capacity file servers, being able to properly back up within a limited time period is critical, as is the ability to restore a single file or an entire file server quickly.

A backup system should be able to back up required files within the amount of time allotted. Even at a relatively fast 11 Mbytes per minute, it would take a single DAT drive with a changer approximately 15 hours to back up a 10-Gbyte file server.

As disk capacities grow, multiple tape drives might be needed to back up a file server in a timely fashion. (Although you can operate multiple tape drives off different networked computers today, I know of no software that will support multiple drives attached to one controller and operating in parallel to back up different disk volumes.)

---

## Reliability

To prevent data loss, you must have reliable backup systems and reliable, tested backup and restore procedures. Reliability means that your backup system provides error-free backups and error-free restores. This is a hardware, software, and procedural problem. The best backup software in the world won't solve problems caused by unreliable hardware. In addition, if your backup procedures are inadequate, then critical data may not be properly backed up.

## What to Look for in a Tape System

Purchasing a tape system is easier if you know what to look for. Here are a few guidelines to help you select a tape backup system:

- The tape system should provide file-by-file backup and restore. It is inefficient (and sometimes impossible) to restore an entire tape to retrieve one lost file, yet this is what streaming, or disk image, tape drives require. (These tape drives take a snapshot of the disk drive's contents on a block-by-block basis.) Disk image restoration may be fine when you need to restore data after replacing a defective hard disk. However, most restorations are the result of lost or damaged files, not damaged disks.
- The tape software should provide multiple methods of

selecting files for backup, including exclusion/inclusion by file and/or directory specifications, date and time, and changed files. Restore options should include selection by file specification and/or directory, selection of specific files during restore (as opposed to a preset list), and date and time.

- The tape system should be able to back up a file server's security and system files, rights information, and all file and directory attributes.
- The system should be able to properly back up all file types stored on a file server. Although some network operating systems support multiple file types, such as DOS, Macintosh, OS/2, and UNIX, most backup systems still only back up DOS files.

• The tape drive and software should allow for unattended backups. If this function is not included, using keyboard macros as well as shareware and public-domain programs such as CHK-TIME, SLEEP, and WAITUNTL often do the same thing. These programs are available on CompuServe and other bulletin board systems.

• Ideally, the tape drive should have a capacity as great or greater than your file server. This allows for unattended backup of the complete system.

• The tape drive should provide error detection and correction during backup and restore. Error detection should be automatic, not an optional operation to be performed after backup. Optional tasks, especially if they are time-consuming, are rarely performed. If the tape software provides a manual verification process, add the time required for the process (usually the amount of time required for a backup) to the backup time.

• The tape drive should be able to maintain constant speed to compensate for variations in tape tension. Some tape systems require regular retensioning of tapes, which rarely, if ever, gets done.

• Mechanical wear should not affect tape-to-head alignment. Recording head misalignment due to wear often means that a tape recorded on one drive will not be readable by another similar drive.

• If unrecoverable errors occur on the tape, the tape system should provide a means to bypass defective sections of tape to continue a restore.

• The vendor should guarantee that you can interchange tape cartridges between two or more tape drives of the same model.

• The vendor should provide recovery services to attempt restoration from damaged tapes. (ADIC, for example, provides this type of service.)

Backup hardware should not break down often. Although MTBF figures can tell you something, they are usually a "best guess" estimate by the manufacturer. In addition, MTBF figures are usually quoted at a rated duty cycle, or percentage of time that the unit will be in use. For example, a rating of 20,000 hours MTBF at a 10% duty cycle means that if you only use your backup device 10% of the time (or 2.4 hours a day), it should last for 20,000 hours before needing repair or replacement. Unfortunately, many backup systems are used at much higher duty cycles than they are rated for, often lowering the real MTBF.

Backup hardware should also provide extensive error correction and detection and be able to bypass or block out defective sections of media during the backup process. Backup media can be damaged between the time files are backed up and the time files are restored. If the media is damaged, the backup hardware should be able to read beyond the point(s) of damage and continue data recovery, even if the data in the damaged areas is lost.

Although many backup devices provide high levels of error detection and correction and the ability to recover data beyond the point of a major media error, much of today's backup/restore software, primarily for tape, cannot take advantage of this capability. In addition, many backup software packages provide inadequate verification capabilities and don't compare the files on the tape to the files on the disk.

However, even the most reliable hardware and software will not make up for improper or inadequate backup and

restore procedures. Backup procedures must be properly planned and should be documented. Unfortunately, even the best backup procedures won't guarantee that you will be able to restore data in a correct and timely manner. Many system administrators who have well-planned, well-executed backup procedures become completely lost when it's time to restore files. In many organizations, restore procedures have never been tested, there are no written guidelines, and all too often the software required to restore files after a server hard disk crash cannot be located.

### File System and Net OS Support

Although file servers can support client machines running different desktop operating systems, most backup software does not work properly with multiple file types. Only a few of the backup programs available for Novell Inc. NetWare file servers, for example, can effectively back up Macintosh files. Finding software to properly back up OS/2 and UNIX files from a NetWare server is even more difficult. Even if the files can be properly backed up, extended attributes or security information added to files or directories by the network operating system may not be preserved properly.

A backup system should have the ability to properly back up and restore all files being stored on a file server, including system files and any extended file and directory attributes, file access dates, and file and directory rights information.

If you have file servers running different network operating systems, you may need a separate backup software package for each network operating system. Currently, most backup software is designed to back up one type of network operating system. There are many backup systems available for NetWare, for example, but few for Microsoft Corp.'s LAN Manager.

Sytos Plus from Sytron Corp. will back up NetWare and LAN Manager servers. Even here, however, you cannot back up both server types in the same session. In addition, because most backup software only supports a limited number of host adapters and tape drives, you may need different hardware for each network operating system.

### Tape Cataloging

As file server storage increases, finding out what data is on which tape becomes more of a challenge. Many backup software packages offer some form of cataloging, but for the most part, this is just a method of locating tapes by label, date, or description, then locating files once a tape has been selected.

Fortunately, there are a couple of exceptions. The most notable is The Network Archivist (TNA) from Palindrome. TNA provides an on-line file history database that contains the archiving history for all files on a volume, allowing you to find and restore any backup version of a file.

### Backing Up Local Drives

What about users backing up their local drives? A simple rule applies here: They don't. Thus, the ability to provide centralized backup of local drives has become a big issue. Most of the major LAN backup systems provide a way to back up DOS-based workstations. However, there are several potential problems with workstation backup.

First, most local-drive backup schemes provide very little security. They require that a terminate-and-stay-resident (TSR) program be run at the workstation to be backed up; this TSR allows the backup station to access the workstation's drive(s). Unfortunately, because the workstation is up and running and NetWare IPX or NetBIOS messages are being sent across the LAN, it is easy for an unauthorized but knowledgeable user to access that workstation also.

Second, backing up multiple local drives is time-consuming. Because of this, if there is a large number of local drives to back up, it is often impractical to back them all up every night.

Third, it may be difficult to enforce the procedures required for effective local-drive backup. Just as it is difficult to get users to perform proper backups themselves, it is also difficult to get them to follow procedures that allow centralized backup, such as running the required workstation software. In the long run, it is probably better to encourage users to keep their data on file servers which are (hopefully) always backed up properly.

### Server- or Workstation-Based Backup?

Some backup systems are connected to a file server, while others are connected to a network workstation. There are advantages and disadvantages to both approaches.

There are two key advantages to server-based backup. First, backup and restore is faster for the host server (the server to which the backup system is attached) than with workstation-based tape systems. Second, if the backup

program is a server-based application, scheduled and unattended backups can usually be accomplished without a user being logged in at a workstation. This can provide a higher level of security than some workstation-based approaches.

On the other hand, there are a few disadvantages to server-based backup that you should consider. If there is a problem with the backup hardware, especially the host adapter or controller, you may need to power down the server. If there is a problem with the server-based backup software, it could affect other server processes.

Although server-based backup can provide faster backups and restores for the host server, this advantage is lost when you need to back up servers other than the host. Another problem with server-based backup is that, in the case of a file server crash, the backup software must be reinstalled before files can be restored. A final point: Many server-based backup systems do not have a workstation-based equivalent to allow you to restore to a local drive or to a server running a different network operating system. This can create problems if you need to restore data to a dissimilar server in an emergency or if you want to migrate to another network operating system.

Workstation-based backup offers several advantages and a couple of disadvantages. On the plus side, backup hardware problems do not affect file server processes or performance. Backup components can be removed and replaced without affecting the file server. In addition, files can easily be restored to other servers and local drives. In the case of a server crash, restore software can be run from a local drive, so it doesn't have to be loaded on the server before files can be restored. If a restore is done on a workstation local drive, the files restored are only accessible by the user of that workstation.

A serious limitation is that server backup using a workstation-based backup system is generally slower than with server-based systems. Also, the backup station must be logged in to the file server in order to back up. Overall, however, I prefer workstation-based to server-based backup.

### Backup Costs

Even with the increasing importance of the data stored on LAN file servers, many companies balk at the \$4,000-to-\$8,000 price tag of a high-capacity backup system. If you calculate the cost of reconstructing even one day's worth of data in most organizations (if it is even possible to reconstruct), the backup system price is a relatively trivial cost.

Media costs for today's high-capacity DAT and 8mm tape drives are far lower per megabyte than media costs for low-capacity QIC tape drives. This is partly because 8mm and DAT can hold more data per foot of tape and partly because 8mm and DAT tapes are high-volume consumer items, while QIC tapes are specialty items. Even without the economies of scale, QIC tape cartridges are more expensive to manufacture.

By establishing proper backup procedures and using automated, unattended backup schemes, the labor costs for backup can be kept to a minimum. This is one area where purchasing high-capacity backup devices can produce immediate dividends.

### Backup Procedures

When selecting your backup system and planning your backup procedures, consider the effort and time it will take



# Secure Workstation-Based Backup for NetWare

Unattended backup for a workstation-based backup system can be set up relatively securely by following the procedures outlined below. Although these procedures are specific to NetWare, the principles can be applied to other LAN operating systems.

1. Create a password-protected user account with supervisor equivalence.
2. Using NetWare's time restriction functions, set time restrictions to allow that supervisor account to be logged in only at the times necessary for backup (see the figure).
3. Set station restrictions so this account can only be logged in from the backup station.

Allowed Login Addresses	
00000100	0000F020272E

4. Create a login script for the backup account similar to the following:

```
BREAK OFF
DOS BREAK OFF
#CASTOFF /A
#KBOFF
MAP F:=SYS:
SYSTEM/TAPE
MAP G:=SYS:
MAP H:=SYS2:
#CHKTIME 02:15:00
-w -p
EXIT "ARCHIVE"
```

In this example, Ctrl Break is disabled using the login script commands BREAK OFF and DOS BREAK. Message reception from other stations and the system console is disabled using the NetWare command CASTOFF /A.

The keyboard is locked using the public domain utility KBOFF. This utility is available from the Novell NOVA forum on CompuServe, and it is part of a compressed file named KBD.ZIP.

The directory containing the backup software is mapped:

Allowed Login Times for User ARCHIVE																							
	AM										PM												
	1	2	3	4	5	6	7	8	9	10	1	1	1	2	3	4	5	6	7	8	9	10	11
Sunday																							
Monday																							
Tuesday		*	*	*	*	*	*	*	*	*													
Wednesday		*	*	*	*	*	*	*	*	*													
Thursday		*	*	*	*	*	*	*	*	*													
Friday		*	*	*	*	*	*	*	*	*													
Saturday		*	*	*	*	*	*	*	*	*													

Saturday 11:30 pm to 12:00 am

```
MAP F:=SYS:
SYSTEM/TAPE
```

The volumes to be backed up are mapped:

```
MAP G:=SYS:,
MAP H:=SYS2:
```

Execution is paused until a predetermined time using a shareware utility called CHKTIME, also available from the NOVA forum. It is part of a compressed file called TIME-R.ARC.

This login script exits to a batch file that performs the backup, then reboots the workstation. This example uses the public-domain program BOOT.COM to reboot the bad workstation. BOOT.COM is available on many

bulletin boards in a file called BOOTET.ARC or BOOTET.ZIP. Other programs, such as REBOOTB.COM from PC Magazine, provide a similar function.

The batch file should be similar to the following:

```
TAPE (parameters)
BOOT COLD
```

This batch file performs the backup and then performs a cold boot of the workstation, effectively logging the backup account out.

to restore data—as well as the complexity of the task—in a worst-case situation. Backups are done at leisure, while restoring is done in a panic. Backups are usually (or at least should be) a matter of routine. Restoring is often required at inconvenient times. You should be able to restore files quickly and easily. (The complexity of restoration varies among products.)

Backup and restore will go more smoothly if you follow these rules of thumb:

- Document your backup procedure. If procedures aren't written down, they're forgotten.
- Document your restore procedure. Be sure the documentation is clear and easy to follow, as file restoration at times may need to be performed by someone with little experience.
- Test your backup system regularly. Since backup systems malfunction, you need to test your system on a regular basis.

Unfortunately, there really is no good way to test a backup system. (Deleting all the files from your server and attempting to restore them might be a good test, but it is a little dangerous!) Try the following approach:

- Set aside an area on your server for test restores.
- Restore the last file on tape to your server restore area and do a file compare with the original using the DOS COMP command, which results in a byte-for-byte comparison.
- Restore representative files of different types to the restore area. Test them by doing file compares with the originals. If they are program files, make sure they execute. If they are data files, make sure they can be opened properly by their applications. Make sure any attributes are properly restored, as well as any dates associated with the file (such as the NetWare Last Accessed and Last Archived dates).

## Backup Cycles

The most effective method of backup is to back up all files, programs, and data every day. This will give you the greatest ability to restore files quickly because all files on the last tape are the most current. This approach also uses the greatest amount of backup media and time.

If you have time constraints, or if you use expensive media, then you might want to perform incremental backups. This means that every day you back up only those files that have changed and once a week you back up all files. This will result in multiple copies of full backups at various stages. This is a workable solution. However, if you need to restore, it usually requires that you go back to the last full backup and then restore each partial backup in order, which makes restoring a slower process.

One potential problem with incremental backup is that the more tapes you have to rely on to accomplish a restore, the more likely that one of them will be unreadable. Some backup systems allow you to back up all files that have changed since the last full backup. This approach means that only two tapes are required for a full restore. Incremental backups have another problem: When you perform a full restore, files deleted since the last full backup will be restored also.

Palindrome uses a third approach to backup cycles. TNA was designed with a complex set of rules to provide what Palindrome calls intelligent backup. TNA's database keeps track of previously backed up files, which are categorized as evolving or stable. Evolving files (files that have changed) are backed up regularly, while stable files (files that have not changed in a specified length of time) are not backed up after they have been saved to three separate tapes. Although this approach saves tape and backup time, it means that full restores can be more complex and time-consuming than with other approaches.

If you are using a traditional full or incremental approach (as opposed to the TNA approach), the following rotation schedule should provide reasonable protection:

- Back up files (all or incremental) every day.
- Back up all files at least once a week.
- Keep your daily backups for at least a week.
- Keep your weekly backups for at least a month.
- Keep your monthly backups for at least a year.
- Archive a tape at the end of each quarter.
- Back up critical files to alternate media on a regular basis. Critical files are those that cannot be reconstructed or restored from original diskettes. These files should be backed up to local drives, another file server, or an alternate tape drive as insurance against a catastrophic tape drive failure. If space is a consideration, utilities such as PKWARE Inc.'s PKZIP can be used to compress the files.
- Perform additional milestone backups before and after potentially dangerous or important events occur, such as accounting year-end reports, software upgrades, or office moves. Use alternate-media backup of critical files as described above.

Backup tapes should be stored off-site on a regular basis to guard against possible disasters such as fire, theft, or earthquake. Backups of original software and weekly backups are good candidates for off-site storage. Smaller companies sometimes store their backup library in a safe deposit box.

Another alternative is to contract with an off-site storage company, which will pick up the backups regularly and store them in environments that are well-protected against disasters. You arrange a regular rotation of tape exchanges with the company, and the storage company returns the tapes immediately when you need to restore.

Unless specifically designed for magnetic media, so-called fire-proof safes do not provide sufficient protection because they are only intended to keep paper from burning. Unfortunately, tape and tape cartridges melt at a much lower temperature than the flash point of paper.

## Scheduling Backups

Although backups can be performed while users are on the system, most backup systems will not back up open files, so it is usually best to schedule backups during off-hours. Most backup systems can be programmed to execute on predefined days and times.

To ensure the backup of all files, you should make sure all users are logged off the network and then disable LOGIN. If your network operating system allows you to restrict the times that users can be logged in, use this capability to make sure all users are logged off during the time period reserved for backup.

Although some packages will back up open files, it is a very risky practice, especially with database files. If you back up database files while they are open, you may lose synchronization between related files.

## Four Software Approaches to Backup

The following are brief descriptions of several of the leading backup software packages for LAN file servers. The discussions of Cheyenne Software Inc.'s ARCserve, Mountain's FileSafe and FileTalk, Maynard's MaynStream, and Palindrome's TNA are intended to provide an overview of the features and capabilities of these products.

### Cheyenne's ARCserve (Version 3.0)

ARCserve is a server-based backup system for NetWare. ARCserve runs as a value-added process (VAP) on NetWare 2.x servers and as a NetWare Loadable Module (NLM) on NetWare 3.x servers. ARCserve provides for attended and unattended backup of DOS and Macintosh files stored on NetWare file servers, and DOS and OS/2 files on network workstations. It does not support the OS/2 High Performance File System (HPFS).

ARCserve provides the same menu interface used by the NetWare utilities and provides a considerable amount of flexibility for selecting files for backup. In addition, ARCserve uses the NetWare queuing functions (the same ones used for print service) to track and manage multiple unattended backup sessions. ARCserve's scripting capability allows frequently used procedures to be saved and reused. Scripts also allow you to define an automatic repeat interval, allowing for repetitive, timed backup.

ARCserve does have several limitations. Because it is server-based, files cannot be restored unless the ARCserve host server is active. In case of a server crash, files cannot be restored to alternate locations, such as to a local drive or another file server. It also means that if the server operating system is changed from NetWare to something else, the

files cannot be restored unless another NetWare server is available. (Cheyenne does have a version called ARCserve Solo that runs on a workstation, but it is only available through OEMs for a limited number of tape drives.)

Another drawback is that full administration requires access to the host file server as well as a workstation. Also, error message documentation is very limited.

ARCserve supports a wide range of tape drives and controllers. It is available from Cheyenne, its dealers and distributors, and OEMs such as Tecmar, HP, and Irwin Magnetic Systems Inc.

### Mountain's FileSafe (Version 5.2.2) and FileTalk (Version 1.10E)

FileSafe, Mountain's workstation-based backup software, has few bells and whistles, but it does have a strong reputation for reliability. FileSafe will back up NetWare-, MS-Net-, and LAN Manager-/LAN Server-based file servers. (LAN Manager/LAN Server rights and security are not backed up, however.)

As for file formats, FileSafe will back up DOS, Macintosh, and OS/2 files (except HPFS files) on file servers, and DOS and OS/2 files (again, except HPFS) from the workstation it is attached to. FileSafe's menu interface makes it easy to select files for backup and restore. FileSafe's restore menu, for example, displays all save sets on a tape, letting you select the one you want to restore from. In addition, FileSafe's command-line mode allows backup procedures to be fully automated using batch files.

FileTalk is Mountain's facility for backing up local hard disks on network workstations; it only backs up DOS files, however. FileTalk supports Novell's IPX protocol as well as the NetBIOS protocol used by many other network operating systems. FileTalk's workstation software, which runs as a TSR program, allows any or all drives on a workstation to be published and uniquely identified. FileTalk can locate and back up all published drives on the backup workstation.

FileSafe does have some limitations. For one, it has no library or catalog capability. Mountain has announced its Data Management System (DMS), which will probably replace FileSafe for LAN applications and, according to the company, features capabilities far beyond cataloging. DMS is designed to provide complete on-line and off-line file management across a multi-platform LAN. According to sources at Mountain, portions of DMS should ship before the end of the year.

Because FileSafe is workstation-based rather than file server-based, care must be taken to maintain security when performing unattended backups. (For tips on how to provide such security, see the sidebar "Secure Workstation-Based Backup for NetWare.") Another drawback is that FileSafe does not back up all OS/2 extended attributes and does not restore the Novell SMODE attribute.

FileSafe is primarily available for Mountain-supplied tape drives, although Mountain will provide retrofit kits for some other vendors' drives. The company supplies drives with capacities from 40 Mbytes to 2.3 Gbytes. Mountain products are available through its dealers and distributors.

### Maynard's MaynStream (Version 3.1)

Maynard's MaynStream is a straightforward, easy-to-use, workstation-based backup system for LAN servers and local hard disks. It is available for NetWare-, LAN Manager-/LAN Server-, and MS-Net-based LANs, and it provides

## For More Information

ARCserve software and controller card ..... \$1,995  
**Cheyenne Software Inc.**  
 55 Bryant Ave.  
 Roslyn, NY 11576  
 516-484-5110  
 800-243-9462

**Mountain Network Solutions Inc.**  
 240 Hacienda Ave.  
 Campbell, CA 95008  
 408-379-4300  
 800-458-0300

FileSafe 1200 ..... \$5,995  
 with 1.3-Gbyte 4mm DAT drive  
 FileSafe 2100 ..... \$7,490  
 with 2.2-Gbyte 8mm Exabyte drive  
 FileSafe 7500 ..... \$3,995  
 with 525-Mbyte DC-6000 drive (all models include FileSafe, FileTalk, and SCSI adapter)

MaynStream software, controller card, and tape drive ..... \$1,295-\$7,495  
**Maynard Electronics**  
 460 E. Semoran Blvd.  
 Casselberry, FL 32707  
 407-263-3500  
 800-821-8782

The Network Archivist software only .. \$995-\$1,195  
**Palindrome Corp.**  
 850 E. Diehl Road  
 Naperville, IL 60563  
 708-505-3300

limited support for Banyan Systems Inc.'s VINES. MaynStream will back up DOS, Macintosh, and OS/2 files from file servers. (HPFS backup is not supported under NetWare.) It will also back up DOS files from local drives on the host workstation and from hard disks on network workstations.

MaynStream's menu interface is very easy to use. Under NetWare, it gives you the option to select specific server volumes by drive letter or to select servers and volumes by name.

IsleLAN, MaynStream's workstation backup facility, supports both Novell's IPX protocol and NetBIOS. IsleLAN consists of two pieces of software. On workstations to be backed up, IsleLAN runs as a TSR program and allows any or all drives on a workstation to be published and uniquely identified. The part of IsleLAN running on the backup workstation can locate and back up all published drives on the network workstation.

Among its limitations is that MaynStream, when combined with IsleLAN, is a RAM hog. IsleLAN on a workstation requires nearly 40 Kbytes of RAM. If you plan to back up local drives, the backup station needs to be an 80386 or 80486 machine outfitted with memory management software such as Quarterdeck Office Systems' QEMM.

Another drawback is that MaynStream does not alphabetize files and directories. This makes it difficult to locate specific files and directories. By default, MaynStream's Tape Directory and Catalog Maintenance screens scroll directories non-stop, which makes it impossible to view them. This problem can be solved by starting MaynStream with the /PAGE parameter.

MaynStream is primarily available for Maynard-supplied tape drives, although Maynard will provide retrofit kits for some other systems. Maynard supplies drives with capacities up to 2.3 Gbytes. Maynard products are available through its dealers and distributors.

### Palindrome's TNA (Version 2.0)

As noted earlier, TNA offers a unique approach to network backup by using a complex set of rules to ensure comprehensive backup. TNA, which is workstation-based, only supports NetWare. It will back up DOS and Macintosh files on the file server and provides for backup of DOS workstations on the network.

TNA is designed to cut down the amount of time required for backup as well as the amount of backup media required. In addition, TNA's extensive cataloging capability keeps a file history database, which records all versions of a file that have been backed up to tape.

TNA provides for multiple backups of each version of a file and for automatic file migration. File migration means moving files that have not been accessed in a certain period of time off the disk and onto tape. TNA uses a complex tape rotation method called "the tower of Hanoi," which uses a different rotation period for each tape set and appends new backup sessions to existing tapes. TNA will back up open files but, because of the potential problems involved, marks them as suspect.

TNA has some limitations. For a full-volume restore, the System Control database must be restored first, increasing the time and complexity of restore. If a full restore is required, the instructions for reloading the software must be followed exactly, or you could easily lose all of your archiving history.

In most cases, at least three sets of tapes are required for a complete restore, again adding to the complexity of the restore process and the time required. In addition, restoring from multiple tape sets increases the chances of encountering a defective tape.

Palindrome supplies complete backup systems, with capacities up to 2.3 Gbytes, as well as retrofit kits for other vendors' drives. TNA is available through palindrome's dealers and distributors.

## Hardware and Software Recommendations

In my opinion, DAT drives using the DDS format have a lot to offer for LAN server backup. DAT has the highest level of error detection and correction of the available tape backup options. It has a capacity of 1.3 Gbytes to 2 Gbytes, depending on the length of the tape (60 meters or 90 meters), and cost-effective tape changers will soon be available to further increase capacity.

In addition, a standard for data compression, implemented in firmware in the tape drive, has emerged. Compression can increase capacity by 100 percent or more, depending on the types of files being compressed. Another advantage is that there is more than one drive manufacturer (whereas 8mm drives are all manufactured by Exabyte), and there is a certain amount of standardization among vendors.

Of the DAT drives I have worked with, the ARDAT drive sold by Maynard, Palindrome, and other vendors is my favorite. It is compact, quiet, and uses direct drive motors instead of belts. DDS DAT drives in general seem to be very reliable.

Although workstation-based backup might be somewhat slower than server-based backup, it provides much more flexibility, especially in case of a file server crash. With workstation-based backup, files can easily be restored to a local drive until the file server is repaired, providing at least limited access to them. Server-based backup requires the file server to be running and the backup/restore software to be installed before files can be restored from tape.

Mountain's FileSafe and Maynard's MaynStream are both excellent choices for workstation-based backup. FileSafe has the edge in one important area—the ease and speed with which single files can be restored. However, Maynard's MaynStream has definite pluses in its cataloging capabilities and its ability to back up file servers by server name and volume name without the need to map drive letters to them.

For a server-based backup package, Cheyenne's ARCserve has a lot going for it. It is easy to use, and it makes scheduled, unattended backup of file servers and workstations a breeze. With the addition of a companion workstation-based product to provide quick restores in case of emergency, ARCserve would probably be at the top of this list with FileSafe and MaynStream.

Although its file management capabilities are impressive, Palindrome's TNA has one major flaw: It makes the job of restoring files dangerously complex. A full restore requires multiple tapes, and if the directions for reinstalling the TNA software are not followed explicitly, your archiving history, which is required for a full restore, could be lost. In addition, if you need to restore files to a different location after a server crash, such as to a local drive or another file server, the file history database must be restored first. This greatly increases the amount of time required to restore.

## Choose What Fits Your Needs

My ideal backup system would transparently back up multiple file types from all attached computer platforms while maintaining all attributes, security, and rights information. It would do it in minimal time, with maximum capacity and zero problems. Unfortunately, nobody supplies that system yet.

I have the opportunity to try a lot of different hardware and software. Like all products, those I mentioned have their good and bad points. If you understand their advantages and limitations, you can select a backup system that will fit your needs. Above all, a backup system should be reliable. In selecting a backup system, my primary criteria are the ability to:

- perform quick, easy, and complete backups
- find and restore single files quickly
- perform a complete restore quickly and easily
- perform a complete restore from one tape

You need to establish your own criteria for selecting a backup system. Bear in mind that the procedures you put in place are equally important. In the not-too-distant future, backup may be as simple as popping a tape into a drive. In the meantime, following the guidelines presented here will help you create a backup system that won't fail you. ■

# Planning for Microcomputer Security

## In this report:

Internal and External Microcomputer Threats.....	2
Micro-to-Mainframe Connection .....	3
Microcomputer Security Planning Considerations .....	8
Disaster Recovery Contingency Planning .....	11

## This report will help you to:

- Identify and address the basic threats to microcomputer security.
- Design a microcomputer security policy in keeping with your organization's strategic information plan.
- Develop a disaster recovery plan that includes microcomputers.
- Institute an information security policy tailored to your microcomputer environment.

## Introduction

In what has been termed the "microcomputer explosion," private companies and government agencies spend millions of dollars annually on microcomputers. Surprisingly, little time or money is spent to protect the information—a company's most important asset—stored on these micros. Many experts observe a growth in information security risks that they attribute to increased reliance on microcomputers. In fact, even in organizations with security policies governing mainframes and midrange systems, microcomputer security is

left up to the user, with no established policy governing security of the information to be processed.

Unfortunately, most U.S. organizations are in a "passive slumber" where microcomputer information protection is concerned. In the past several years, many of these organizations have been jolted out of this complacency by an unfortunate incident. Many organizations will institute safeguard measures based on a single security breach and return to "business as usual." What these organizations are overlooking is sound micro security planning that will put them in a proactive rather than a reactive mode where information protection is concerned.

This report was written exclusively for Datapro by Reed Phillips Jr., Director of Information Resources Management for the U.S. Department of Commerce. It has been revised by Datapro editors to reflect current technologies and security threats.

---

## Information Threats and the Microcomputer Revolution

A few years ago, most microcomputers operated in a standalone environment with little or no communication with other micros or larger computers. Most applications were standard off-the-shelf software packages. These micros with specifically tailored application packages were isolated to particular departments within the company. Theft protection extended only to the micros and software; little concern for data security existed, since the applications being processed were basic and included very little critical or sensitive data. Now as more critical and sensitive information is processed on microcomputers, primary emphasis must be shifted to information protection, with concern for hardware, software, and other areas becoming secondary.

Recent strides in microcomputer technology have increased the vulnerability of information associated with this equipment. The newest generation of microcomputers have standard memory capacities of one megabyte, a level that is sure to triple in the next few years as OS/2 gains wider acceptance. Due to the tremendous competition among hardware vendors, a fully configured system including a printer and modem can cost anywhere from \$1,800 to \$18,000. The use of local area network (LAN) technology is allowing the networking of two or more microcomputers to achieve file transfer. Barriers to the micro-to-mainframe connection are being removed. These include data formats, communications protocols, and operating systems. Data communication paths capable of carrying several hundred thousand bits per second will be commonplace, as will on-line storage systems that contain billions of data bytes. It can be speculated that a future microcomputer will have as standard equipment its own dedicated communications board and a channel for a coaxial cable.

---

## Internal and External Microcomputer Threats

It is estimated that by 1993, 60 million additional microcomputers will be sold, doubling the number of those now in use. More and more, information previously processed on paper is being handled on computers. Increased use of microcomputers presents a new range of information security and privacy problems that must be planned for now. This

includes dealing with the new breed of knowledgeable users who are capable of exploiting the vulnerabilities of microcomputers.

Microcomputer security controls work with a different set of variables than those for mainframes. Control over program changes, data security, system documentation, backup and recovery plans, and system testing is inherent in most mainframe environments, but microcomputer systems seldom have this protection. We are also faced with two very basic problems—most organizations don't know who their users are, and worse, they lack a reliable system to determine which users are interfacing with the mainframe.

We are also witnessing increasingly sophisticated packaged software and unique programs. Many systems are being built using relational database packages. To stay competitive, financial institutions and some other organizations are beginning to, or will soon have to, allow customers to access the company's computer files using micros. For example, an automotive parts manufacturer may allow customers to submit orders directly into its mainframe systems. Because these customers may be business competitors, it is important to safeguard against their gaining access to each other's files through the manufacturer's system.

One of the most serious security issues, one that has been compounded by the micro and LAN revolutions, is a lack of awareness of executives and users to the vulnerability of their critical and sensitive information. Industrial espionage, which previously involved paper files, has shifted with changing technology and now includes computer files. Companies in fierce competition for a large market share have discovered that valuable information can be obtained through monitoring competitors' microcomputers or by stealing diskettes containing it.

---

*Industrial espionage, which previously involved paper files, has shifted with changing technology and now includes computer files.*

---

In addition, hackers using microcomputers and communications links have, over the last few years, pirated millions of dollars worth of software, misused credit card information to obtain millions

of dollars worth of goods and services, fraudulently obtained airline tickets, and even acquired automobiles by ordering them and erasing the information from the financing firms' credit files.

Authorities believe that the actual number of crimes involving micros is significantly higher than reported, because uncovering them is very difficult. The percentage will most likely increase dramatically as micros become more prevalent in the business world.

---

### Micro-to-Mainframe Connection

Information managers were concerned early on that with the proliferation of microcomputers, users would want to gain access to mainframe computer files. Now that it is technologically feasible for micros to upload to and download from a mainframe, management is presented with a whole new area of unresolved problems of security, data integrity, and software reliability.

The IBM Personal System/2, along with upcoming editions of OS/2 and other software developments, will make micro-to-mainframe communications easier than ever. Eventually, micro-to-mainframe links will be the norm, rather than the exception.

Most information managers who haven't established micro-to-mainframe links admit security is the reason. They recognize that there must be carefully made policy revisions and planning to ensure that security is satisfactorily provided. Although managers and users may be anxious for the technology that would allow them to use files and databases on the mainframe, proper planning is necessary to avoid risks.

Many information managers who either ignored the security problems associated with the micro-to-mainframe connection (doubting that problems would occur in their organizations) or did not believe they would happen so rapidly, did not plan the necessary safeguards and thus found themselves in a reactive mode.

### Uploading Data

Two distinct aspects to consider in micro-to-mainframe connection are *data integrity* and *data security*. The first concerns data restriction to those who are fully authorized; the second concerns restricting the ability to modify data. These concerns have been so overwhelming in some organizational

environments that mainframe access from a micro has been granted to very few users.

*Data integrity:* Micros have made powerful computational and analytical tools available to users throughout many organizations. Increasingly important business decisions are being made based on information processed by micros and uploaded to mainframe systems. Information generated on large mainframe systems has been for years subjected to extensive critical reviews and error checking, both during system development and normal processing. This has resulted in confidence in the information produced from these systems. These rigorous controls are usually not applied to micros, and this lack of controls could become a significant problem. Mainframe files changed by a micro user cannot be checked for accuracy, and mistakes or data alteration can seriously undermine the database's integrity.

*Data Security:* From a data security standpoint, organizations are confronted with the difficulty of preventing user programs from intentionally or accidentally accessing or modifying portions of the operating system and circumventing security otherwise provided. Data uploading also provides a temptation for a user to access another user's files or the corporate databases and make modifications including entering unauthorized information.

---

### The Nature of Microcomputer Security Threats

In introducing microcomputers into the work environment, companies and governmental organizations are making the same security mistakes that were made with mainframe computers. A lack of substantial advance planning has resulted in significant vulnerabilities that exist in most environments, threatening the confidentiality, integrity, and availability of information associated with such systems. To ensure effective protection of critical and sensitive information, users and managers must be aware of the existing threats—and the protective measures that can be applied through proper planning.

---

*Microcomputers have unique security problems that must be understood if rational and effective security measures are to be implemented.*

---

Microcomputers have unique security problems that must be understood if rational and effective security measures are to be implemented. These problems, related to information security, include:

- Physical Accessibility
- Hardware
- Software
- Data Communications/Networking
- Uploading of Data
- Media
- Disaster Recovery/Contingency Planning

#### **Physical Accessibility**

Data processing centers protecting mainframe computers consist of layers of physical security, making unauthorized access virtually impossible. Microcomputer networks now handle just as much critical and sensitive information as mainframes but are more easily accessed within organizations.

Because it is seldom feasible to build a protective fortress around an individual or a cluster of micros, security against damage, hardware modifications, or unauthorized access is difficult to provide.

#### **Hardware**

Most microcomputers lack built-in hardware mechanisms necessary to isolate users from sensitive, security-related system functions. They also lack many significant security features available on mainframes: memory protection mechanisms that prevent unauthorized access to sensitive portions of the system; privileged instructions that limit access to functions such as reading and writing to disk; and multiple processor states that provide separate domains for users and system processes. Without these features, it is almost impossible to prevent user programs from accessing or modifying parts of the operating system and circumventing

any intended security measures. Although some technical knowledge is necessary to take advantage of these weaknesses, experienced microcomputer users are acquiring this knowledge.

A significant hardware problem in many organizations is microcomputer and/or component theft involving personnel within the company as well as outsiders. Some of these components are easy to carry away in a purse, briefcase, or coat pocket. Organizations that lack accurate or current inventories of their microcomputer components and peripherals are particularly vulnerable.

A situation similar to the automobile "chop shops" is becoming prevalent in the microcomputer industry. Black market sales of "hot" microcomputer parts are flourishing because of the popularity of home and business micros.

With an increase in industrial espionage and technology that makes it easier to eavesdrop, there is a growing concern over critical and sensitive information being leaked to a business competitor. Computer screens, cables, printers, and other peripheral devices broadcast radio-wave frequencies that can be intercepted with monitoring devices such as spectrum analyzers. Microcomputers can generate a clear signal that eavesdroppers can tune into from a nearby building or a van in the parking lot. Without proper shielding, these signals can travel several hundred yards.

#### **Software**

Information processed and stored on micros is normally linked with one person or a small well-defined group, it is often more sensitive and accessible than that found on larger, multiuser systems. Such sensitive information can be in the form of reports, spreadsheets, memoranda, or listings that are easily accessible using software tools familiar to all microcomputer users. This information is usually in a well-organized format as opposed to the unanalyzed or unprocessed raw data found on mainframes. The accessible format makes the task of searching for specific information simpler than on large systems with a multitude of users and data files.

With numerous generic applications software packages developed for micro use, security is threatened because detailed knowledge of specific application systems is not necessary to access potentially sensitive information. Many different databases can be accessed using one or two generic



applications. Also, in most smaller systems, all files are available to all users. The more advanced systems provide tree-structured file directories, which offer some degree of file segregation but no real protection.

*Viruses:* The growing number of known virus attacks—including that of a programmer who destroyed 160,000 sales commission records on his former employer's system—has left a number of organizations sadder but wiser. A virus could change data within a file, erase a disk, or direct the computer to perform useless system-slowng calculations. Viruses replicate and spread quickly, and are spread by downloading programs from outside sources, introducing untested software onto the system, or communicating with an infected computer through a network or by telephone.

The seriousness of the computer virus threat should not be underestimated. Viruses can have devastating effects if an information manager fails to assess the organization's risk of virus attack and institute policy changes to reduce those risks. Further, measures to detect and eradicate viruses should be in place before the need arises.

*Piracy:* Unauthorized copying of licensed software, or piracy, is a problem that has plagued software developers and the organizations they license since the development of the personal computer. As with microcomputers and their components, software is a prime target for employees owning compatible computers. Because of the inconvenience that antipiracy methods cause licensed users, most software developers have removed them from their products, thereby increasing the risk. Pressure from the major software companies has induced many organizations to adopt strict policies regarding piracy, but it continues to be a large-scale problem.

*Data Residue:* Another problem inherent to microcomputers is data residue left on disk or in memory. Data residue (or "magnetic remanence") is stored data on erased media that has been released for reuse. Such data can often be read by subsequent users. Thus, it is often dangerous to share with other users files on diskettes that contain erased, sensitive data. This problem exists for hard disks, as well, since the data remains potentially accessible to subsequent system users.

### **Data Communications/Networking**

An entirely new set of vulnerabilities exists when organizations begin adding micros to communications links and networks. This is apparent from the ease with which hackers, using a micro and modem, are able to gain access to, and use, communication links and networks without paying service charges. Many organizations, to transmit information among the various micros, connect them on LAN systems. The inherent danger with LANs is that all of the network's nodes can read all the transmitted messages.

When transmitting sensitive or critical information over unprotected communication lines, an organization risks signal interception and the possibility of an interloper masquerading as an authorized user actively receiving information. Dial-up access to mainframes from micros creates a special problem. On company or government property, all access can be guarded, both physically and administratively, but when users go outside these networks the protection disappears.

### **Media**

Magnetic media, the primary repository of users' information on micros, is also most vulnerable to damage. The two most prevalent types of magnetic media are hard disks (fixed disk devices) and flexible diskettes. The newer CD-ROM optical disks resist damage much better than magnetic media.

*Hard Disks:* A hard disk is usually found in a self-contained sealed unit that is relatively protected from the environment. Numerous advantages to hard disk use include faster program loading and quicker disk access resulting in reduced execution times. Also, the need to shuffle diskettes in and out is reduced or eliminated. One of the micro's few moving parts is the hard disk. Care must be exercised when physically moving a hard disk because of the danger to the read/write heads or other internal components.

*Diskettes:* Diskettes are highly pilferable because they can be used on home personal computers. Many times, diskettes containing data are taken when blank ones cannot be found. Often diskettes are not labeled, and the thief is not aware that valuable information is being taken. There have

been many instances where sensitive or critical information was contained on a stolen micro diskette. Diskettes are also vulnerable to ordinary contaminants in a working environment, and their surfaces can be damaged by rough handling.

**CD-ROM Optical Disks:** The CD-ROM optical disk boasts extremely high storage capacities. CD-ROM disks and their drives are adaptations of the compact disk technology common in home stereo systems. Unlike magnetic media, CD-ROMs are impervious to magnetic fields, head crashes, and contaminants. In addition, as read-only media, CD-ROM disks resist viruses as long as the original information stored on the disk is virus free. Because CD-ROMs are removable, they must be protected from theft, particularly since they are likely to contain an organization's entire database. The data contained on CD-ROM can be copied onto magnetic media by anyone with microcomputer access.

#### **Disaster Recovery/Contingency Planning**

Although many organizations have data center disaster recovery/contingency plans, few have considered micros in the planning process.

Disaster recovery/contingency planning consists of those activities undertaken in anticipation of potential emergencies or disastrous events that could cause serious adverse effects. This process includes failure assessment and emergency, recovery, and maintenance and testing procedures and backup resources, preparations, and operations. A vivid example of the lack of backup planning for microcomputers was the U.S. Postal Service Headquarters fire on October 16, 1984. The fire gutted the building's ninth floor, resulting in more than \$50 million in damage. Smoke and water damage was extensive throughout the building. The fire destroyed a computer room containing a Wang minicomputer and damaged 600 microcomputers, over 4,000 micro diskettes, and numerous data-filled hard disks. *Although the Postal Service had a disaster recovery/contingency plan which covered the primary data center located in the building's basement, it had no such plans for the minicomputer and microcomputers.*

Too little thought has gone into micro files reconstruction in the event of a disaster. Some users back up their micro files to protect against a hardware malfunction or media failure. But they

usually store the backup media in the same physical location as the working media, forgetting that all copies stored in the same physical location could be destroyed in a disaster.

Another problem for micros is *power surges* or *power spikes*. These can be caused by electrical motors operating on the circuit, power company load shifting, or lightning. Although lasting only a fraction of a second, electrical power variations can destroy microcircuits, damage relays, alter programs, or wipe out a microcomputer's memory.

---

### **Planning Successful Microcomputer Safeguards**

This section will examine approaches to effectively managing information security in a microcomputer environment. *A very important, often overlooked, factor is that the information security officer and the topic of information security must play a role in the entire planning process.* Otherwise, should an emergency or disaster occur, the organization will be unprepared and possibly suffer monetary loss and embarrassment because of strategic errors in the planning process.

#### **Policy and Procedures**

Although many companies have developed microcomputer security plans in the last two or three years, they are still in the minority. In many cases, security plans were brought about due to a breach of the company's security or the recognition that this was possible due to the publicity such breaches generated. The foundation of any strategy is a *policy statement* that must clearly define the responsibility for information security. This policy must be flexible yet enforceable.

All organizations should develop policies and procedures governing micro acquisition and use. These must comply with the overall corporate short- and long-range information system goals and security and control requirements. After implementation, they should be monitored regularly to assess adequacy and continued effectiveness.

In the past, it was very difficult to get user participation during any aspect of the information planning or policy development process. This attitude is changing, however, as users become involved to eliminate what is perceived as arbitrary restrictions imposed by an insensitive data processing department. This new source of input should

be welcomed by the information and security policy developers. *A hidden benefit of user involvement is that it heightens security awareness.*

Security policy for micros should be as simple as possible and relatively easy to implement. In addition, it should not establish a whole new organizational structure nor be time consuming. Once organizational microcomputer policy with security features is approved, it should be developed with the specific intent to protect against the micro security vulnerabilities discussed earlier.

A corporate policy addressing micro-to-mainframe compatibility should include within its documents or standards provisions for maintaining data security and data integrity. The policy document should also specify which organizational data must reside only on specific computers and is not to be downloaded or migrated to micros.

Some specific considerations might include encrypting and authenticating sensitive and critical information, and/or using Tempest-certified micros. The policy might also discuss the types of information that can be taken home and processed and specifically what types cannot be taken off the premises. Guidelines for storage of the micros and media (including backup) should be included. Procedures for backup should be developed. Policy on password security and implementation should be spelled out. Controls over use of access logs, inventories, etc., should be discussed. Procedures must be developed for reporting all suspected or confirmed micro security violations. To reduce risk of viral infection, a microcomputer security policy should define permissible software sources, bulletin board use, and the types of applications that can be run on company computers. The policy should also provide testing for unknown applications and limit diskette sharing. If a viral attack does occur, a policy should indicate what actions should be taken and by whom. Since the steps taken to eliminate a virus can be very disruptive to the organization, the security policy should designate an individual with the authority to make those decisions.

In today's microcomputer environment, new CD-ROMs can hold volumes of information. At the same time, the storage capacity of magnetic media has quadrupled, while disk sizes have decreased. Storage media manufacturers are promising higher information densities in the future, for even greater capacities. Since this information can

easily be duplicated, it is necessary to plan for restricting access, to formulate policies and procedures for removing disks from the premises, and to restrict internal handling.

At a minimum, an organization's policy and procedure documents should discuss the following aspects of microcomputer security: physical security; personnel security and training; information security; procedural security; communications security; software security; and disaster recovery procedures.

### Planning Strategy

Organizations must now position their micros to be an integral part of the overall strategy for information management. Users are becoming interested in moving applications off the mainframes and onto the micros where they are less expensive to run, but these moves should be carefully planned. Because of their age and inefficiency, many of these systems may need to be redesigned. Controls and security features should be built into the redesign of systems being migrated to the micros. This also provides an opportunity for the data processing manager to assert greater control over the way micros are used in the organization.

*In any security strategy, management should concentrate on protecting information and the means of processing or conveying the information.* Centralized control should be maintained over all micro hardware and software acquisitions. This is often done through the microcomputer stores set up by organizations. This centralized control makes policy and procedure enforcement easier.

A major concern among security officers is the acquisition of additional security products and the ability to integrate them with existing controls. A strategy must be carefully laid out before acquiring additional security products. It should include determination of information to be protected and evaluation of the various products available to provide the protection.

### Strategic Information Planning

An organization's information security officer should participate in the strategic planning process. This is normally when the information architecture is described, showing such activities as the integration of information systems, plans for implementing or expanding a telecommunications network,

developing a microcomputer-based distributed network, and the method for integrating office automation with data processing or building local area networks. This is the blueprint that shows how everything ultimately fits.

The information security officer must understand the information architecture in order to commence planning for entire system protection. This planning is heavily influenced by how the micros fit into the overall computer system. This person must also be informed about architectural changes. In conjunction with the architectural development, an information flow analysis must be performed to provide a clear understanding of how information moves throughout the system, to understand its interfaces and how it is being stored, and to eliminate system redundancy. The security officer can then analyze needs based on knowledge of the system's weaknesses or vulnerabilities.

The information strategic plan should be derived from the companies' overall strategy; the overall data processing, telecommunications, and office automation objectives; and the program objectives of the departments receiving information services support. More than ever, companies must have an understanding of the users' program objectives and their need for micros to support those objectives.

When determining the resources needed to support the strategic planning effort, line managers must express their priorities because normally the budget is not large enough for everything to be done at once.

Companies do their outyear budget estimates based on the strategic plan requirements and activities. *Therefore, it is important that information security officers fully understand the strategic planning process and what activities will be pursued, and the resources that will be needed to provide the necessary protection.* Including the various types of safeguard products in the outyear budget estimates ensures that they can be acquired when needed. If this is not done in the planning stages, security officers will be fighting for dollars throughout the actual budget year with less likelihood of success.

### **Operational Planning**

Microcomputers have an impact on operational planning. Effective operational planning translates

the strategic plans into meaningful action. The operating plan focuses on specific resource requirements and commitments during the coming year. It is important that the specific micro security strategy be reflected, since operating plans tend to quantify alternatives and prioritize various competing projects against realistic funding levels.

During operational planning, details such as micro interconnection, what information each system delivers, and how, when, and where it is delivered should be part of the planning process.

*An essential operational planning ingredient from the security standpoint is a systematic approach to identifying and implementing security requirements.* Throughout the organization, procedures must be developed for identifying and classifying information processed on micros and other assets (i.e., hardware, software, media communications, etc.) that require protection. The emphasis must be on information security problems. Once the analysis is accomplished, an assessment can be made regarding the specific vulnerabilities of the micros, diskettes, and other physical assets. Measures should be considered to provide cost-effective loss control, and a means of reviewing ongoing activities should be instituted to evaluate their continued effectiveness.

---

### **Microcomputer Security Planning Considerations**

Formal assignment of responsibility for the entire organization's information security is essential. This responsibility should apply to information in any format whether in paper records or on mainframes, minicomputers, or microcomputers. The basic responsibility for information security should be assigned to those who own and use the information. The concept of ownership means that every organization must wage a security awareness campaign to ensure that these owners and users are knowledgeable about information vulnerabilities and what safeguards can and must be taken to protect it.

Few users are aware of the full range of threats to, and the countermeasures available to protect, microcomputers and the information they process, transmit, and store. Providing adequate

training, assuring consistent procedures, and minimizing duplication of effort are significant issues that make the microcomputer environment unique.

Following are some specific planning considerations to safeguard information on microcomputers and their environment.

### **Restricting Physical Accessibility**

Because so many organizations have experienced the loss of micros and associated information, they have adopted stricter accountability procedures. Microcomputer distribution is now centrally controlled in many organizations. If possible, micros processing critical or sensitive information should be consolidated in areas that have physical access controls such as locks on doors, staff present during working hours, and guards checking during nonworking hours.

In organizations where micros are spread throughout the environment, surveillance measures and fraud forces have been increased. Some have restricted areas for working with sensitive data. These measures not only protect access to the micros but also assist in reducing theft of the micros and auxiliary equipment.

Micros located in secure offices that contain other expensive office equipment such as word processors, electric typewriters, and copying machines may require only recording of the equipment serial numbers for adequate insurance coverage. A few organizations authorize only certain personnel to use specific micros.

### **Protecting Hardware**

Organizations must now plan to ensure that there are lockable equipment enclosures, lockable power switches, fasteners, and heavy cables to secure micros. Several suppliers provide devices that fasten micros to desks much the same way typewriters are bolted to tables. Perhaps a more secure system is a series of properly sized boxes that are fastened to each other and then to a desk or a table. The market offers numerous other micro hardware security devices. Some of these devices are attached to movement-sensitive alarm pads. Some sound an alarm when disconnected from a wall socket. Others trigger an alarm at the mainframe CPU when the micro is disconnected without advance notice.

In some companies, policy compels users to disconnect and store the power cables in a safe place when the micro is not in use.

A system developed by the Department of the Navy for shipyards to track valuable tools could possibly apply to microcomputers. This concept is predicated on having *crystal oscillators* embedded in the micros with various frequencies cut to several hundred thousand frequencies. Antennas would be positioned to send out a signal that the oscillator would pick up and retransmit back to the antenna, alerting an observer that a micro is being moved.

One of the best systems available is a specially designed cart for mounting a micro. The cart has hidden power and connecting cables and provides locked diskette storage. Covers on the carts protect the micros from airborne contaminants.

Standardized *inventory forms* should be used throughout the organization. At a minimum, these forms should contain information on micro location, the person responsible for the equipment, and any modifications made to the micro. If possible, the inventory should be automated and maintained on a micro.

To protect the micros from possible moisture damage, consideration should be given to providing inexpensive plastic equipment covers that also provide dust protection.

For extremely critical or sensitive information, organizations are beginning to provide a protective shielding, referred to as Tempest, that prevents radio frequency emissions from traveling beyond the computer. Because printers, cables, and other peripheral devices also broadcast radio-wave frequencies, they must be Tempest protected as well. The major drawback is cost. A Tempest-protected micro costs approximately three times the cost of a micro without the shielding. Increasing supply and demand, however, should bring prices down.

Another approach rapidly gaining popularity for protecting against radio frequency emissions is the use of *elastomers*—electrically conductive plastics and paints that absorb radio emissions. These create a more aesthetically appealing form of protection than Tempest.

### **Protecting Software**

Many companies offer microcomputer security packages that prevent micro users from accessing

each other's files and protect disk files by encrypting them. If the operating system or programs are to recognize files containing sensitive information, internal labels must be present. Standard file management capabilities of most micros provide only the filename. Nonetheless, it is possible to store files in specific directories, thus providing file segregation by user or by data sensitivity.

When software security access control systems are used on both the micro and mainframe, it affords dual information protection. When a micro is accessing a mainframe, no modification of the access control package (e.g., CA-ACF2, RACF, CA-Top Secret) is needed, because the software package usually has security features. The threat to data security is reduced in a direct micro-to-mainframe connection because both system and data access can be controlled by the mainframe software.

Many microcomputer security packages provide data encryption; however, separate encryption packages can be purchased for \$250 or less. These products protect critical or sensitive data. The strength of crypto systems depends upon the quality, integrity, and secrecy of the keys used to encrypt and decrypt information. *It is important in any micro environment to keep system access logs that reveal who is accessing the systems and what information is being extracted.*

For shared micro situations, the user should be authenticated in some manner. The authentication process requires an explicit interaction between system and user, achieved through some type of "logon" procedure by which the user provides a nonsecret identifier such as name or account number and an authenticating password. User logon should be required whenever the system is powered up or a new user attempts access. Passwords should be randomly generated and, at a minimum, should be six-character codes. Hackers use as many as 12 to 18 character codes on their secret pirate boards to provide a greater degree of security. Passwords should be changed at least once a month—weekly is better. Multilevel password schemes will not significantly delay a legitimate system user, but they will help protect the system from intrusion.

It is important in any micro environment to keep *system access logs* that reveal who is accessing the systems and what information is being extracted. These logs should be monitored regularly

and procedures established so that unusual activity can be immediately investigated.

Logical controls are those enforced by computer software. The identification process that includes authentication of the person logging on to the computer is a logical one.

Each program designed for a micro should be thoroughly tested with sample data to ensure the software is performing its intended tasks. The test results should be compared with the results produced by the current system, and any problems should be resolved before using the program.

Most packages that allow file transfers include security provisions that bar users from sensitive files or restrict a particular user's access only to the data needed to perform specific functions.

Perhaps the most effective means of protecting information when connecting micros to mainframes is to have multiple levels of software security. This provides greater protection than a single countermeasure that an unauthorized person might be able to circumvent.

Because most micros have residual memories, it is important that users clear the memory after processing sensitive information. Memory clearing requires removing the storage media, powering off the micro to clear any volatile memory (that which is lost when power to the micro is off), powering the unit back on, and either proceeding with the next job or overwriting the portion of the permanent storage area with random patterns.

### **Communications/Networking Protection**

If a communications link from a micro is handling sensitive data, it is best to plan to encrypt the data. Although expensive, it can be justified by the cost of unauthorized disclosure. There are now both software packages and hardware devices available for encrypting users' data files. Some companies are using devices incorporated into the micros (i.e., expansion boards) that have complete communications and cryptographic protocol functions built in. Encryption device costs also are coming down.

The use of local area networks (LANs) for networking micros offers more security than modems. Most LANs have some data security measures built into the networking software, but these measures are not nearly as sophisticated or well developed as those available in the mainframe environment. As a result, highly sensitive data should generally not be stored on LANs. LAN designers

are beginning to recognize the importance of developing measures to make LAN-resident data more secure.

Some companies allow data uploading from a micro to mainframe only to a buffer area; the data is later incorporated into the production database by data processing center personnel.

### Protecting Storage Media

In day-to-day operations it is critical that diskettes be labeled before use and that the diskette be securely stored when not in use. Two essential label ingredients are the creator's name (or the person who owns the diskette) and the format—this is especially critical in multivendor organizations. One vendor manufactures color-coded diskettes, which assists in physically identifying media with sensitive data stored on them.

One method of preventing diskette theft is to hide signaling devices in the plastic disk jackets similar to those used by libraries and retail stores. If someone tries to depart the building with a diskette, it triggers an alarm.

Diskettes and any other media should be stored in closed and, where possible, locked containers to protect them from contaminants, unauthorized access, disclosure, damage, modification, or destruction. The containers also provide some protection from soot damage and water damage in the event of fire. Diskette backup must be a routine procedure, with storage off-site if possible.

### Future Technology Planning

Published materials provide an idea of the general trends in micro security technology for up to five years in advance. Even though micro protection is available from some vendors, it's not widely offered. The reason is that microcomputer design technology is ahead of computer security technology. Micro designers tend to consider security a separate concern. Security also increases the price and in many cases slows down processing speeds and impedes access. While computer security technology has decreased in price and become more widely available over the last few years, it is still expensive and usually the last thing a micro buyer thinks about.

Within the past several years, protection features in hardware and software capable of supporting multilevel security on larger mainframes have

increased. For micro-based systems, however, difficulties have been encountered in designing multi-level security. Virtually all off-the-shelf MS-DOS micros lack one or more of the capabilities such as memory mapping, multiple processor execution states, and instruction trapping features necessary to implement multilevel security. OS/2 should change that over the next few years as it gains acceptance and development shifts away from MS-DOS.

Each company must ask, "Are our internal controls, policies, procedures, protection plans, and safeguards for micros adequate to protect us in the future?" It is very important to assess potential micro security product candidates to ensure that the product chosen meets the necessary requirements.

---

## Disaster Recovery Contingency Planning

### Why It Is Necessary to Plan for Disaster

A disaster recovery/contingency plan helps organizations respond to the variety of events that might affect processing availability. The plan must include responses to natural catastrophes (such as fire, floods, or storms), hardware failures, power irregularities, strikes, and bomb threats.

---

*The primary objective of disaster recovery/contingency planning should be the continuity of business activities, not of computer processing.*

---

The primary objective of disaster recovery/contingency planning should be the continuity of business activities, not of computer processing. Systems users should be encouraged to protect themselves by developing and maintaining their own fallback procedures.

The task of backup and contingency planning in a microcomputer environment is essentially the same as for other data processing activities. For organizations with microcomputers, midrange systems, and mainframes, planning should be an integrated process. There will be special considerations

for micros because the equipments are widely dispersed and many people are involved.

While recovery planning for micros should be easier than for mainframes, without proper planning it can be more difficult. It is becoming increasingly important to plan for regular and systematic backup of micro files, because such backup can no longer be accomplished centrally and systematically as done on minis and large mainframes.

A significant lesson learned from the U.S. Postal Service fire where 600 micros were involved is the importance of good documentation including complete updated micro inventories and identifying sources that could provide assistance if an emergency or disaster strikes. Equally important is a backup plan for storing critical information off-site.

#### **Microcomputer Disaster Recovery Procedures**

Microcomputer disaster backup and contingency planning requires substantial emphasis among users because it concerns problems and speculation regarding future events. In situations where locally stored backup copies would be lost with the originals, consideration should be given to storing periodic archival copies at some location unlikely to be jointly affected by common emergencies such as fire and flooding. Where micros are connected to a mainframe, some organizations have the micros upload to the mainframe, where the information is included in its backup files. In other situations where micros are connected to a data communications network, it may be possible to establish procedures to make backup copies on a separate device, such as a remote host or file server.

The method and frequency of backup media must be determined by each user, based on the storage media and the volatility of the data involved. Some commercial off-site storage services are becoming available to handle micro backup media at competitive rates and they provide pick-up and delivery services. Also popular with larger companies is sending backup daily to a branch site for storage; this can often be done using regularly scheduled internal mail deliveries. Smaller companies may desire to keep backup media in a fireproof container or have an arrangement with another company in the same city for backup storage. In organizations where the data processing

magnetic tape library is well secured and properly protected, the library is used as a backup site for micro media.

Some micro users use diskettes exclusively and, if this is the case, a copy should be made at day's or week's end for backup. Many companies maintain three copies of all microcomputer information, referred to as grandfather, father, and son. The son is the working copy; the father is kept close at hand (it is the backup needed most frequently); and the grandfather is placed off-site in a location that the company can reasonably access. The grandfather copy should be stored in a fire-proof corporate vault located on a different floor, in another building, or in a bank safe-deposit box. If the father and son are destroyed, all but the most recent transactions should be available on the grandfather.

For hard disk systems, it is usually impractical and normally unnecessary to build full disk backup copies daily. Usually only designated sensitive or critical data requires capture for backup. Storing the entire contents of a typical 30-megabyte fixed disk requires approximately twenty-one 1.44M-byte diskettes and two to four hours, assuming that the diskettes are already formatted. Although still relatively expensive, some organizations use a tape cartridge backup system capable of storing from 10 to 60 megabytes on a single tape in a matter of minutes. Although they are slower, a removable hard disk or high-capacity diskette system can also be used for backup.

Often, users run commercially available software with built-in privacy protection mechanisms that link the software to a given micro or system disk. This may cause problems when trying to conduct backup operations on different equipments or with alternate software versions.

In areas where the power supplier is unreliable, producing frequent outages, voltage spikes, or large fluctuations in voltage or frequency, companies are expanding their uninterruptible power supplies (UPS) to support their micros and networks. UPS systems can be purchased for as little as \$400, up to about \$800. These devices automatically provide power as soon as the output from the normal source deviates from specified levels.

Surge suppressors protect personal computers from sudden voltage increases that might cause errors. Organizations may wish to install a spike



protector between the power source and the microcomputer. A spike protector keeps momentary power surges from damaging the micro. Some spike protectors are actually power regulators that smooth out current fluctuations in both directions, while others trip a fuse or circuit breaker and are often built into a multioutlet bus bar. The important difference is their reaction time. A voltage spike can pass through a standard circuit breaker fast enough to burn out the micro circuit without tripping the breaker. A surge protector dissipates surges until the circuit breaker has time to function, and some surge protectors also drop off the line if the power dips too low.

---

### Conclusion

Security planning and countermeasures introduced must be accompanied by a strong awareness training program. Otherwise, an overall security program or even a micro security program will accomplish little. It is extremely important to create an awareness of security and control concerns and inform users of procedures needed to maintain adequate safeguards.

The cause of most data security problems is lack of management concern. Security will always be a managerial rather than a technical problem. Unfortunately, this fact is not recognized in many organizations. To guard against a costly and embarrassing security breach, management must clearly establish and strictly enforce security controls. Many organizations are in the process of making information management responsibility an integral part of each manager's job. Every organization should do the same for information security.

Every organization member bears responsibility for security, and management must lead the way. Users must accept the task of protecting the sensitive and critical organizational data they handle through their microcomputers.

The key to ensuring that security is built into microcomputer networks and micro-to-mainframe systems is advance planning. An organization will not let a security obstacle prevent it from gaining the productivity benefits of using micros. Management must lead the way in achieving both productivity and security by promoting security awareness and in giving full backing to the organization's security policy, plans, and procedures. ■



# Developing Microcomputer Security Awareness

## In this report:

Microcomputer Security Problems.....	2
Suggestions for Better Backups.....	4
Suggestions for Sensitive Data Protection .....	4
Suggestions for Maintaining Data Integrity.....	5

## This report will help you to:

- Realize the information security risks in the transition from host-based to microcomputer-based computing.
- Understand the role management must play in implementing and maintaining an information security plan.
- Take steps to preserve the accuracy of microcomputer-generated information.

## Introduction

During earlier times when an organization's data processing was done solely on mainframe computers housed in computing centers, the computer security function could focus its attention on well-established physical, technical, and personnel security measures. Physical security measures kept unauthorized persons from direct contact with computing equipment and utilities. Only computing operations personnel were normally allowed unrestricted access to computing centers. Access control software ensured that only authorized users

could obtain on-line computing services with limited access to certain files and programs. The traditional internal controls of separation of responsibilities were relatively easy to enforce. The operators ran the computers, the programmers developed the software to run the systems in accordance with the owner's requirements, and users entered data and/or commands to utilize the system. The separation of responsibilities was a key element in making it difficult for one person to manipulate the system for personal benefit. Normally, collusion of two or more persons was necessary to compromise a system.

The move toward distributed processing, which originally involved the use of minicomputers, complicated the security problem to some extent. The minicomputers performed local processing of data downloaded from the mainframes, and the results were then uploaded to

---

This report was written exclusively for Datapro by Mr. Hal Tipton, manager for information security with Rockwell International Corporation, Seal Beach, CA.

the mainframes. The complications resulted if physical security were relaxed because of the expense of providing the necessary protection at a number of remote locations. As a result, some programmers and even users were allowed direct contact with the minicomputers to run their own jobs, deteriorating the separation of responsibility concept and increasing potential security exposures. Also, most minicomputers did not support the kind of sophisticated software necessary to provide adequate technical security.

While security officers were dealing with the initial challenges of distributed processing, microcomputers began gaining wide acceptance. Concurrently, the number of mainframes in use decreased due to increasing mainframe capacity, and the work load moved to hundreds and thousands of microcomputers at remote locations from the data processing center. Today, the power of a micro is equivalent to that of the mainframes of 20 years ago.

As a result, the responsibility for exercising computing security now extends to the individual microcomputer user. The situation creates a number of potential security problems and exposures.

---

### **Fallout from the Microcomputer Explosion**

Microcomputers have decentralized computer-based information. Thus, sometimes the user, the programmer, the systems analyst, and the operator is the same person. Immediately, the carefully constructed and maintained series of checks and balances, which separates responsibilities, disappears. The atmosphere may become so relaxed that the user ignores all controls.

A major problem is the general lack of user awareness about the vulnerability of computer-based information. Users now have much more information at their fingertips. For instance, microcomputer networks often allow users to access previously protected corporate data such as official financial or legal information.

---

*. . . the widespread use of microcomputers brings practices that can undermine security.*

---

On the other hand, microcomputers provide users with several productivity enhancements including:

- A better response time during local processing.
- The capability for the user to work at home.
- The capability to easily manipulate and analyze information.
- A user friendly interface making computer use easier.
- Frequent use of the microcomputer resulting in increased familiarity and therefore, better and more effective use of computing resources.
- A graphics capability not readily available on mainframes.

Along with gains in productivity, however, the widespread use of microcomputers brings practices that can undermine security. One of the disadvantages stems from the inadequate documentation provided with computing equipment. This contributes to potential user errors that could lose, damage, or destroy data or programs. Errors also result from the failure to provide sufficient training for the rapidly expanding user population. Inadequate training and limited equipment cause many users to fail to properly back up their files, and they are unable to recover quickly in case of damage or destruction of the active files. Another problem is that in their haste to bring equipment to market, vendors tend to provide inadequately tested software. Under certain circumstances, software faults cause a system failure that destroys data. Also, because microcomputers are designed for single users, few security features are provided. Finally, because microcomputers are scattered throughout an organization in or near users' desks, it is seldom feasible to protect them using standard physical security measures.

---

### **Microcomputer Security Problems**

Microcomputer security concerns can be grouped into four general categories: microcomputer access control problems at the host end, access control problems at the microcomputer end, problems caused by decentralization, and unauthorized copying of software.

Access control problems related to the host computer involve downloading, uploading, and

dial-up access. Files that are well protected at the host often undergo significant exposure when downloaded to microcomputers, which usually lack file protection. Uploading data to the host from microcomputers risks host database corruption, either accidentally or intentionally, because of the weaker access controls on the microcomputer. Allowing dial-up access to the host from the microcomputer leads to several exposures, including the possibility that hackers could invade the system. Dial-up capability also allows users access to the host from any microcomputer equipped with a modem, thereby permitting access from unsupervised locations.

Access control problems related to microcomputers are magnified by their relative unsecured location in the workplace, the limited capacity of available access control software, unprotected diskettes that are easily stolen, and their easy portability that makes them difficult to track. Unsecured workplace locations make it easier for intruders to use your microcomputers and take advantage of the limited access control software to compromise the system. The exposure to theft of diskettes also threatens the confidentiality and availability of the information they contain. Microcomputer portability can compromise the availability purpose of the information security program.

---

*... unauthorized copies of software may be contaminated with a virus that can destroy or disrupt the entire system.*

---

A variety of problems relate to the decentralization of microcomputing. First, the personnel using microcomputers are not usually data processing professionals trained to take regularly scheduled backups of programs and files. Another significant problem related to microcomputers is the risk to data integrity caused by incorrect data entry, static discharge, dirty media, hardware failure, etc. It is important that frequent backups be made to ensure a quick recovery capability. Also, off-site backup

---

## Microcomputer Security Checklist—Risk Analysis

When identifying vulnerabilities, the following system characteristics should be considered:

- Are multiple users on the system?
  - Do users input the data being processed?
  - Does the data come from another computer system?
  - Are assets controlled by the system?
  - Are system reports used in decision making?
  - Does the system update the master files of another system?
  - Does the system contain company-sensitive information?
- 

storage may be advisable under many circumstances. With larger amounts of mainframe processing now being performed on micros, it is often necessary to have similar disaster recovery plans established.

The unauthorized copying of software is a serious and sensitive issue. Studies indicate that about half of the commercial products in use are unauthorized copies or bootlegged. This is dangerous from two viewpoints. One is that the organization may be the target of a lawsuit if unauthorized copies are found. The other is that unauthorized copies of software may be contaminated with a virus that can destroy or disrupt the entire system. Frequently, users are encouraged to make copies of each other's software or download software from bulletin boards because of the aggravation often involved in obtaining organization approval to purchase their own copy. This aggravation is seen in extensive cost justification requirements and/or the time involved in completing the procurement cycle.

These microcomputer security problems can be summarized into the following three general categories:

## Microcomputer Security Checklist—Administrative

The following are some operational factors to consider when defining a microcomputer security policy:

- Is there a company policy to cover the acquisition, use, and security of PCs?

- Is there a clear statement in the company policy advising users of their responsibilities/accountabilities for PCs and PC data in their work area?

- Is a PC coordinator with responsibility for coordinating PC use and PC security appointed for each work area?

- Is a separation of responsibilities for data entry, computer operating, and programming maintained?

- Is mainframe database access by PCs authorized by specific individuals?

- Is adequate PC training provided to users?

- Are users instructed in the proper care of PC media?

- Are users instructed how to sanitize data stored on PC media?

- Are users trained in security awareness and security procedures?

- Does policy forbid users to keep diskettes in their desks?

- Are PC users cautioned against copying proprietary programs?

- 
- Loss of critical data due to lack of proper backups.
  - Disclosure of sensitive data left exposed to theft.
  - Loss of data integrity due to a lack of a quality control mechanism.

### Suggestions for Better Backups

The following policies and procedures are recommended to address the need for proper data and program backups.

**Require Backup Equipment.** Organization policy should require a streaming tape or removable cartridge backup subsystem for all new microcomputers.

**Mandate Data and Program Backup.** Organizational policy must require program and data backup at regular intervals. Users usually will follow a policy that they understand and consider reasonable. This policy should specify data and programs that require no backup, those that require a secure on-site backup, and those that require both an on-site and off-site backup.

**User Training.** Training on backup management procedures should be provided to all users.

**Interaction with Users.** Organizations should make their security officers available to consult with users on backup policies and procedures so that misunderstandings can be identified and corrected before bad habits are formed.

**Supervisory Support.** Users should know that their supervisors fully support the backup policy, and supervisors should be required to audit microcomputer backup activity regularly. This procedure can be more effective if the supervisors' performance of this responsibility is reflected in performance evaluation comments.

---

### Suggestions for Sensitive Data Protection

The following policies or procedures can be used to minimize the problem of sensitive data being inadvertently exposed to theft. They include user guidance as well as physical and technical security measures.

**Define Sensitive Data.** Organizational policy should define what is meant by the term sensitive data. Included should be a discussion of the types of sensitive data. Additionally, a custodian for each sensitive dataset should be identified. This policy will enable the users to better understand the scope of their responsibilities and more easily identify the level of protection to be provided.

**Identify Where Sensitive Data Can Be Stored.** The information security plan should prohibit sensitive

data storage on fixed disks. This is necessary because access to data on hard disks is very difficult to control—they cannot be removed and locked up when the user is not using the microcomputer. Instead, sensitive data should be stored on specially colored diskettes or designated/marked removable hard disks. This policy makes it easier for supervision and security personnel to identify media containing sensitive information and take corrective action if it is left exposed to theft or removed from the facilities.

**Provide Secure Storage Areas.** A locked area or cabinet should be provided for all media containing sensitive information. The media should be under the control of and accessed only by a designated data custodian. The custodian should maintain a log of media entered into or removed from the storage facility.

**Secure Microcomputers Working with Sensitive Data.** Microcomputers that are used to work with sensitive data should be placed in private offices that are locked when not occupied by authorized users of the data.

---

### Suggestions for Maintaining Data Integrity

The following procedures should be employed to minimize the problems maintaining data integrity in a microcomputer system.

**Validate All Customized Programs.** All customized microcomputer programs used in supporting the organization's decisions must be validated and certified for accuracy. This is important to ensure that the output used in making decisions is not inadvertently skewed, thereby adversely affecting the accuracy of the decisions.

**Time-Stamp Database Modifications.** Organization databases should include date and time fields that are accessible by the users to indicate when the database was last modified. This is an important feature to advise users of the currency of the data, particularly if data from another source is being combined. It is extremely difficult to keep data from different sources in synchronization with

---

## Microcomputer Security Checklist—Physical Security

Physical security should provide the following coverage:

- Is an ID tag firmly attached to each piece of hardware?
- Is the equipment adequately secured against theft?
- Are fire extinguishers provided?
- Do guards check for authorization to remove equipment or storage media from company premises?
- Are PCs protected with power surge protectors, line filters, and uninterruptible power if processing critical applications?
- Is backup storage media kept off-site?
- Are diskettes stored under lock and key?
- Are diskettes left in machines unattended?
- Are diskettes labeled: contents, classification, department, and company ID?
- Are internal components protected from removal by lockable covers or similar devices?
- Are reports, file layouts, file dumps, etc. locked up when not being used by authorized persons?

---

each other without this kind of update information. Lack of synchronization results in decision credibility problems.

---

*... no corporate decisions should be made on the basis of microcomputer-produced data unless data integrity conditions have been met.*

---

**Mark Microcomputer Reports with Production Dates.** All microcomputer output should be marked with the time and date of its production. If a report is from an external database, the time and

## Microcomputer Security Checklist—Technical Security

Additional measures to consider when securing the microcomputer include:

- Is access control installed to protect data on hard disk from unauthorized access?
- Is it possible to alter financial data without producing an audit trail?
- Can outdated or incorrect files be inadvertently processed?
- Is diskette access tightly regulated?
- Can users create data for use outside their departments?
- Is an access control system provided?
- Is plain-text version of encrypted files deleted?
- Are copies of all user-written production software, purchased software, and all data files stored off-site?
- Have all applications been properly documented?
- Are passwords adequately composed, changed regularly, and protected from access or exposure?
- Is dial-up access control provided?
- Is PC utilization recorded for review by management?

date mark for that database should also be available. This procedure relates to the same synchronization concerns as the previous item.

**Validate Data and Formulas in Reports.** An independent validation of data input and embedded formulas used to create analytical data reports should be conducted to verify the work of the originator. There are many horror stories about major corporate decisions that were based on faulty computerized information that could have been avoided by using this procedure. Similarly, all analytical models should be checked for conceptual

and clerical accuracy. The auditors should be required to initial the reports that they have checked and include a note of the time and date that the check was made.

The final word on the data integrity assurance issue is that no corporate decisions should be made on the basis of microcomputer-produced data unless all of the above conditions have been met.

### Microcomputer Security Issues

The following security issues specific to the use of microcomputers need to be resolved within the organization's security program.

**Limit Access to the Microcomputer.** The need to limit access to the microcomputers as well as the data and programs they contain can be addressed by a combination of policies and implementation of a spectrum of control measures.

A rather sensitive issue is that of allowing employees or contractors/consultants to bring their personal computers onto company property instead of using company-provided equipment. If possible, a firm organizational policy should be established prohibiting the use of non-organization-owned microcomputers in organization facilities. There are several reasons for this policy, the first being a concern about viruses which can be brought into an uninfected environment via outside hardware and/or software. Another is the difficulty in preventing personally owned equipment or software from being upgraded with organization-owned components. Also, there is a question of whether the organization is responsible for the loss or damage of personally owned microcomputers on company property. Finally, the temptation exists for users to perform personal business on company time. Each case of personally owned micros on company property should be thoroughly justified and approved by senior management.

Control methods that should be implemented include both physical and technical elements. The physical methods consist of using security personnel to patrol areas where microcomputers are located, using badges to make it easier to recognize intruders, using locks on equipment and doors, and using a secure place to store diskettes.

Technical methods include using access control software to prevent unauthorized personnel



from starting up the micro, limiting user access to specific directories or files, and employing encryption to prevent unauthorized personnel from reading confidential data.

**Identify Dial-Up Entry Points.** Dial-up entry points are tempting targets for hackers and are serious exposures. The usual dial-up entry points include digital switches, dial-up hosts, and dial-up micros. While switches and hosts can usually be well controlled, it's a different situation with microcomputers because the ease of attaching a modem creates an unauthorized and dangerous entry point. Therefore, it is necessary that supervisors routinely check microcomputers to ensure that unauthorized dial-up entry points are not created.

**Limit Scripted Logon Sequences.** Users frequently code their mainframe log on sequences into their microcomputers so they can log on to the mainframe by pressing a key, thereby allowing intruders to logon using the same key. Users should be discouraged from including their passwords in programmed logon routines.

**Protect Proprietary Software.** The need to protect proprietary software is a particularly difficult problem that is currently best attacked by user education and supervisor action. To assist in monitoring this issue are special programs that will indicate the presence of unauthorized software in a system. These programs are also useful as a virus deterrent by showing the existence of virus code.

---

### Potential Security Exposures

Several potential exposures can undermine an otherwise well-designed and -implemented microcomputer security program. Some of these that should be monitored are:

- Sophisticated applications software packages that may inadvertently threaten data security. For instance, particular skill is no longer required of a user to access sensitive information. Users are now capable of entering many different organization databases by using generic applications. This exposure is expected to worsen as new integrated software now being developed will compound the issue. These applications

---

## Microcomputer Security Checklist—Mainframe Security

To effectively control access to the mainframe, the security policy should address these issues:

- Is access control software and/or callback hardware installed to protect mainframe dial-up ports from access by unauthorized PCs?
- Is the mainframe logon sequence programmed on the PC?
- Are the number of unsuccessful password attempts limited?
- Are unsuccessful password attempts logged?
- Are uploading and downloading controlled and logged?

---

make it difficult to maintain access controls on the need-to-know basis that is required to limit exposure.

- The hiring of outsiders to provide training and support in the use of new software and hardware can be a dangerous exposure because security breaches rise in direct proportion to the number of people who have access to or knowledge of an automated operation.
- File contents remain following the execution of the DOS delete command since only the file pointers are deleted. This is a potential exposure to organizations because the file contents can be recovered by intruders using readily available disk utilities. Also, RAM is seldom completely cleared of its contents unless the microcomputer is turned off—a tactic that is usually not convenient for users to employ.
- Local area network (LAN) messages that can be read by other nodes despite specific network addresses. Through user-implementable modifications, all nodes can read data being transmitted on the LAN, and information should not be

assumed to be limited to the sender and designated receiver. Also, nodes can disguise themselves as other nodes to defeat efforts to selectively address messages; therefore, information sent over a LAN should be carefully screened to limit possible exposure.

---

## Conclusion

In summary, certain countermeasures will minimize the potential risk posed by the widespread use of microcomputers. First, an effective and continuing user security awareness program must be designed and implemented. Individual users must fully understand the need for good security practices and must be responsible for implementing and maintaining good security practices.

Second, users must be provided with appropriate security tools so that they can effectively perform their security responsibilities. These tools include, but are not restricted to, access control

software to limit access to sensitive data, hardware for fastening equipment to desks to make pilferage more difficult, lockable cases for diskette storage, and understandable policy/procedures/guidelines that are reasonable in their expectations.

Finally, management is ultimately responsible for fostering microcomputer security awareness throughout the organization—while at the same time maintaining a user-friendly data processing environment. Organizational security policy must make it clear throughout the organization that local management is responsible for microcomputer data security. Experience repeatedly proves that unless management takes the lead—displaying and demonstrating a consistent interest in security—users will not take the challenges of security seriously. ■

# A Sample Microcomputer Security Policy

## In this report:

System and Data Access Control .....	2
Software and Data Integrity .....	2
Software and File Backup .....	4
Equipment Protection .....	4
Personal Computer Networks and Terminal Emulation .....	5

## This report will help you to:

- Understand the need for a microcomputer security policy for your organization.
- Develop, publish, and implement a microcomputer security policy.

**NOTE:** *The microcomputer security policy presented in this report is a follow-on to Report 5820, "Developing Microcomputer Security Awareness." As with any prewritten sample policy, a word of caution is in order. The best strategy is not to circulate the sample policy to security committee members or upper management until the problems and needs of the individual company have been thoroughly analyzed and the components needed in a security policy have been identified. At that point, a sample policy could be circulated for review. If the sample policy is brought into the proceedings earlier, human nature and time factors may spur premature acceptance without sufficiently analyzing the sample's applicability to local conditions.*

## Purpose

This policy applies to the use of all microcomputers, except for those incorporated into deliverable hardware sold to customers. The purpose of the policy is to provide internal control over microcomputers in all locations of the company, including subsidiaries.

Microcomputers and various user-controlled programming and report writing tools have become an integral part of computing activities and have provided substantial productivity improvements. Technological advancements have facilitated the use of these resources and have resulted in an increase in user-developed and -maintained, computer-based applications.

While the use of these resources is encouraged, it is essential that their application be subject to each user location's system of internal control. Adherence to the following general

This report was written exclusively for Datapro by Mr. Hal Tipton, manager for information security with Rockwell International Corporation, Seal Beach, CA

guidelines for all computing resources handling corporate data, financial and nonfinancial, is absolutely mandatory.

- Control procedures must be in effect to ensure appropriate protection of computing resources and corporate data in accordance with the requirements specified in Corporate Directives and Policies.
- Integrity and accuracy of information must be ensured. Adequate internal control and data integrity procedures must be incorporated with the design, development, and operation of all computer-based applications for corporate data.
- Local management is responsible for developing standards and guidelines for microcomputing and user-controlled systems. The appropriate equipment facilitates its effective and accurate utilization and ensures the security of equipment assets, programs, and data files.

The following, specific guidelines are categorized for easy reference and, while not necessarily mandated in every case, they should be given serious consideration by local management during the development of local standards and guidelines.

---

## Guidelines

### System and Data Access Control

Security problems exist in access control because most micros are designed for single users and provide for unrestricted access. Also, it is easy to overcome software-based security devices, system components and diskettes are easy to access and interchange, and users can access all system facilities and cannot be prevented from using utilities to read/modify files and file directories. These problems are particularly serious when sensitive data or proprietary software is downloaded from a host where protective measures were more effective or when sensitive data/software is created or maintained on the micro.

The following security measures should be considered to prevent unauthorized access to sensitive data sets.

- Install access control software to prevent unauthorized access to data on nonremovable media (hard disk).

- Establish standards specifying that no sensitive data will be stored on nonremovable media. Sensitive data should be stored on specially colored/identified removable media that is kept physically secure when not in use by an authorized user.
- Establish physical access control so that unauthorized persons cannot operate the computer system containing sensitive data or make modifications to the equipment or software. These physical controls can include placing the computer in a locked room, installing a lockable keyboard, etc.
- Install an encryption capability so that sensitive data is stored encrypted and the decryption key is protected from unauthorized use. When encryption is used, it is necessary to ensure that plain text is erased or written over during the encryption process so that special utilities can't be used to bypass the encryption by accessing the plain text.
- Provide a logon procedure that requires a non-secret identifier (logon ID) and a secret password for user authentication to access the computer system. User reauthentication should be required whenever the user may have changed. This can be accomplished by causing a system reset (reboot) either manually (user must perform reset when leaving the machine), automatically (the application performs a reset when processing is completed), or automatically time out (changing the operating system to cause a system reset after a certain period of system inactivity).
- Install a program to purge all file data as part of the deletion process on systems that are shared between users. This will avoid the problem of data residue being accessible by subsequent users if the delete command only sets a file-deleted indicator in the file directory instead of physically erasing or writing over the data.
- Institute positive dial-up access control measures.

### Software and Data Integrity

Security problems can exist in software and data integrity when either downloaded data is manipulated on the PC and uploaded back to the host or decisions are based on information derived from

software and data residing on the PC. These integrity problems result from accidental or intentional acts by users that cause the software or data to be unreliable.

Accidental acts (errors and omissions) by users occur more frequently in a PC environment because:

- The amount of error checking available on the PC is much less than that normally employed on mainframes.
- The amount of data processing and training provided to PC users is often inadequate compared to that of data processing professionals (computing center employees). Also, there is a lack of user awareness about the vulnerability of computer-based information.
- It is not unusual for privately available software, such as that obtained from bulletin boards or PC Club programmers, to have unsuspected errors or contain hidden routines that destroy or damage entire databases.
- The use of application software packages may provide broad access to previously protected, sensitive, legal, and financial files without requiring highly skilled users.
- Most micros are designed for single users who can access all systems facilities and can't be prevented from using utilities to read/modify files and file directories.

Intentional acts (data manipulation, fraud, etc.) by users can occur more frequently in a PC environment because, in addition to some of the above reasons:

- Separation of responsibilities is difficult to attain because in microcomputing the same person is often the user, programmer, computer operator, and systems analyst.
- Dial-up access can be accomplished from unsupervised locations.
- There is a lack of audit trail capability on microcomputers.

The following security measures should be implemented to protect software and data integrity before corporate decisions are made based on PC-based data.

## Definitions

**internal control**—A plan of organization designed to safeguard the corporation's assets, verify the accuracy and reliability of data, promote operational efficiency, and encourage adherence to prescribed managerial policies.

**microcomputers**—Small, programmable, microprocessor-based computers (also referred to as "personal computers", "desktop computers," or "intelligent workstations") possessing a self-contained range of word processing, data processing, and transaction processing functions. Microcomputers are capable of operating in a standalone mode, as well as interactively with other computers.

**company-sensitive information**—refers to both proprietary information and company official information.

**proprietary information**—Information applicable to research, development, and production technology which is generated by, or on behalf of, the company and which is useful to the company and would adversely affect the company's interest if not properly protected. It may or may not be in documentary form and includes computer software programs and databases.

**company official information**—Information applicable to the business personnel and financial and legal affairs of the company which is generated by, or on behalf of, the company and which is, by reason of its sensitivity, to have limited dissemination.

- Avoid the use of software obtained from private (bulletin boards, etc.) and untested commercial sources until it is thoroughly examined for errors and hidden, undesirable, routines.
- Validate and certify for accuracy all customized microcomputer programs.
- Establish user-accessible date/time fields showing the last time the database was modified.
- Mark all PC output with the date/time of production.
- Perform independent validation of data input and embedded formulas used in analytical reports.

- Evaluate operational procedures including data preparation and input handling procedures, program execution procedures, media procedures, and output handling and distribution procedures.
- Establish formal controls over software development, testing, and data integrity when important functions are being performed on personal computers. These controls should be applied to the use of generic software tools (e.g., spreadsheet and database management systems) to build complex applications as well as systems being designed and programmed in traditional programming languages.

### Software and File Backup

Security problems exist in software and file backup because of the uncertain backup/recovery procedures practiced by microcomputer users. Personal computers are exposed to data loss as a result of incorrect data entry, static discharge, dirty disks, and hardware failure. Personal computer users often are not aware of the need to take frequent backups or trained in the proper procedures. Critical data loss due to the lack of proper backups is one of the most serious problems related to personal computer usage.

The following security measures should be considered to minimize the backup exposure.

- Establish standards and guidelines requiring the purchase of a backup capability (streaming tape or external hard disk) with the purchase of PCs with internal hard disks. Also, when purchasing an external hard disk, another removable hard disk and drive or a streaming tape should be available for backup.
- Establish standards specifying data and program backup requirements. Three categories of backup (no backup, secure on-site backup, and on-site plus off-site backup) should be addressed.
- Provide user training on backup management procedures to include the use of hard and diskette systems as well as how to make and store backups of programs and data. Backup approaches include full-volume backups, incremental backups, and application-based backups.
- Require supervisors to audit PC backup activity and make ISSOs available to consult with users.

### Equipment Protection

The problem of protecting PCs and associated equipment from theft and physical damage is similar in many respects to that of protecting other valuable office equipment. PC component portability increases exposure to theft. Also, it is necessary to prevent unauthorized access to the inside of the PC equipment itself in order to protect against component theft, maintain configuration control, and protect the integrity of installed system security elements.

Environmentally, PCs are sensitive to the quality of the electrical power source. Static electricity discharges from personnel can damage integrated circuit components or semiconductor memory. Magnetic media, particularly diskettes, are vulnerable to damage.

The following security measures should be considered to protect PC equipment and components.

- Place personal computers in areas that have basic physical access control, such as locks on doors and employees present during working hours.
- Use equipment lockdown devices that secure the equipment to a table or other fixed object. Some devices also prevent unauthorized access to the system power switch and, therefore, can be used to protect against unauthorized PC usage.
- Place equipment on workstation enclosures that can be closed and locked when the equipment is not being used. These enclosures should also protect documentation, diskettes, and other equipment.
- Provide locked cases for diskette LAN storage and advise users of the following procedures for diskette care.
  - Always store in the protective jacket.
  - Protect from bending or similar damage.
  - Insert carefully into the drive.
  - Maintain a temperature range between 50 and 125 degrees Fahrenheit.
  - Avoid direct contact with magnetic fields.
  - Do not write directly on diskette jacket or sleeve.
- Conduct daily operations with a backup copy of all important disks. The recording surface of

diskettes may deteriorate in time; therefore, new master copies should be created every 18 months.

- Use antistatic sprays, carpets, or pads and instruct users to discharge built-up static charges by touching a grounded object other than the PC (a sign posted on each machine to remind users can deter potential trouble.)
- Place PC equipment on an isolated power source or install protection against power surges.

### **Personal Computer Networks and Terminal Emulation**

Serious security problems exist when PCs are connected in a network or when terminal emulation software is installed on PCs. In local area networks (LANs), all nodes can read data being transmitted (information is not limited to sender and receiver). Although logon information is usually suppressed, network diagnostic equipment is readily available that enables any node to read logon and other sensitive data. This is particularly serious in cases where a node is logging on to a host and, therefore, sending ID and secret password data through the network.

Terminal emulation software installed on PCs often provides for the logon sequence to be programmed and automatically transmitted to the computer when a function key is pressed. This practice can bypass logon security because anyone could initiate the logon of an authorized user simply by pressing the correct key.

The following security measures should be considered to minimize exposures when PCs are connected in a LAN or used as terminals.

- Establish standards that passwords will not be stored on PCs and automated logon routines will not include the user's password.
- Establish policy that sensitive information transmitted in a LAN will be encrypted so that only receivers with the appropriate decryption key can read it.

## **Responsibilities**

### **Operating Location (Operations/Group/Division/Plant)**

The protection of data, information, and reports is the responsibility of the senior line executive of each operating location of the corporation who controls the collection, maintenance, and use of the data.

The physical protection of computer hardware, software, and data stored within the computing facilities is the responsibility of the organization to which such resources are assigned for control and operation, as set forth in Corporate Directive F-10.

The location controller is responsible for ensuring that adequate internal controls are established to safeguard the location's assets.

Senior line executives at each operating location are responsible for protecting all computing resources assigned or controlled by their location. This responsibility is delegated to the location controller and is carried out through the location's Information Systems function, with Industrial Security or the equivalent function supplying protective services.

Location management must develop, implement, and maintain appropriate local standards and guidelines to ensure that the use of microcomputers is included in the location's system of internal controls.

### **Information Systems Center**

The corporate Information Security function will provide technical assistance and guidance to location management in its development, implementation, and maintenance of microcomputer security standards and guidelines. Also, Information Security will advise location management of technological advances in microcomputer security and/or emergent security exposures that might affect the adequacy of microcomputer security standards and guidelines.

**Corporate Offices—Internal Audit**

Will determine and apprise management of the following:

- Adequacy and effectiveness of security measures being applied to microcomputers and microcomputer networks.
- Results of evaluations of location's microcomputer security standards and guidelines.

**The Senior Vice President and Advisor to the President and CEO**

Will supervise implementation of and compliance with this Finance Policy and will issue supplemental procedures as required. This responsibility will be carried out through the Vice President and General Manager of the Information Systems Center. ■



# Microcomputer Data Security Solutions

## In this report:

Software Security Programs: Batten Down the Data Latches .....	2
Building Your Data Fortress .....	6
Virus Protection: Strong Medicine for a Fast Cure .....	6
Disk Mirroring: When Two Disks Are Better Than One .....	10

## Datapro Summary

Computer virus infections threaten the most sensitive of company data, and always take their victims by surprise. Three classes of software products—security, antivirus, and disk-mirroring programs—are designed to ward off and prevent viruses from infecting an entire system. Presented herein is a competitive look at security, antivirus, and disk-mirroring programs designed for PC, Macintosh, and UNIX computer platforms.

The first distress call Joan (not her real name) made to her computer dealer was just after electronic balls suddenly started bouncing across her PC screen. After more phone calls to the dealer and an unnecessary repair bill, Joan's Washington, DC law firm still found itself infected by a virus.

On the advice of her dealer, Joan had backed up her hard disk and paid the company \$100 to reformat the drive. In the process, she unwittingly infected the backup software and each of the floppy disks to which she transferred data. Because the virus resided in the master boot record, the intruder was untouched by the reformatting. Joan then decided to phone the National Computer Security Association (NCSA).

"Fortunately, the virus proved to be benign and didn't damage any of her data before I destroyed (the virus)," says David Stang, NCSA director. "But how she got the virus is a good story. Her law firm had been given an infected application disk that had infected three law firms in all. Viruses normally travel on bootleg stuff. If I had to pick

the most common cause of virus transmission, it would be that."

From private law firms to international corporations, organizations that rely on computers are vulnerable to downtime, data destruction, and monetary losses due to viruses, equipment breakdowns, or human intervention. In a recent survey, 53% of *BYTE* readers said their companies had suffered losses of critical data that cost an average of \$14,000 per occurrence. In addition, 28% said their companies had been victims of a computer virus that damaged or destroyed program and data files.

The best protection for critical data is an organization-wide implementation of security products and techniques that control access to hardware, software, and sensitive files. In this report, we look at three components of a data-security strategy, each of which operates as a supplement to standard procedures such as regular data backups.

First, we examine software security systems for PCs and Macintoshes. These programs provide a front-line defense against unauthorized access by restricting individual users from specific files. Although software solutions offer only very basic security and thus are not for everyone, many companies will find such programs the most economical and easiest security systems to install.

Second, we evaluate antivirus products for PCs and Macs. Although the threat of

This Datapro report is a reprint of "Rx for Safer Data" by Stanford Diehl, Stan Wszola, Bradley Kiewer, and Larry Stevens, pp. 218-235, from *BYTE*, August 1991. Copyright © 1991 by McGraw-Hill, Inc. Reprinted with permission.

viruses may outstrip actual incidences, prudent computing in the 1990s demands a simple but effective barrier to destructive programs.

Third, we look at disk-mirroring products for PCs, Macs, and UNIX, presenting a subset of redundant-storage techniques that keep disk drive failures from bringing companies to a standstill.

## Software Security Programs: Batten Down the Data Latches

The best DOS and Mac security products can help keep data from falling into the wrong hands or, more likely, from passing before curious eyes. In addition, security packages can provide you with outstanding control over your computer resources. Look for programs that let you selectively restrict access to files, directories, floppy disk drives, and even external ports. You should also be able to track program usage with extensive audit logging and prevent software piracy by making floppy disk drives read-only. In general, security software gives you a good idea of how your computer resources are being used.

The BYTE Lab looked at 12 software-only security programs for Macs and PCs (see Table 1). Except for the most demanding security requirements, software products are sufficient. Many hardware products offer a higher level of control, but they are more expensive and are not a practical solution for large installations. Also, software-only solutions provided an acceptable level of security for many *BYTE* readers: Only 7.4% blamed sabotage as the cause for lost data, while only 5.7% were victims of data theft. Software packages can deliver a significant layer of protection even if they cannot deter advanced hackers.

Security programs can make the hard disk drive "inaccessible" on a boot from a floppy disk (this is called *boot protection*). Nevertheless, an experienced programmer is still able to see the disk drive as a physical device and look at raw disk sectors.

LANs amplify security problems. Because most LANs send packets around from station to station, a data thief need only install a program to intercept the packets. *Data encryption* can solve this problem. Encryption is simply a means of encoding data so that it is unreadable. The recipient of the data must have a similar program to decrypt the data on the other end. Regard any unencrypted data sent over a LAN as fair game for thieves or the overly curious.

To deter the advanced intruder, we suggest a program of data encryption. If you intend to subvert even the most savvy intruders, we cannot recommend simple access control by software. In addition to a program of encryption, a hardware solution may help to meet your rigid security requirements. For most security needs, the products reviewed here will do the job.

In developing test scenarios for the security products, we did not expect any of these software solutions to be unbreakable. We sought only to ensure that they could not be bypassed by readily available tools. All these products successfully withstood the scrutiny of disk utilities such as the Norton Utilities and Mace Emergency Room. We also tried other obvious avenues, such as booting from a floppy disk and breaking out of the normal boot-up routine. In other words, we made sure that breaking the security capabilities of our test programs would not be a trivial matter.

We also wanted to see how well these products could protect your organization against willful destruction of

data. We came to a clear conclusion: They can't. If intruders are bent solely on destruction, without concern for recovering any usable data, they can succeed with little difficulty. In this case, no software access-control product that we reviewed can protect you. For the most part, we just performed a low-level format to erase a protected drive.

## PC Solutions

### Access II

Rather than go through the tedious step-by-step, directory-by-directory approach to administration that is taken by some security programs, Access II from Kinetic Software allows you to select multiple directories in one easy operation. In fact, the interface takes this tack for most administrative chores. When you set up an application, you can immediately select every user who should have access to it. It works the same with access rights. You simply fill in a table, giving all your users access rights in one easy step. Access options include floppy disk drives, serial and printer ports, and the system timer.

Access II also supports a Directory Assistance feature. When you're setting up directories, pressing a function key will call up all the directories on the disk, allowing you to pick and choose with simple cursor movements. In this way, you select as many directories as you need at one time.

Setting up an organized menu structure is also a simple task, or you can grant direct access to the DOS command line. All directory and resource restrictions will remain in force.

Access II takes up a good chunk of RAM (68 kilobytes), but you can select an option to decrease this to 10 KB. Kinetic also offers a hardware complement to Access II. If you anticipate heavy administration needs, Access II can make that job much easier.

### OnGuard

We ran into serious problems when working with OnGuard from United Software Security. The company insisted that some of the problems could not be occurring on the system. However, after many reinstallations and reconfigurations, we could not resolve the problems we encountered.

For instance, when we enabled boot protection, the floppy disk drive didn't work correctly. In fact, simply running a DIR command made the system hang. The program did not work properly with Windows. It successfully kept us out of protected directories, but then the entire system locked up.

Some security loopholes also caused us concern. We were able to TYPE files that had no rights assigned to them, and we could even change protected files. For example, we could not delete or edit the AUTOEXEC.BAT file when it was protected. However, we could copy the file into an unprotected directory, edit the file there, and then copy it back to the root. The program does have some nice touches, but its operation is too flaky. We cannot recommend it.

### PC/DACS

The PC/DACS system from Pyramid Development offers a decided enterprise-wide solution. It's clear from the organization of the program that setup and administration are geared toward large installations and multiple users. One feature, called "deploy," lets you set up a system configuration and then install preconfigured users across a

LAN. The software, by its very structure, promotes a complete program of security consciousness.

Setting up a configuration is easy. The program gives you a complete listing of directories (including subdirectories). You can select a directory with simple cursor key movements. Once you have highlighted the chosen directory, you can designate the entire directory for protection or select individual files. Next, you're given a selection of access rights to enable as required. The whole process proceeds simply and logically.

PC/DACS also supports a range of wild cards for more flexibility. For example, C:\\*. \* matches only files in the root directory, while C:\= includes all directories under the root. The same scheme will work with extensions (C:\\*.COM protects only COM files in the root, while C:\=.COM protects all COM files on the C drive).

Pyramid also provides a network solution with NET/DACS. A special windows module supports Windows 3.0 operation. From Windows or from DOS, users will simply not see any restricted files or directories. You can also specify encryption areas for automatic encryption of files.

### PC Watchman

PC Watchman from Harcom Security Systems works differently from many of the other programs reviewed here. Instead of specifically protecting files and directories, PC Watchman grants access to selected tasks. Even after you grasp this philosophy, the program is difficult to use.

Each user must be assigned to a group, and each group has a set of tasks assigned to it. This causes some annoying limitations. Assume you have a group called "accounting" with a set of users assigned to it. Suppose that you have one user who needs all the tasks assigned to "accounting" plus an extra task that you don't want the other accounting members to have. PC Watchman can do this, but the procedure is complicated.

Because you must assign a user account to one (and only one) group, the only way to give a single user access to multiple groups is to set up a new account each time. If a user needs access to a variety of groups, that user must remember passwords for every group. We did not object to the task philosophy so much, but we disliked PC Watchman's implementation of it.

### Protec

Protec from Sophco lets you control directory access, but its main strength is controlling specific programs. You can set up applications and, unlike with PC Watchman, easily assign the applications to users. By configuring data directories separately from the program directory, you can force applications to store data files in custom directories.

You enable many of Protec's features through independent programs. This makes the program somewhat difficult to use. For example, to set up applications for all users to execute, you must run the NOEXEC program and set up an INCLUDE list. You would then administer a command line such as -fc:\lotus\123.com -r -d i@c:\user\,menu\include.txt. Protec would be more inviting if this kind of operation were included in the standard interface. There is a nice file management utility for encrypting files.

Some serious limitations prevent us from recommending the software. With Protec running, we were unable to load all the necessary files for network access. Also, the program offers no access protection for floppy disk drives, external ports, or individual files. Other products that we reviewed provide more features and a smoother interface.

## Preventing Virus Infections

There are no known ways to make a general computing system completely immune from virus attacks, but the following practices can help you decrease the risks.

- Avoid using programs whose origin is unknown.
- Don't allow others to run their programs on your computer.
- Use only shrink-wrapped software packages and check them for viruses.
- Make regular antivirus checks of all your files.
- Back up your hard disk and store the backup in a safe place.
- Make backup copies of all your program and data files so that you can easily replace infected files.
- Beware of all programs downloaded from BBSes, or use only BBSes where all software is checked before it is posted (e.g., BIX).

### Security Guardian

Security Guardian from Command Software Systems is a powerful and flexible program for data security and control. Unfortunately, the very features that make it so powerful and flexible also make it difficult to learn. For each user, you can start off with all directories locked and then selectively enable them, or you can start with all the files enabled and then selectively lock individual ones. You can set up a directory table for the system or for individual users and switch back and forth among the configurations. Perhaps you can see the problem: Because there are so many different ways to set up the system, the process can quickly get confusing.

Menus are easy to establish, and by restricting access to DOS, you can make these menus the primary interface. This enhances control and ease of use. Other system controls prevent a secondary program from accessing DOS, disable Control-Break and Control-C, prevent writing to the hard disk, and allow complete read/write control over the floppy disk drive. You cannot use Security Guardian to restrict the use of printer or communications ports.

Security Guardian can also keep extensive logs. Besides the usual tracking of resource use, the program will record inactive time to show how long the system remains unused. This comes in handy when you want to determine PC utilization. Security Guardian consumes a scant 5 KB of RAM and is a solid product with outstanding flexibility.

### Watchdog

Watchdog from Fischer International Systems has a strong, well-deserved reputation for security. If your principal concern is safeguarding sensitive data, Watchdog is a top choice. In addition to an impressive software approach, Watchdog supports a hardware option for those workstations requiring stronger access control.

**Table 1. Security Program Features**

	PC Programs						
	Access II	OnGuard	PC/DACS	PC Watchman	Protec	Security Guardian	Watchdog
<b>Price</b>	\$165	\$295	\$249 (single user)	\$195 (one site)	\$295	\$250	\$295
<b>RAM used (min./max.)</b>	10/68 KB	20 KB	8/39 KB	12 KB	54 KB	5 KB	17/58 KB
<b>Number of users</b>	16	24	Unlimited	Unlimited	52	Unlimited	Unlimited
<b>Boot protection</b>	●	●	●	●	●	●	●
<b>Time-out</b>							
Log-out	●	●	●	●	●	●	●
Screen save	●	○	●	●	●	●	●
Continue processing	●	○	●	●	○	●	●
<b>Restrict</b>							
Directories/folders	●	●	●	●	●	●	●
Files	●	●	●	○	○	●	Option
Serial ports	●	○	●	●	○	○	●
Printer access	●	○	●	●	○	○	●
Floppy disk drives (R/W)	●	●	●	●	○	●	●
Network drives	●	○	●	●	○	○	○
<b>Encryption</b>							
Proprietary	Option	Option	●	●	●	●	●
DES	Option	Option	●	●	●	●	●
Automatic	Option	○	●	○	●	●	●
<b>Menuing interface</b>	●	○	●	●	●	●	●
<b>Transparent DOS interface</b>	●	●	●	●	●	●	●
<b>User log-in scripts</b>	○	●	●	●	○	●	●
<b>Hide restricted directories</b>	●	○	●	○	○	●	●
<b>Works with Windows 3.0</b>	●	○	●	○	●	●	●
<b>Audit trail</b>							
Invalid log-in attempts	●	●	●	○	●	●	●
Program usage	●	●	●	●	●	●	●
Track time on system	●	●	●	●	●	●	●

Note: The best security programs restrict access to files, directories, floppy disk drives, and external ports, as well as providing extensive audit trails (● = yes; ○ = no; N/A = not applicable).

(1) All four Kent Marsh products are designed to work together as a single security system.

(2) Yes when used with QuickLock.

(3) Yes when used with NightWatch.

However, we weren't entirely happy with the Watchdog interface. You must first add directories to the system table. Once a directory is added, you then give users specific rights to it: Adding all these directories is tedious, especially since you can't call up a listing of current directories on your hard disk.

Watchdog grants direct access to the DOS command prompt, and, if you prefer menus, it has an excellent menu builder. With Watchdog's global libraries, you can give all users access to files in a directory while still protecting those files from unauthorized changes. Transparent data encryption is another big win on the security front.

Watchdog offers impressive resource control, including restrictions on printer and communications ports. Establishing these restrictions is a breeze. Watchdog worked well with Windows 3.0.

## Mac Solutions

### AME

AME (Access Managed Environment) from Casady & Greene is the most complete of the Mac software packages in this review. Besides protecting your hard disk files, it takes control of serial ports and printers as well as the floppy disk drive. Another nice feature is AME's "Trusted Software" list. The administrator can specify that certain applications on the hard disk are safe for the general user base to access. Running any other application from either the hard disk or a floppy disk will bring about a security violation. Violations cause the activation of an alert box and/or an audio alert. If you activate all the port and file security, you can sleep better at night, knowing your data is secure.

**Table 1. Security Program Features (Continued)**

	Mac Programs							
	AME	DiskLock	Empower II	FileGuard	FolderBolt (1)	MacSafe II (1)	NightWatch (1)	QuickLock (1)
<b>Price</b>	\$279	\$189	\$296	\$249 (one user)	\$129.95	\$189.95	\$149.95	\$59.95
<b>RAM used (min./max.)</b>	164 KB	256 KB	1 MB	200 KB	384 KB	384 KB	384 KB	384 KB
<b>Number of users</b>	Unlimited	1	Unlimited	Unlimited	Unlimited	Unlimited	255	N/A
<b>Boot protection</b>	●	●	●	●	N/A	N/A	●	N/A
<b>Time-out</b>								
Log-out	●	○	●	●	N/A	N/A	○ (2)	○ (3)
Screen save	○	●	●	●	N/A	N/A	○	●
Continue processing	●	●	●	○	N/A	N/A	○	●
<b>Restrict</b>								
Directories/folders	●	●	●	●	●	N/A	N/A	N/A
Files	●	●	●	●	●	●	N/A	N/A
Serial ports	●	○	○	○	○	N/A	N/A	N/A
Printer access	●	○	○	○	○	N/A	N/A	N/A
Floppy disk drives (R/W)	●	○	●	●	○	N/A	N/A	N/A
Network drives	○	○	○	○	○	N/A	N/A	N/A
<b>Encryption</b>								
Proprietary	●	●	●	●	●	●	○	N/A
DES	●	●	●	●	○	●	○	N/A
Automatic	●	Option	●	●	N/A	●	○	N/A
<b>Menuing interface</b>	●	●	●	●	●	●	●	N/A
<b>Transparent DOS interface</b>	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
<b>User log-in scripts</b>	●	○	N/A	○	○	○	○	○
<b>Hide restricted directories</b>	●	●	●	●	●	N/A	N/A	N/A
<b>Works with Windows 3.0</b>	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
<b>Audit trail</b>								
Invalid log-in attempts	●	●	●	●	N/A	N/A	●	○
Program usage	●	○	●	○	N/A	N/A	○	N/A
Track time on system	●	○	●	●	N/A	N/A	○	N/A

Note: The best security programs restrict access to files, directories, floppy disk drives, and external ports, as well as providing extensive audit trails (● = yes; ○ = no; N/A = not applicable).

(1) All four Kent Marsh products are designed to work together as a single security system.

(2) Yes when used with QuickLock.

(3) Yes when used with NightWatch.

Even with all these features, we simply can't recommend AME. All security software has to interface with the Mac operating system at a fairly low level. Most of the software packages manage to tell you about security violations in a polite, controlled manner. AME, on the other hand, brings up a dialog box and then returns an error to the operating system. The Mac dutifully displays a message along the lines of "Unable to . . . (source disk was modified during copy)" and then aborts the operation. A simple click in the wrong spot shouldn't trigger such horrific alerts.

Even worse is the sequence that occurs if you boot an AME-protected Mac from a floppy disk. The Mac reports that your hard disk has to be formatted. If you tell it to proceed, the dialog boxes suggest that it's actually formatting the drive. Again, other packages handle errors much

better. AME may be excellent at protecting data, but the error displays are enough to scare even the most knowledgeable Mac user.

### DiskLock 2.0

DiskLock 2.0 from Fifth Generation Systems (the Suitcase II people) is a simple utility that prevents anyone else from booting your Mac or accessing your files. When you boot the machine, a dialog box asks for the password. If you don't know it, the machine won't boot. You can also opt to use the FolderLock feature to encrypt entire folders. Even if you choose to let someone else use your Mac, you can still restrict access to specific folders. The FileProtect feature uses your choice of three security levels: FastLock (simple but speedy), QuickCrypt (somewhat slower), and DES (slow but secure).

If you leave the machine idle for a prescribed amount of time, a screen saver kicks in and locks the machine. When

you return, you simply reenter your password to start working again. You can opt to have DiskLock automatically relock any FolderLocked folders before it blanks the screen.

DiskLock is simple yet effective protection for a Mac that is primarily used by one person. It should prove to be more than enough for many people's security needs.

### Empower II

Empower II from Magna gives you boot protection, file and folder protection (à la AppleShare; see FileGuard, below), and controlled access to the floppy disk drives. It won't do anything to protect your serial ports or printers.

The system administrator creates the users and groups and sets up specific access rights. Each folder receives Owner, Group, and Everyone rights, just as with FileGuard. If you walk away from the Mac, the screen saver blanks the screen and waits for a password. Data encryption is optional, and so is a complete audit trail of file access, user log-ins, and security violations.

Empower II's operation is transparent, and security violations are handled without making you feel like a criminal. Empower II is one of those rare products that do exactly what they advertise without any surprises.

### FileGuard

FileGuard from ASD Software uses an interface familiar to anyone who's ever used an AppleShare file server. Each folder allows different access levels to the folder's owner, a specific group, or everyone. For each of the three user categories, you specify whether a user can make changes (including deleting files), see the files (read and execute rights), or simply see any contained folders. The FileGuard administrator creates all the users, assigns them to groups, and grants access to public folders and applications. If you like, FileGuard will encrypt selected documents to make them harder for other users to access. If you choose the automatic encryption feature, FileGuard will bring up a dialog box and ask for a password whenever you save a file. After you leave your application, any "automatic" files will be reencrypted by the program.

One especially nifty feature is the automatic time-out on applications. The administrator can decide that a particular application can be accessed only a set number of times or run for a preset period of time. After that many executions or that time period, you are alerted with a dialog box and the application is terminated.

The software protects files only on your hard disk—if you want to prevent people from inserting floppy disks in your machine or accessing your serial ports from communications software, you will have to look elsewhere. Another potential security leak is that you can't restrict the format of the passwords (they are always greater than three characters, and any combination of letters is permitted).

### NightWatch

Kent Marsh has a modular solution to Mac system security. To keep other people from starting up your Mac, you use NightWatch. This product simply modifies your hard disk drive so that if you boot from a floppy disk, the hard disk won't mount. When you are ready to shut the machine off, a shutdown utility relocks the hard disk drive to secure it. To reboot the machine, you need a User Disk, which contains a database of users and passwords. You boot from this special floppy disk and identify yourself by name. If

you enter a correct password, the hard disk drive is unprotected and your Mac reboots. This approach is kludgy, but it works.

Another module is QuickLock—a security screen saver. If you leave the machine unattended, the screen goes blank. You have to enter a password to awaken it. If you don't know the password, the machine will shut down. If you're also running NightWatch, QuickLock will automatically execute NightWatch's shutdown facility.

Unfortunately, this procedure has a serious bug. Before NightWatch will reprotect your disk, it brings up a screen and gives you the opportunity to cancel. Let's say you don't know the machine's password. The machine will try to shut down and activate the NightWatch screen. You simply click on Cancel, and you're back at the Finder. Actually, it's worse than that. During testing, this sequence repeatedly destroyed the System and Finder, causing the screen fonts to disappear.

The last two pieces of Kent Marsh's security solution are MacSafe II and FolderBolt. MacSafe is an application that lets you create a "safe" in which to store specially encrypted files. Anyone can see the safe, but no one except you can use or delete it. FolderBolt is a file access-control package similar to DiskLock's FolderLock. The modular approach is a good idea, but the pieces simply don't work well together. FolderBolt is worth a look, as is QuickLock. We didn't like NightWatch's key disk implementation, and it was dangerous to have NightWatch and QuickLock working together.

## Building Your Data Fortress

No piece of software alone will solve your microcomputer security needs. But for establishing a complete program of resource control for your corporation's data, you won't find a better platform than PC/DACS. No other system provides such a complete solution for program control. PC/DACS emphasizes an overall approach for numerous installations, as reflected in the "deploy" feature as well as in the design and philosophy of the software as a whole.

PC/DACS is flexible enough to meet the needs of a wide range of users. It also does an excellent job of staying out of your way as you work. Novice users will see only those resources pertaining to them. More advanced users can command impressive levels of control. And the whole security mechanism is tied into a consistent backbone of user management, resource control, and audit logging. With PC/DACS as a guide, you'll find it easier to build a total security solution.

Depending on how you use your Macintosh, either DiskLock or Empower II should do the trick. Neither one has the company wide support of PC/DACS, but both do the job. DiskLock is our choice for a single-user machine, and we'd pick Empower II for any machine with multiple users who aren't concerned about I/O access control. If you need better control Empower II provides, your only choice is AME; if you go with this program, be prepared for lots of panicky phone calls from your users.

## Virus Protection: Strong Medicine for a Fast Cure

The surge in computer communications has spawned a great deal of confusion and fear about computer viruses. BBSes are awash with talk—some based on fact, some on

## Affordable Mirroring for UNIX

Data protection through disk mirroring isn't new; most of the ideas now available for PCs and workstations have been used for years on minicomputers and mainframes. Powerful, expensive UNIX file servers commonly have disk mirroring, as well as automatic backups and power protection, as part of their typical configuration. But most people moving into UNIX today don't have such deep pockets. Fortunately, there are solutions that can satisfy both your needs and your budget. As an example, I'll describe one system that provides a well-rounded data-integrity solution.

The Altos system 5000 is a 33-MHz 486 EISA-based system that is tuned for file and compute serving. It is also built to run UNIX exclusively. Altos has its own "private-label" version of SCO UNIX System V (which is derived from AT&T System V release 3.2). Altos UNIX 5.3.2, as it is called, includes several enhancements added by Altos with data protection in mind.

At the heart of Altos's failure protection is support for multiple SCSI channels. While mirroring multiple drives on the same SCSI host adapter is fair insurance against drive failures, it can't do much to guard against host adapter or cable failure. The System 5000 can support up to three independent SCSI channels, each handling up to seven devices. One channel resides on the system's base I/O board, and the other two are part of the optional High Performance File Processor.

Altos's disk-mirroring capability is part of the standard operating-system software. Drives are mirrored by *divisions* (i.e., blocks of disk cylinders, similar to DOS partitions). An Altos utility, *vdutil* (virtual disk utility), presents a menu-driven interface to the mirroring system.

Through *vdutil*, you can also stripe and span multiple drives, enhancing performance and extending available contiguous storage.

When mirroring is first applied, the source division can be one that already exists or a new division on a freshly formatted hard disk drive. I'm glad this choice is available—you can forgo mirroring at first and phase it in as your needs grow without reloading your data.

Once you have enabled mirroring, the "from" and "to" divisions are combined into a single UNIX virtual device. From then on, the mirrored divisions can be treated like any physical UNIX device. Since the mirroring is part of the operating system, no external utilities or background processes are needed to maintain the mirroring facility.

All types of disk failures are trapped by the mirroring system. As long as it can respond (i.e., the system hasn't crashed, and the SCSI and system buses haven't frozen from a short or some such incident), the mirroring system reports the failure to the system console and shuts down I/O on the faulty drive of the mirrored pair. Applications using that drive, if they

are actively involved in I/O, will receive an error during the change. The switch doesn't take much time, so applications that are built to automatically retry failed I/O operations will hiccup and then recover.

I was a bit disappointed that the system doesn't make the switch without applications being aware of it, but an error notification is preferable to the passing of bad data. And only processes that are actively reading from or writing to the failed drive will be affected.

Altos's mirroring does carry measurable overhead; after all, all data has to be written twice. The versatility of SCSI improves things considerably, though: Write operations are slowed by about 20%, while reads are about 10% slower. Performance would have been better if I had mirrored between two controllers (they would operate asynchronously), but even these numbers aren't too frightening considering the level of protection offered and the solid performance of Altos's file I/O.

I tested the mirroring software by hooking up a pair of drives: a Plus Development Impulse 3½-inch 168-megabyte drive and a Micropolis 5¼-inch 330-MB drive. I placed them on the System 5000's internal SCSI channel following the system's single internal drive and configured the new drives for 168 MB of mirrored space. It took only about 10 minutes to get the drives fully on-line (they were already formatted).

I invented a simple test that created a huge (16-MB) file and then read its contents repeatedly until interrupted or an error occurred. I tested

the failure recovery by pulling the power connector out of one of the drives while the read cycle was under way. Sure enough, the system spat out error messages until the threshold I had set was reached, and it then took the primary drive off-line and re-routed everything to the "to" side of the mirrored pair. My test program sensed an error at the time of the switch (the switch will not take place if the drive returns to service before the threshold is exceeded), but no bad data was passed, and the file was in precisely the same state following the switch as it was when the failure occurred.

Once the system takes a drive off-line, it remains off until the system administrator returns it to service. In some cases, the primary drive can be reactivated after a cable is tightened (or, in my test, the power is returned), and users will experience no interruption of service. At some point, however, the data on the two drives will have to be reconciled. This involves a high-speed copy of the entire division between the drives. The *vdutil* program takes the mirrored pair off-line for this; I'd have preferred the option of an on-line restore. Even so, Altos's goal of minimal downtime is largely realized. And, more important, no data is lost.

Disk mirroring is combined with on-line diagnostics and power-failure handling (with the optional UPS) to give the System 5000 a well-rounded data protection solution. Altos's solution is not the only one, but its transparency and affordability place it among the best.

myth—of viruses and the damage they cause. The following is a rational look at viruses and the latest antivirus software for PCs and Macs (also see the sidebar "Preventing Virus Infections").

First things first: A computer virus is a program that can alter, without your knowledge, the way your computer operates or modify the programs and data files stored on

your computer. The virus copies itself onto other executable programs by adding to or overwriting the existing program code and thereby damaging the program. Whenever you run an infected program, the virus code is executed first. The virus then goes on to infect other programs. Some viruses operate as TSR programs and can hide in RAM.

**Table 2. Antivirus Arsenals**

PC Programs	Price	Updates	Viruses Detected and Removed							
			1701	1704	Izrael	Musician	Vienna	W13_A	W13_B	Jocker
Central Point Anti-Virus	\$129	BBS/ quarterly	●	●	●	●	●	●	●	○
Certus 2.1	189	BBS	●	●	●	●	●	●	●	○
Data Physician	49	BBS	●	●	●	●	●	●	●	○
Dr. Solomon's Anti-Virus Toolkit	279.95	Quarterly	●	●	●	●	●	●	●	●
Norton Anti-Virus 1.0	129.95	BBS	●	●	●	●	●	●	●	○
Virex PC	\$129.95	Quarterly	●	●	●	●	●	●	●	●
VirusCure	99.95	BBS	●	●	●	●	●	●	●	●
VirusSafe	80	\$60/ quarterly	●	●	●	●	●	●	●	○
Vi-Spy	250	Quarterly	●	●	●	●	●	●	●	●
Viruscan	15-35	BBS	●	●	●	●	●	●	●	●
<b>Mac Programs</b>			<b>Modm</b>	<b>WDEF</b>	<b>nVIR</b>					
Disinfectant 2.4	Free	BBS	●	●	●					
Symantec Anti-Virus for the Macintosh	99.95	BBS	●	●	●					
Virex	99.95	\$75/year	●	●	●					
VirusDetective/ VirusBlockade	Share- ware	BBS	●	●	○					

Note: The test results show that some antivirus programs couldn't recognize the Jocker virus (● = yes; ○ = no).

Antivirus programs detect, identify, and remove the intruders. Typical antivirus programs will scan RAM and hard and floppy disks for infections. These programs can identify the virus infecting the system and tell whether the virus resides in memory, the boot sector, the partition table, or a file. You can then delete the infected file or disk sector, or you can repair the file by deleting only the virus program code.

An antivirus program searches the system for program-code sequences or patterns that are unique to each computer virus and then reports their presence. This method works for viruses that the antivirus program recognizes. Many programs let you enter the identifying characteristics of a new virus into a data file to help you cope with newer strains.

Many of the antivirus packages provide for system protection and immunization. A TSR antivirus program will constantly monitor your system looking for virus activity. Some antivirus programs will log all the program files on your hard disk and calculate a cyclic redundancy check (CRC) on each file. The original checksum of the file is then compared to the current checksum for discrepancies and possible infections.

In this roundup we chose 10 PC and four Mac products (see Table 2). We limited our tests to software solutions because of the effectiveness, ease of installation, and economy of these programs compared to hardware solutions.

#### Isolation Test

To test the packages, we set up a PC and a Mac in an isolation area in the BYTE Lab. On the PC side, we ran tests

using eight of the most pervasive and destructive viruses in circulation (see Table 2). All the programs identified and controlled most of the viruses. A couple of the programs, such as Viruscan on the PC and Disinfectant 2.4 on the Mac, caught all our test viruses. This is a function primarily of how often the software is updated. New viruses appear every day. When choosing an antivirus program, find out how often updates are issued and how easily you can receive them. For most people, downloading updates from a BBS is the quickest way to stay ahead of evolving viruses.

#### DOS Protection

Among the 10 PC-based antivirus packages we tested, Dr. Solomon's AntiVirus Toolkit from Ontrack Computer Systems, Virex PC from Microcom, VirusCure from International Microcomputer Software, Viruscan from McAfee Associates, and Vi-Spy from RG Software Systems identified all our test viruses. These five also represent the high and low in prices, ranging from \$279.95 for Dr. Solomon to \$35 for all the modules in the shareware version of Viruscan.

#### Dr. Solomon

Dr. Solomon's Anti-Virus Toolkit's menu-based front end integrates several programs to scan a disk for a virus, check for viruses on system boot-up, prevent a particular virus from infecting your disk, remove boot sector and partition sector viruses, and view a hard or floppy disk sector or view a file. The documentation details virus types and ways to remove them.



## Company Information

**Altos Computer Systems**  
(System 5000)  
2641 Orchard Pkwy.  
San Jose, CA 95134  
(800) 258-6787  
fax: (408) 433-9335

**ASD Software, Inc.**  
(FileGuard)  
4650 Arrow Hwy., Suite E-6  
Montclair, CA 91763  
(714) 624-2594  
fax: (714) 624-9574

**Casady & Greene, Inc.**  
(AME)  
22734 Portola Dr.  
Salinas, CA 93908  
(408) 484-9228  
fax: (408) 484-9218

**Central Point Software, Inc.**  
(Central Point Anti-Virus)  
15220 Northwest  
Greenbrier Pkwy.  
Beaverton, OR 97006  
(800) 888-8199  
(503) 690-8090

**Certus International**  
(Certus 2.1)  
13110 Shaker Sq.  
Cleveland, OH 44120  
(800) 722-8737  
(216) 752-8181

**Command Software Systems, Inc.**  
(Security Guardian)  
1061 East Indiantown Rd.,  
Suite 500  
Jupiter, FL 33477  
(800) 423-9147  
(407) 575-3200  
fax: (407) 575-3026

**Corporate Systems Center**  
(FastCache 32)  
730 North Pastoria Ave.  
Sunnyvale, CA 94086  
(408) 737-7312  
fax: (408) 737-1017

**Digital Dispatch, Inc.**  
(Data Physician)  
55 Lakeland Shores Rd.  
Lakeland, MN 55043  
(800) 221-8091  
(612) 436-1000  
fax: (612) 436-2085

**EliaShim Microcomputers, Inc.**  
(VirusSafe)  
520 West Highway 436,  
Suite 1180-30  
Altamonte Spring, FL 32714  
(800) 771-7233  
(407) 682-1587  
fax: (407) 869-1409

**Fifth Generation Systems, Inc.**  
(DiskLock 2.0)  
10049 North Reiger Rd.  
Baton Rouge, LA 70809  
(800) 873-4384  
(504) 291-7221  
fax: (504) 295-3268

**Fischer International Systems Corp.**  
(Watchdog)  
4073 Merchantile Ave.  
Naples, FL 33942  
(800) 237-4510  
(813) 643-1500  
fax: (813) 643-6357

**Golden Triangle Computers, Inc.**  
(DiskTwin 2.0)  
4849 Ronson Court  
San Diego, CA 92111  
(619) 279-2100  
fax: (619) 279-1069

**Harcorn Security Systems**  
(PC Watchman)  
130 William St.  
New York, NY 10038  
(212) 766-1802  
fax: (212) 732-0596

**International Microcomputer Software, Inc.**  
(VirusCure)  
1938 Fourth St.  
San Rafael, CA 94901  
(415) 454-7101  
fax: (415) 454-8901

**Jeffrey S. Shulman**  
(VirusDetective/  
VirusBlockade)  
P.O. Box 1218  
Morgantown, WV 26507

**John Norstad**  
(Disinfectant 2.4)  
Northwestern University  
2129 Sheridan Rd.  
Evanston, IL 60208

**Kent Marsh, Ltd.**  
(FolderBolt, MacSafe II,  
NightWatch, QuickLock)  
1200 Post Oak Blvd.,  
Suite 210  
Houston, TX 77056  
(800) 325-3587  
(713) 623-8618

**Kinetic Software Co.**  
(Access II)  
240 Distillery Commons  
Louisville, KY 40206  
(502) 583-1679  
fax: (502) 583-1104

**Lomas Data Products, Inc.**  
(LDP Cache II)  
182 Cedar Hill St.  
Marlborough, MA 01752  
(508) 460-0333  
fax: (508) 460-0616

**Magna**  
(Empower II)  
2540 North First St.,  
Suite 302  
San Jose, CA 95131  
(408) 456-2500  
fax: (408) 943-0651

**McAfee Associates**  
(Viruscan)  
4423 Cheeney St.  
Santa Clara, CA 95054  
(408) 988-3832

**Microcom, Inc.**  
(Virex, Virex PC)  
P.O. Box 51489  
Durham, NC 27717  
(919) 490-1277

**Ontrack Computer Systems, Inc.**  
(Dr. Solomon's Anti-Virus  
Toolkit)  
6321 Bury Dr., Suite 15-19  
Eden Prairie, MN 55346  
(612) 937-1107  
fax: (612) 937-5815

**Pyramid Development Corp.**  
(PC/DACS 2.02)  
70 Inwood Rd.  
Rocky Hill, CT 06067  
(203) 257-4223  
fax: (203) 257-4245

**RG Software Systems, Inc.**  
(Vi-Spy)  
6900 East Camelback Rd.,  
Suite 630  
Scottsdale, AZ 85251  
(602) 423-8000  
fax: (602) 423-8389

**Sophco, Inc.**  
(Protec)  
P.O. Box 7430  
Boulder, CO 80306  
(303) 444-1542  
fax: (303) 444-1454

**Symantec Corp.**  
(Symantec AntiVirus  
for the Macintosh/  
Norton AntiVirus 1.0)  
10201 Torre Ave.  
Cupertino, CA 95014  
(800) 441-7234  
(408) 253-9600  
Virus Newsline:  
(408) 252-3993

**United Software Security, Inc.**  
(OnGuard)  
8133 Leesbury Pike,  
Suite 380  
Vienna, VA 22182  
(703) 556-0007  
fax: (703) 734-3368

**Unitrol Data Protection Systems, Inc.**  
(Immunity Plus)  
815 Hornby St., Suite 604  
Vancouver, BC,  
Canada V6Z 2E6  
(604) 681-3611  
fax: (604) 687-0814

**Other products used in our testing:**

**Adaptec, Inc.**  
(AHA-1542B)  
691 South Milpitas Blvd.  
Milpitas, CA 95035  
(408) 945-8600  
fax: (408) 262-2533

**Micropolis Corp.**  
(1684)  
21211 Nordhoff St.  
Chatsworth, CA 91311  
(818) 709-3300  
fax: (818) 709-3396

**Plus Development Corp.**  
(Plus Impulse 170S)  
1778 McCarthy Blvd.  
Milpitas, CA 95035  
(800) 624-5545  
(408) 434-6900

### Virex PC

Virex PC is a two-part package. The first program scans RAM and your floppy disk looking for a virus. The scanner can also remove a virus and restore a file. The other half is a TSR monitor program. You register all the programs you normally use with this TSR to grant them access to your hard disk.

The Virex program also calculates a CRC checksum to create a signature for the registered programs. These checksums are stored in a data file. The Virex TSR will alert you if an attempt is made to format your hard disk, if any attempt is made to write to the hard disk, if any program attempts to terminate and stay resident, if an unregistered program is run, or if a registered program's checksum is modified, and it will alert you if any user-specified operations are attempted.

### VirusCure

You can use VirusCure as a standalone virus scanner, or you can install it to run automatically on boot-up. The installation program initially scans your hard or floppy disks for viruses. You have the option to remove a virus from an infected file and repair the file. VirusCure also creates CRC checksum signature files for all the files in your system. The program alerts you if any changes are made to the boot sector, partition table, DOS, or other critical system files. You can repair the boot sector and partition table using a reconstruct option. VirusCure also features two TSRs that constantly monitor your system and check for viruses every time you boot up. VirusCure is based on the highly regarded McAfee software.

### Viruscan

The Viruscan series of shareware programs from McAfee Associates includes a virus disinfection program, a scanning program, and an automatic log and file check. The shareware programs frequently updated to incorporate information about new viruses. You also can try the programs before you buy them.

### Vi-Spy

Vi-Spy is designed to run from a floppy disk or be installed on your hard disk. The floppy disk version lets you check several computers. The hard disk version uses the Disk Watcher TSR to scan the RAM and your hard disk and monitor the system's activity. If a virus is detected, Vi-Spy will display the filename, size, date and time, and the name of the virus. Vi-Spy then asks if you want the infected file wiped out. Vi-Spy scans the boot sector and partition tables and can repair those areas on the hard disk.

### Mac Protection

Disinfectant 2.4 from John Norstad is a freeware program, widely available on many BBSes, that is frequently updated to incorporate information about new viruses. You can run Disinfectant from a floppy disk or install it on your hard disk. Clicking on the No Viruses icon pops up a menu from which you select a scan of your hard and floppy disks, remove viruses, and install a protection INIT to prevent a reinfection. On-line documentation is included. (VirusDetective and VirusBlockade from Jeffrey S. Shulman are also freeware products available on BBSes.)

Symantec AntiVirus for the Macintosh consists of two components: SAM Intercept and SAM Virus Clinic. SAM Intercept consists of an INIT and a cdev. The INIT alerts you to any activity on the system that might be a virus attempting to infect your files. SAM Virus Clinic is a

stand-alone program for scanning files, folders, and hard disks for the presence of known viruses.

Virex from Microcom also consists of two components: the Virex INIT/cdev and the stand-alone Virex scanner. The Virex INIT examines floppy disks whenever they are inserted into the disk drive and compares each file with its prerecorded checksum to look for changes. The Virex scanner lets you examine selected files and volumes and remove detected viruses.

### The Prescription

For the PC, one of the best sets of programs is the shareware Viruscan series. The programs caught all the test viruses, the price is reasonable, regular updates are available on many BBSes, and you can try the software before you buy. The highly regarded McAfee technology is also the basis for commercially available packages.

For the Mac, Disinfectant 2.4 did equally well. It caught all the test viruses, the price is right, and regular updates are available on BBSes. The screen display is straightforward and easy to use, and on-line documentation is included.

## Disk Mirroring: When Two Disks Are Better Than One

As large hard disk drives measure their capacities in gigabytes, backing up becomes an increasingly important issue. On a single-user system, daily backups place as much as one day's worth of work at risk should the disk crash at quitting time. But on a busy network, several people can each lose up to a day's worth of work—a potentially expensive and frustrating problem.

As a solution, several companies in the microcomputer industry have borrowed a concept from the world of larger systems: RAID (for redundant arrays of inexpensive disks).

RAID is defined in five levels. At the simplest, the technique mirrors data to two drives. Mirroring systems are convenient because they require no action on the part of the user. Data backup is continuous—once a failure occurs, the system can be up and running with little delay. And finally, as drive prices continue to fall, mirroring is becoming price competitive with removable media for some systems.

Currently, several mirroring systems are available for PCs, Macs, and UNIX platforms (see the text box "Affordable Mirroring for UNIX" for one example of Unix-based mirroring). Although not as powerful as full-blown RAID systems, these systems do provide an extra measure of real-time data security. We looked at a representative sample of mirroring products for the PC and Mac platforms. As we discovered during our tests, the different implementations are better suited to particular systems or tasks.

### PC Mirroring Choices

When using Immunity Plus from Unitrol Data Protection Systems, you must format the disk drives with DOS FDISK or ONTRACK Disk Manager, and the operating system must be DOS. This keeps you from running on servers such as NetWare, OS/2, or UNIX, which use their own disk formats and drivers.

The resynchronization (i.e., initialization) took about half an hour on our system. Our first drive was one cylinder shorter than the second, which caused a mirroring problem (one partition could not be mirrored). It would be nice if the utility could mark small reserved areas on the

primary partition to reduce the size of the partitions and allow mirroring of the entire drive. Of course, in a situation like this, you could leave a tiny partition at the end of the disk or swap the drives, but that is inconvenient.

### Immunity Plus

Immunity Plus was the only tested system that allowed partial mirroring. It is thus possible to mirror only critical data and increase the total space available on the drives. It also allows you to use drives of different sizes to their full capacity. For example, you could mirror a 40-megabyte drive to an 80-MB drive and use the remaining 40 MB on the larger drive for unprotected storage.

We could not test a total drive failure because whenever we removed power from a drive, the drive controller quit working. Thus, the software never got a chance to "fix" the problem. We did disable mirroring, made changes to one of the drives, and attempted to reestablish mirroring. The software caught the discrepancy, flashed a warning message, and refused to enable the mirroring. A similar process is supposed to occur during a real failure. Note that it is impossible to reboot if the primary drive fails, since mirroring is not enabled until the device driver has been loaded.

### LDP Cache II

LDP Cache II from Lomas Data Products is register compatible with the Western Digital WD1003-WA. As a result, it can be used without special drivers (an important consideration for systems running UNIX, OS/2, or NetWare). However, it can also run in an enhanced mode that improves performance when special drivers or the on-board ROM is used.

The controller cannot implement parallel reads. However, its 4 MB of RAM cache negates much of the need for such a feature. The cache can be enabled for write-through mode (the drive is updated immediately) or buffered mode (writes are delayed until the buffer is full or the drive is not busy). Because both drives write simultaneously, the write performance is the same as with a single drive.

When we cut off power to the primary drive under OS/2, OS/2 system response went dead. The system would not boot from the second drive, although we were able to get up and running by removing the "dead" primary drive and restrapping the secondary as the primary. The system then responded normally (other than giving a warning message at boot time that a drive error had occurred).

To get a better feel for the process, we tried the procedure again under DOS while running a disk search with Norton Utilities. When we removed power from the primary drive, the search continued, but at a greatly diminished pace (at the rate of one sector every few seconds, rather than many sectors per second).

LDP Cache II continues operations on both drives until a write failure occurs. Thus, when a drive totally fails, the system may respond quite slowly as it first attempts to read the primary drive and then switches to the secondary drive on each read failure. Of course, with a partial failure, such as a bad sector, the performance difference should be negligible, and you retain the benefits of continual backup for the remaining sectors.

### FastCache 32

FastCache 32 from Corporate Systems Center represents something of a hybrid hardware/software approach. The mirroring function is controlled by the BIOS, so it is actually software controlled. But, unlike Immunity Plus, it

does not use space in RAM. Because writes are not cached and access is sequential (writing first to one drive and then the next), mirrored writes are not as fast as with LDP Cache II. Nevertheless, writes outperformed our MFM-based system by a significant margin. FastCache 32 is not register compatible with the WD1003-WA, so we could not use it with OS/2 (instead, we used IBM PC LAN 1.3).

Like LDP Cache II, FastCache 32 has a 4-MB RAM cache and does not read from both drives. When we removed power from the primary drive, read time increased but was not nearly as slow as with LDP Cache II. Although the system continued to run, we could not get it to reboot without restrapping the secondary drive as primary. On the other hand, you probably would not want to boot from only a secondary drive because of the performance disadvantage.

The floppy disk drive always reported errors during our system's power-on self test. However, the drive functioned normally. Since we were using an IBM AT with an Inboard 386, the error might have stemmed from a timing problem.

## Mac Mirroring Choices

### DiskTwin 2.0

For immediate and almost transparent disk backups, DiskTwin 2.0, a hardware plus software disk-duplexing system from Golden Triangle Computers, is the only solution available on the Macintosh. Unlike mirroring applications, which write to disks sequentially, first to the primary disk and then to the backup disk, DiskTwin writes to both simultaneously. As a result, we noticed no degradation in the speed of disk operations when using DiskTwin. However, stopwatch in hand, we determined that most disk writes actually took about 4 percent longer with DiskTwin than without it.

The hardware part of DiskTwin is a NuBus card. (The company recently also released an SE/30 version.) The software part of DiskTwin is a Control Panel device (cdev). Once you install the card and the cdev, setting up the system is easy. If you have more than one disk drive, you can daisy-chain them. You then indicate, via the Control Panel, which disk or disks (connected to the Mac's SCSI port) are to be primary and which (connected to the DiskTwin's SCSI connector) are to be their twins. DiskTwin then "synchronizes" the paired disks, creating an exact duplicate of the primary disk on the twin. Synchronization took just over 3 minutes for an 80-MB drive. From that point on, every write to the primary disk is automatically also written to its twin.

If the primary drive fails, you simply disconnect the twin from the NuBus card and reconnect it to the Mac's SCSI port (in essence turning the twin into a primary drive), and you're back in business. The entire process takes less than 2 minutes. DiskTwin's biggest drawback is that the twin has to be at least as large as the primary drive. If the twin is larger, its excess capacity cannot be used. (The company has recently provided a partial solution to this problem by bundling another cdev, PartitionTwin, that lets you choose partitions to replicate instead of complete disks.)

When using DiskTwin, your disk operations will only be as fast as your slower drive—primary or twin. When using a 40-MB 12-millisecond Quantum drive as primary and a 46-MB Seagate rated at 32 ms as the twin, performance decreased about 7 percent.

While disk duplexing ensures continuous protection against mechanical disk problems, it provides no help

when files are accidentally trashed or corrupted, say through a virus, since the problem would occur on both the primary and twin disks simultaneously. (DiskTwin has a standby mode, which creates a copy of the primary disk on command, thus allowing you to archive whenever you want. But you can archive that function just as easily using the Finder.)

DiskTwin was recently upgraded, and version 2.0, which we received mid-evaluation, adds an important feature: Automatic Cutover. This is supposed to automatically and instantly activate the twin when the primary disk fails. The screen then displays a flashing Apple icon to indicate the failure. (On a network, the icon flashes on the

server, where it might not be noticed. The company just released a utility that will send messages over Microsoft Mail.)

In our tests, the Automatic Cutover feature worked only about 65 percent of the time; at other times, the Mac bombed (system error) or froze. Admittedly, the way we were forced to test Automatic Cutover—by powering down the primary drive—is much more likely to cause a Mac to bomb than is a head crash, which is the most common way a disk fails.

In spite of its minor shortcomings, DiskTwin is an efficient way to ensure continuous backup of your data. Once installed, it's barely noticeable. And if the backup drive is needed, getting it on-line takes virtually no time. ■

# Security and Integrity in Micro-Mainframe Networks

## In this report:

The Survey.....	2
Results.....	2
Implementation.....	5
Conclusion.....	8

## This report will help you to:

- Evaluate the risks to system security and data integrity in micro-mainframe networks.
- Determine the cost/benefit trade-offs involved in implementing various computer and data security measures.
- Implement recommended security measures to ensure data integrity in specific types of micro-mainframe networks.

## Abstract

This report examines the impact on computer security and data integrity of linking personal computers in user departments with the corporate mainframe computer. It describes the results of a survey of experts in computer security and integrity. In this survey, participants identified those security and integrity controls that become critical because of the micro-mainframe link. The risks associated with three ways of implementing the link are evaluated and procedures for controlling these risks are suggested.

This Datapro report is a reprint of "Implementing Security and Integrity in Micro-Mainframe Networks" by J.L. Boockholdt, pp. 135-144, from *MIS Quarterly*, Volume 13, Number 2, June 1989. Copyright ©1989 by The Society for Information Management and the Management Information Systems Research Center at the University of Minnesota. Reprinted with permission.

## Introduction

As MIS users have increasingly incorporated personal computers into their day-to-day activities, they have expanded their expectations of these systems. Many now recognize the value of linking personal computers with a mainframe computer in order to access the organization's database. This creates a micro-mainframe network.<sup>1</sup>

As user requests for micro-mainframe networks increase, MIS managers must address the issues raised concerning data security and integrity. Some organizations have begun to implement appropriate procedures over their micro-mainframe networks that have been described in professional literature.<sup>2-11</sup> However, no systematic study has examined the effects of this new technology on computer system security and data integrity.

**Table 1. Network Environments Used for Evaluation**

<b>Case A:</b>	A host mainframe is linked to intelligent terminals using standard teleprocessing software such as CICS.
<b>Case B:</b>	A host mainframe is linked to personal computers using a software link. The link provides online access, with access privileges identical to the terminal that the PC replaced.
<b>Case C:</b>	A host mainframe is linked to personal computers using a software link that provides <i>download</i> capability only. That is, files may be transferred from the mainframe to a PC, but may not be transferred from the PC to the mainframe.
<b>Case D:</b>	A host mainframe is linked to personal computers using a software link that provides both <i>download</i> and <i>upload</i> capability. Files may be transferred from the mainframe to a PC, changed by the PC, and transferred back to the mainframe.

This report describes a survey undertaken to address these concerns. It attempts to answer three related questions. First, what methods of achieving data security and integrity are critical when a personal computer replaces a terminal in a network? Second, what forms of micro-mainframe network create the greatest threats to data security and integrity? Finally, what procedures are available for implementing the critical methods?

### The Survey

In this study, 85 Certified Information Systems Auditors were surveyed about their views on security and integrity in micro-mainframe networks. The Certified Information Systems Auditor (CISA) designation is awarded after an individual attains a minimum amount of related work experience and passes the CISA examination. This exam was established by the EDP Auditors Foundation to measure knowledge in auditing computer-based information systems. Approximately 58 percent of the questions on the examination concern computer controls, data integrity, and computer security.<sup>12</sup> Thus, all participants in the survey had demonstrated an expertise in these subjects.

The questionnaire was developed from the security and integrity guidelines described in the most recent edition of *Control Objectives*, published by the EDP Auditors Foundation.<sup>13</sup> Thirty-one of these consisted of those guidelines that, in my opinion, are likely to be affected by the link's existence. One additional guideline, which should not be affected by the link, was included as a validator. These control objectives and guidelines are listed in the Appendix.

Respondents were asked to evaluate the criticality of each of the 32 guidelines in four different cases. Case A consisted of a typical teleprocessing

environment, with a host mainframe computer connected to multiple local and remote terminals. It was used as a reference for comparison to the other cases involving micro-mainframe links. In Case B, the respondents were asked to assume that each terminal in Case A was replaced by a personal computer (PC) with access privileges identical to the terminal's. In Case C, each terminal was replaced by a PC that provided its user with file download, but not upload, capability. In Case D, PCs replaced terminals, and the links provided both download and upload capability. Table 1 summarizes these four cases.

Respondents evaluated the criticality of 32 guidelines in each of the four cases. This produced a relatively long, although simple, questionnaire consisting of 128 evaluations. Statistical techniques were used to analyze these responses and identify those guidelines that respondents considered critical in micro-mainframe networks.\*

### Results

Table 2 summarizes the results of the statistical analysis to determine the critical security and integrity features in a micro-mainframe network.\*\*

Of the 32 features, or guidelines, in the survey, 20 had statistically significant differences in criticality. In other words, these features were more critical in a micro-mainframe-link environment than in a teleprocessing environment using only

\*Eighty-five useful responses were obtained from a randomly chosen sample of 414 CISAs. These were analyzed using multivariate statistical methods and the F statistic. Significance was determined at the  $\alpha = .05$  level.

\*\*The *Control Objectives* study designated control objectives by letter of the alphabet and designated control guidelines by number. Thus, for example, control Q08 was the eighth control guideline under objective Q, Access and Physical Security. These letter and number designations are used in Table 2, Table 3, and the Appendix.

**Table 2. Critical Security and Integrity Guidelines in a Micro-Mainframe Network**

Guideline Number	Guideline Description
<b>Control Objective O: Effectively Manage Resources</b>	
03	Periodic equipment maintenance
05	Inventory logs for all computer files
06	Standardized file identification
<b>Control Objective Q: Access and Physical Security</b>	
08	Protect against unauthorized access
09	Assign responsibility for access and security
11	Control access to terminals or PCs
12	Record and review terminal or PC access data
14	Security control based on classification of data
15	Do not publicly identify IS facilities
16	Protect files from destruction and unauthorized use
18	Train mainframe operators in security procedures
19	Periodically test security plans
<b>Control Objective R: Backup and Recovery Plans</b>	
25	Backup CPU and other hardware
26	Backup procedures to minimize recovery requirements
27	Periodic test of the backup plan
<b>Control Objective W: Control in Database Systems</b>	
28	Consideration of integrity and security of shared data
29	Written procedures for data dictionary changes
30	Access to sensitive data, concurrent access controls
31	Sufficient written data recovery procedures
32	Adequate procedures for database integrity

terminals. The Appendix provides a more thorough description of these guidelines.

### Critical Guidelines

Variable number 1, which was Control Guideline N in the EDPAF study,<sup>13</sup> states that adequate resources should be planned and managed. There is no reason *a priori* to expect resource planning and management to be more critical in a micro-mainframe network than in a teleprocessing one. This guideline was included on the questionnaire as a validator; if a statistically significant difference between environments was found on it, one might question the validity of all the results of the survey. Fortunately, the survey results indicated that no such difference exists. The rest of this section discusses the critical guidelines that have been categorized into four objectives (see Table 2): effective management of resources, access and physical security, backup and recovery plans, and control of PCs in database systems.

The study identified three critical guidelines associated with the *management of resources*. CISOs felt that periodic equipment maintenance, inventory logs for computer files, and standardized file identification methods were each more critical

in a micro-mainframe network than in a teleprocessing one. Equipment maintenance becomes important because PCs, with memory and disk drives, are more complex than terminals. In a micro-mainframe network, files are stored at user locations as well as on the mainframe. Thus, methods of identifying and controlling these files become important.

Eight *access and physical security* guidelines were identified as more critical when PCs are used. Participants perhaps concluded that a PC's programmability enables its user to circumvent security methods that might be sufficient in a teleprocessing network. Therefore, more controls would be needed to prevent physical access to equipment or to limit access to data on the mainframe.

Another objective concerns *backup and recovery plans*. The study identified three associated guidelines as critical in a micro-mainframe network. Backup plans and procedures are critical because users maintain their own files at the PC site. These files may not be subject to the routine backup procedures employed on a mainframe, so a backup plan must prescribe appropriate procedures for the PC. A periodic test of backup plans can determine if users are implementing the plan.

**Table 3. Summary of Critical Guidelines and Forms of PC Access in a Micro-Mainframe Network**

Guideline Number	Description of Critical Guideline	Record/Field Access	File Download	File Download and Upload
<b>Control Objective O: Effectively Manage Resources</b>				
3	Periodic equipment maintenance	X		
5	Inventory logs for all computer files			X
6	Standardized file identification			X
<b>Control Objective Q: Access and Physical Security</b>				
8	Protect against unauthorized access			X
9	Assign responsibility for access and security			X
11	Control access to terminals or PCs		X	X
12	Record and review terminal or PC access data	X	X	X
14	Security control based on classification of data		X	X
15	Do not publicly identify IS facilities		X	X
16	Protect files from destruction and unauthorized use			X
18	Train mainframe operators in security procedures			X
19	Periodically test security plans			X
<b>Control Objective R: Backup and Recovery Plans</b>				
25	Backup CPU and other hardware	X		
26	Backup procedures to minimize recovery requirements			X
27	Periodic test of the backup plan			X
<b>Control Objective W: Control in Database Systems</b>				
28	Consideration of integrity and security of shared data			X
29	Written procedures for data dictionary changes			X
30	Access to sensitive data, concurrent access controls			X
31	Sufficient written data recovery procedures			X
32	Adequate procedures for database integrity			X

Other guidelines evaluated in the study provide integrity and security in *database supported systems*. Respondents identified all five related guidelines as critical in a micro-mainframe network. In these networks, users access mainframe data and use it for applications on the PC. Thus, procedures for data changes and the maintenance of a data dictionary should be written and communicated clearly to all users. When mainframe data are used at a PC, they can be printed or copied to a floppy disk. Controls over access to sensitive data at the mainframe level thus become more important than when the user employs a terminal. Procedures for database recovery and data integrity are more critical because a PC user may make changes to a mainframe database that are not documented by an audit trail on the mainframe.

Statistical analysis identified those teleprocessing security and integrity features that are most critical in a micro-mainframe network. However, features in certain forms of the network are likely to be more critical than in other forms. In the survey, respondents were asked to evaluate three of them. Further analysis then investigated the effect of the form of network on the criticality of those control objectives and guidelines.

#### Form of Micro-Mainframe Network

The questionnaire asked respondents to evaluate security and integrity features in networks using three different micro-mainframe links. These were identified as Cases B, C, and D. Statistical methods then compared the criticality of features in these networks to the same ones in Case A, a teleprocessing system using only terminals.\* Table 3 summarizes the results of these comparisons. It identifies the forms of micro-mainframe networks that are most vulnerable.

#### Record or Field Level Access

Case B contains PCs with access privileges identical to those of the terminals in the teleprocessing system. In most systems, this provides access at the record or field level to data stored on the mainframe. Comparing Case B with Case A shows whether network vulnerability is increased simply because of the existence at user locations of the greater memory, local storage, and printing capability of the PC.

Table 3 shows that three guidelines are more critical in a micro-mainframe system providing record or field level access. These are the variables

\*For this analysis univariate t-tests were used at a significance level of  $\alpha = .02$ .



numbered 3, 12, and 25 in the Appendix. Because PCs have more mechanical parts than terminals, regular maintenance and backup hardware are more important. The respondents also felt that recording and reviewing terminal (or PC) access records on the mainframe is more critical in this form of network. Perhaps they were suggesting that, because of the PCs' greater capability for circumventing mainframe security features, more reliance should be placed on detecting unauthorized access after it has occurred than on preventing unauthorized access.

### File Download Capability

Cases C and D describe networks providing access to entire mainframe data files. The network in Case C provides only download capability—the ability to transfer files from the mainframe to the PC. In Case D, the PC user has both download and upload capability; that is, files can be transferred from mainframe to PC, altered at the PC, and then transferred back to the mainframe.

The guidelines considered more critical in the download only case are: restricted access to terminals, security classifications for data, scheduled hours of operation for terminals connected to sensitive data, and avoiding public disclosure of the location of computer facilities. Thus, the respondents considered this form of network more vulnerable than the teleprocessing network because of the ability of PCs to download greater quantities of data.

### File Download and Upload

The most vulnerable case is Case D, in which data download and upload capabilities are provided. Table 3 shows that respondents identified all but two of the guidelines as critical when the micro-mainframe network provides both download and upload capability. Critical features include all the access and physical security guidelines and all those listed for database systems. Respondents also felt that file controls and backup procedures are important in this kind of network.

The respondents' concern with data integrity appears justified. Using this form of link, a PC user can download a data file from a mainframe, make changes to all or part of the file, and replace the mainframe file with the changed one. This procedure may circumvent any validation procedures that exist in the mainframe-based application that

processes the file. Two guidelines identified as critical—file identification (No. 6) and file inventory and logs (No. 5)—may be useful in detecting this process after it occurs.

Of the three micro-mainframe networks evaluated, respondents considered Case B (record and field level access) as the least vulnerable and Case D (data download and upload) as the most vulnerable. These results were expected. They demonstrate to MIS managers those features that are most critical in a specific form of network. The results also show the importance of knowing procedures for implementation. MIS managers adopt procedures that are cost-effective for the organization and its form of network.

---

## Implementation

This study identifies security and integrity guidelines for teleprocessing systems that are critical when PCs are used. This section of the article provides guidance concerning how to implement these guidelines. As discussed previously, they have been categorized in Tables 2 and 3 according to four objectives. They are: effective management of resources, access and physical security, backup and recovery plans, and control of PCs in database systems.

### Effective Management of Resources

The survey identified several critical practices associated with the effective management of resources. These include standardized file identification, an inventory of files, scheduled periodic maintenance, a smooth production schedule, input/output controls, reasonable chargeout rates, and safe storage for data files.

### Risks

Effective management of a network that provides both file download and file upload is especially important because of risks introduced by file upload capability. Changes can be made to the file at the PC level, producing data on the mainframe that are incorrect or subject to misinterpretation. For example, a clerk can create a data file at a PC and upload it to the mainframe without its undergoing adequate validation. Or a manager may download a file for analysis, make any changes needed for his or her purpose, and then upload the file for later use. Another employee may produce

reports from the file without understanding the assumptions made in the analysis.

### **Establish Policies**

Appropriate management policies can substantially reduce the risks of file uploading. All production data files stored on a mainframe should go through a validation program on the mainframe, whether the data originate from a mainframe application or from a PC. PC users may be allowed to store PC files on the mainframe, but these should never be used as inputs to production programs. No report should ever be produced from data derived from a non-mainframe source. If all data are validated when uploaded and then downloaded only when needed, no data used in a report are ever more than one generation removed from validated data.

An alternative policy is to mark all mainframe data with the date and time of their creation. Any report produced on a micro must show this date and time or identify any external information sources used. If this information is missing from a report, a reader is alerted to the possibility of corrupted data.

### **Access and Physical Security**

Other objectives in teleprocessing systems are to limit access to computer resources and to provide physical security against unauthorized use, damage, loss, or modification. Survey respondents identified several practices for these objectives that were critical in PC networks. They included designating a security manager, classifying security for data, limiting and logging access to PCs, and protecting computer files and facilities.

### **Data Classifications**

All PC and mainframe data files should be categorized by security classification, and different protection policies should be implemented for each classification. In many organizations, the use of two data security classifications is adequate: non-critical and critical. Others may establish a third for sensitive data.

*Non-critical* data include memos, analyses, and newsletters for which neither security nor integrity are a major concern. *Critical* data are the records that are essential to the operations of the organization. They include accounting, statistical, and inventory data. The major concern with critical data is preserving their integrity. *Sensitive* data

includes personnel salaries, private personal data, and proprietary information for which very limited access is desired. Because it is subject to theft, controls over sensitive data should provide both integrity and security.

### **Controlled Access**

In order to implement any security classification method, an organization must identify individuals who have a need to access each data set. Employees should justify their requests for upload or download capability in writing, and this request should be approved by management. Then procedures should be put in place to restrict access to only those individuals.

For example, user profiles, in the form of passwords and user ids, are stored on the mainframe. These profiles specify which parts of the mainframe database are accessible to each PC user and are maintained by the security manager.

One respondent to the survey stated that, in many organizations, application of security on an individual basis creates an undesirable amount of overhead. This has led these organizations to broaden access categories to include heterogeneous groups rather than individuals. Such a compromise, although easier to implement, increases the risks to data security and integrity.

Survey participants felt that access to terminals or PCs should be controlled, and those connected to sensitive data should have specifically scheduled hours of operation. One respondent suggested that restricting physical access to PCs is not practical in a micro-mainframe network.

These physical limits must be replaced by individual user profiles, accompanied by preprogrammed days and hours of authorized mainframe access. In this way, the PCs are functional at all times. The mainframe is accessible to individuals only in their predetermined periods and modes of operation.

### **Protect Files**

Controlled access should be accompanied by routine logging and monitoring of attempts to gain unauthorized access to files. The mainframe operating system should log any attempt by an unauthorized person to access data files, execute sensitive programs, or access system utilities that

enable the user to copy, modify, or access files. The security manager should investigate all such attempts.

File protection at the mainframe level is implemented with appropriate data security software. At the PC level, it is dependent on user training and discipline. More restrictive protection policies should be established for sensitive data. For example, sensitive PC files can be stored only on specially colored floppy disks and locked up when not in use. These files should always be encrypted before storing them on a mainframe or a PC hard disk, and preferably, a removable hard disk should be used.

### **Protect Facilities**

Survey participants indicated that the locations of information system facilities should not be identified publicly. However, most organizations have PCs located throughout the organization. When this is true, the location of those PCs linked to the mainframe should not be publicly disclosed. Furthermore, non-user employees in the same work area should not know the access privileges of any linked PC. This limits the ability of an unauthorized person to use the PC to gain unauthorized access to the mainframe. PC keyboards should be equipped with locking switches to prevent their use when unattended.

### **Backup and Recovery Plans**

The survey identified three critical guidelines associated with file backup and recovery plans. They concerned backup hardware, backup procedures, and a periodic test of the backup plan.

#### **Backup Hardware**

A backup plan should provide for backup CPU and other hardware resources. In a teleprocessing environment, this is frequently provided by an identical mainframe at a different location. In a micro-mainframe network, backup hardware is necessary for PCs as well. The relatively low cost of a PC enables an organization to maintain spares for use as replacements. To maintain software compatibility, spares should be identical to those PCs in daily use. If the technical expertise is available, spare components such as motherboards, power supplies, and hard disks can be maintained.

One respondent to the survey stated that having available "hot recovery" facilities, which allow

immediate resumption of vital processing, is the critical form of hardware backup. It is impractical to have on hand facilities to completely restore permanent facilities following their destruction.

#### **Backup Procedures**

Backup procedures that minimize recovery requirements should be established. Procedures should allow a user to easily recover files that have been lost or damaged.

At the PC level, users may lack the training or discipline to employ procedures that are routine on a mainframe. One survey participant stated that most backup and recovery plans focus on disaster at the data center and ignore the potential for disaster at the user site. At the PC level, he suggested, a greater and more costly exposure may exist. An effective recovery plan recognizes the entire data flow, from information origin through processing and data center backup and recovery.

In this regard, the ability to upload files, identified by this study as a source of risk in a network, also provides an advantage. Data files created and used at a PC can be uploaded to the mainframe by communications software. Standard file backup procedures that exist for the mainframe then protect the PC user from lost programs and data.

#### **Test the Backup Plan**

Not only should the backup and recovery plan consider the entire data flow, but it should also be tested periodically to ensure that it functions properly. PC users who have never experienced a major data loss are frequently unaware of the importance of routine backups. They become careless about making them, even when management policy requires them. A periodic test of the backup plan in user departments, perhaps conducted by internal auditors, serves as a reminder.

### **Security and Integrity in Database-Supported Systems**

The study identified five critical topics in dealing with database-supported systems. Three topics (access to sensitive data, database recovery, and database integrity) have already been discussed; the two remaining ones are data dictionary maintenance and concurrent access to data.

### Data Dictionary Maintenance

In a micro-mainframe network, procedures related to data description, data changes, and data dictionary maintenance should be established and stated in writing. When a network allows data file upload, individual PCs may create data sets for processing by a mainframe. The data dictionary designates standard data names and formats required by the application. Written procedures prevent an individual PC user from developing data sets that are incompatible with those of another user. They also prevent one user from making changes that would affect another.

### Concurrent Access

A database-supported system should provide features that address problems created by concurrent access. Such a system can produce conflicting information if one user accesses a data item or file while another is making a change to it.

Concurrent access can be prevented by certain software packages that implement the micro-mainframe link. These packages provide a "check-in check-out" facility that is similar to a record lockout in a teleprocessing system. Whenever a file has been downloaded to a PC, it is designated as checked out. It cannot be accessed by anyone else until it is checked in.

These suggestions provide ways of implementing improved security and data integrity in micro-mainframe networks. However, before attempting to implement them, MIS managers must evaluate the tradeoff between the costs of a feature and the benefits it provides at the managers own installations. If the benefits are small, or if the likelihood or a security or integrity failure are remote, some features may not be worthwhile. MIS managers must exercise judgement in evaluating them.

### Conclusion

This report describes a survey of persons knowledgeable in computer security and data integrity. The survey was undertaken to summarize their opinions concerning the impact in these areas of the use of PCs in computer networks. Statistical methods were used to identify those features that the survey population considered critical. Three different kinds of micro-mainframe networks were considered.

The study found that few differences exist whenever the PC simply replaces the terminal. In this case, however, a PC can be programmed to bypass mainframe security features. Thus, logging and reviewing attempted file accesses become critical. Because of a PC's mechanical parts, maintenance and hardware backup are also more important.

Whenever the PC is able to download entire files, three additional access and physical security controls become critical. Security classifications for data should be established and different access restrictions implemented for each classification. Locations of PCs with access to a mainframe should not be publicly disclosed, and their access should be restricted by software and user profiles.

The most vulnerable form of a micro-mainframe network allows file transfer both from the mainframe to the PC and also from the PC to the mainframe. The study identified eighteen guidelines that are significantly more critical in a network of this kind. These existed in four areas: effective management of resources, access and physical security, backup and recovery plans, and control in database-supported systems. Methods of implementing these guidelines were suggested.

As more users acquire PCs and as current PC users become more sophisticated in their use, requests from them for a micro-mainframe link will increase. MIS managers cannot afford to implement such a link without first evaluating its impact on computer security and data integrity. This study identifies the security and integrity features that are critical in micro-mainframe networks.

### Acknowledgment

The aid of the EDP Auditors Foundation and the financial support from the University of Houston are gratefully acknowledged.

### References

- <sup>1</sup>Nestor, J.H. "An Introduction to the Micro-to-Mainframe Connection," *Journal of Accounting and EDP*, Winter 1989, pp. 4-9.
- <sup>2</sup>Cook, M.G. "Unravelling PC Data-Security Confusion," *Internal Auditor*, December 1988, pp. 49-53.
- <sup>3</sup>DiPerna, M. "Auditing a PC—It's Not So Difficult," *EDP Auditor* (I), 1985, pp. 23-25.
- <sup>4</sup>Lehman, J.A. "Microcomputer Use of Mainframe Databases," *Journal of Systems Management*, January 1986, pp. 18-22.
- <sup>5</sup>Leitheiser, R.L. and Wetherbe, J.C. "Approaches to End-User

Computing: Service May Spell Success," *Journal of Information Systems Management*, Winter 1986, pp. 9-14.

<sup>6</sup>Mills, V. and Viggiano, W. "Do You Have to be a Whiz Kid to Audit Data Security?" *EDP Auditor* (II), 1986, pp. 39-45.

<sup>7</sup>Opliger, E.B. "Identifying Microcomputer Concerns," *EDP Auditor* (I), 1985, p. 43.

<sup>8</sup>Parker, R.G. "Microcomputer Security and Control," *EDP Auditor* (I), 1988, pp. 13-20.

<sup>9</sup>Schultz, N.O. "Microcomputer Control Strategy: An Empirical Study of Its Development," *EDP Auditor* (II), 1985, p. 5.

<sup>10</sup>Schultz, N.O. and Redding, R.J. "A Survey of Microcomputer Control and Support Policies," *EDP Auditor* (I), 1985, p. 5.

<sup>11</sup>Skinner, B.F. "Technology: How Is It Changing the CFO's Job?" *FE Magazine*, May 1986, p. 23.

<sup>12</sup>EDP Auditors Foundation. "Certified Information Systems Auditor Study Guide," January 1981, p. 3.

<sup>13</sup>Li, D.H. *Control Objectives, Control in a Computerized Environment: Objectives, Guidelines, and Audit Procedures*, EDP Auditors Foundation, Carol Stream, IL, 1983.

### Appendix: Guidelines Selected for CISA Evaluation

Guideline Number	EDPAF Control*	Description
1	N	Resources adequate to support the IS objectives should be planned and managed.
2	O	Computer resources in the IS function should be utilized effectively by maintaining a smooth production schedule, providing adequate input/output controls, allowing reasonable chargeout rates, and keeping data files in safe storage.
3	O3	Periodic equipment maintenance should be included in the overall schedule.
4	O8	Responsibilities for data storage should be assigned, and housekeeping procedures to protect library contents be established.
5	O9	All computer files should be inventoried and controlled by appropriate logs.
6	O10	Computer files should be uniquely identified according to installation standards.
7	P	Systematic procedures should be followed to identify, select, program, implement, maintain, and control systems software acquisition and usage.
8	Q	Access to computer resources should be controlled, and physical security should be provided to protect them against unauthorized use, damage, loss, or modification.
9	Q1	Responsibility for access control and physical security should be assigned. A security manager should be appointed where appropriate.

### Appendix: Guidelines Selected for CISA Evaluation

Guideline Number	EDPAF Control*	Description
10	Q2	Access to the mainframe computer room should be restricted to authorized personnel.
11	Q4	Access to terminals should be controlled. Access to terminals connected to sensitive data should have specifically scheduled hours of operation.
12	Q5	Terminal access data should be recorded and reports of terminal activity should be reviewed periodically.
13	Q6	Access to the library should be restricted to authorized personnel.
14	Q7	There should be access security control based on classification of file data.
15	Q8	The location of IS facilities should not be identified publicly.
16	Q10	Computer files should be protected against destruction and unauthorized use.
17	Q11	Security measures should be provided for source documents and forms to ensure privacy, confidentiality, retention, and availability for backup.
18	Q12	Mainframe computer operations personnel should be trained in the application of security controls and procedures for mainframe computer operations.
19	Q13	Security plans should be tested periodically.
20	R	Adequate plans should exist for the backup of critical computer resources and for the recovery of 15 services following unanticipated interruptions.
21	R1	There should be a documented backup plan for processing critical jobs in the event of a major hardware or software failure or destruction of mainframe facilities.
22	R2	The backup plan should contain a predetermined priority for application processing.
23	R4	The backup plan should identify critical production and operating systems and the files needed for recovery.
24	R5	The backup plan should contain instructions for re-establishing communications.

**Appendix: Guidelines Selected for  
CISA Evaluation**

Guideline Number	EDPAF Control*	Description
25	R6	The backup plan should provide for backup CPU and other hardware resources.
26	R8	Procedures for backup files to minimize recovery requirements should be established.
27	R10	The backup plan should be tested periodically to ensure its workability.
28	W	In the development of database-supported systems, the control, integrity, and security of data shared by multiple users should be considered. All components making up a database environment should be addressed.

**Appendix: Guidelines Selected for  
CISA Evaluation**

Guideline Number	EDPAF Control*	Description
29	W3	Procedures related to data description, data changes, and data dictionary maintenance should be established and stated in writing.
30	W4	Procedures related to access to sensitive data and control over concurrent access to data, should be addressed by the DBMS.
31	W5	Recovery procedures for the database should be sufficient to minimize failures and recover the database to the point of failure. Procedures should be written.
32	W6	Procedures should be adequate to ensure the integrity of data maintained within the databases.

\*As identified in Control Objectives.<sup>13</sup> ■

# Protecting Against Computer Viruses

## In this report:

The Trouble With Antivirus Software.....	3
Vaccines.....	3
Antidotes.....	4
Virus Scanners.....	5
Memory-Resident Antivirus Programs.....	6

## Datapro Summary

Computer virus paranoia hit the U.S. in 1988, and ever since, countless antivirus software programs flooded the market. The problem is, many of these so-called prevention and detection systems have their share of problems, and are rarely foolproof. Knowing which features to seek, and which to avoid, will help you select the most secure antivirus program for your system.

Ignoring the reality of computer viruses is clearly a direct road to disaster. Adopting a "that kind of thing will never happen to me" attitude is playing right into the hands of rogue programmers. It is highly possible that somewhere in the vast computing wonderland there exists a computer virus with your name on it, or the name of someone with whom you share files.

The way to prevent disaster is to keep abreast of the latest safe-computing practices. Exercising good disk and file management and religiously employing effective antivirus software programs will assure your ability to avoid viral crises. Knowing how to identify and eradicate newly discovered viral infections is also a definite asset, although viral experts are always available (for a fee, of course) to handle that kind of task.

Entrusting your systems' security to most of the countless antivirus software programs on the market today is much like asking Elmer Fudd to guard the carrot patch against Bugs Bunny. Rogue programmers have consistently proven their ability

to outwit, with relative ease, software-based security systems. Most antivirus software systems available today are subject to viral short-circuiting. While antivirus programs may make users feel more secure, most of them are not delivering any real protection.

## Evaluating Antivirus Software

By far, the single most frequent thing users do, once they finally decide to arm themselves against computer virus attacks, is purchase antivirus software programs. This act in itself plays right into the hands of the world's rogue programmers. They know full well that users "mate" to their antivirus software in much the same way programmers mate with their programming editors, accountants with their spreadsheets, and writers with their word processors. In other words, computer people tend to place the utmost trust in the software they use. Most times, that trust is warranted. In the realm of antivirus software, however, such trust is often dangerously misplaced.

Antivirus software comes in as many flavors and varieties as do computer viruses. All antivirus software systems can, however, fall into two broad categories: prevention systems and detection systems.

## Prevention Systems

*Prevention systems* attempt to stop rogue software attacks on a real-time basis. Some also try to prevent unauthorized programs and users from accessing system hardware.

This Datapro report is a reprint of Part Two, Chapter Five, "Measures to Take, Measures to Avoid," pp. 49-64, from *The Computer Virus Handbook* by Richard B. Levin. Copyright © 1990 by McGraw-Hill, Inc. Reprinted with permission.

By identifying and blocking illegal disk accesses as they occur, by stopping rogue programs from loading into memory, or by employing password protection schemes to keep unauthorized users from accessing hardware and software, prevention systems can and do enhance the level of system security.

All prevention systems are, by necessity, memory-resident programs. Like those popular hot-key programs, they terminate and stay RAM-resident after loading. Always on-line and operational, they rely on the constant monitoring of DOS interrupts to detect and intercept software-driven command requests such as "load program," "read from disk," or "write to disk." When questionable activities are encountered (for example, when a program loads and secretly requests permission from DOS to overwrite to the boot sectors), prevention systems jump into action. They intercept the DOS calls and advise users of the trapped events. Users may then be queried as to whether the intercepted actions should be permitted to proceed. Of course, in the case of disk boot sector write requests, the actions should be prevented at all costs (unless users are themselves reinitializing the protected disks).

While this approach of real-time system monitoring is, in theory, a good one, in practice it fails to provide adequate and efficient protection for end users. First and foremost, memory-resident protection systems gobble up valuable RAM space. Under DOS, application programs are limited to less than 640K for program code space, and users are always better off keeping as much code space available as is possible. In addition, TSR programs, infamous for their tendency to cause compatibility conflicts with other programs, can be even more annoying when applied as anti-rogue systems. They interrupt work in progress with warnings of attempted program loads, disk reads, and disk writes. These are normal, constant activities of everyday computer usage, and TSR antivirus programs have no way of knowing which activities are initiated by users and which by rogue programs. In addition, the mere act of remaining memory-resident and intercepting virtually all disk read and write activities is bound to cause problems with other program software.

Moreover, just as utilities like Bloc Publishing's Pop-Drop, Helix Software's Headroom, TurboPower Software's TSR Utilities, and other TSR management systems can detect and disable memory-resident programs on demand, so can computer viruses as they scrutinize their hosts before striking. When antivirus prevention systems are located and identified, computer viruses can shut them off, perform virus-related activities, and then return the prevention systems to their active states.

To defeat just such attacks, some prevention systems have gone so far as to add "heartbeat monitors" to their antivirus code. By perpetually watching computers internal clocks, prevention systems can ascertain when and for how long they have been illegally turned off. Of course, it's a simple matter for computer viruses to stop computer clocks and restart them after deactivating antivirus prevention systems. By doing so, computer viruses can easily slip by the defensive shield provided by clock monitoring.

Worst of all, anti-rogue TSR programs cannot detect the direct manipulation of disk controllers, a severe and potentially fatal flaw. This means that all well-designed viruses, once loaded, have the capability to bypass DOS (and thus bypass antivirus prevention systems) and conduct disk access directly through disk controllers. While such deadly viruses are either loading, replicating, or damaging

disk-based data, prevention systems sit idly by, unaware of the goings-on in their own backyards. This is the primary reason memory-resident antivirus prevention systems should not be relied upon as the principal weapon in a defense against computer viruses. At best, they provide a modest shield against poorly designed bombs and viruses.

### Detection Systems

While prevention systems monitor software goings-on while programs are active, *detection systems* check program code before it's run. Detection systems complement prevention systems; they allow intruders to breach systems and then rely on sophisticated examination algorithms to isolate them. Users, when advised of inspected code's potential for harm, can intelligently decide whether the checked programs should be used, evaluated further, or discarded.

When you compare prevention and detection systems, the latter are by far friendlier, more compatible, and more reliable. Detection systems load, run, and exit just like other normal application programs. Unlike prevention systems, they do not permanently retain large chunks of memory nor do they intercept or otherwise interrupt the operation of normal programs.

Detection systems can be one of two types: *antibomb detectors* and *antivirus detectors*. Antibomb detectors scan programs, looking for hidden messages and destructive program commands buried within the programs' code. Antivirus detection systems specialize in isolating viral infections immediately after they have occurred. Both strategies have their advantages and disadvantages; neither scheme is foolproof.

#### Antibomb Detectors

Most antibomb detectors are capable of scanning individual or multiple file sets and searching for destructive routines embedded in the target files' executable code (for example, file erasure commands or disk reformatting program calls). Some antibomb detectors extract text messages stored in programs, looking for overt indications of rogue activity (statements like "Arf! Arf! Gotcha!" or profanity). Most share the same, unsettling drawbacks: they flag legitimate programs as bombs, they can't locate or display encrypted text messages, and they often fail to catch truly deadly program commands.

#### Antivirus Detectors

Due to the very nature of viral activity—viruses cause detectable changes to otherwise static (nonchanging) executable files—antivirus detection systems are considerably more effective at detecting viral activity than are their prevention-based and antibomb cousins. This level of effectiveness, the highest among antivirus software types can be further enhanced by the quality of the detection algorithm put into play.

Antivirus detection programs fall into two distinct classes: *program-specific detectors* (commonly called *virus scanners*) and *generic detectors*.

**Program-Specific Detectors:** Program-specific detectors search for a limited number (presently well under 100) of known viruses. They probe their target files, looking for identifying features (*viral signatures*) of the viruses they are programmed to detect. Program files containing known viral signatures will cause virus scanners to produce messages notifying users of their discoveries.



Program-specific detection systems appear to be a good and logical notion until their conceptual flaws are revealed:

- Program-specific detectors can only recognize a limited number and fixed set of known viruses. This means that new or modified viruses can be spreading themselves across hard disks like warm peanut butter on fresh bread, slipping by any out-of-date program-specific detectors.
- Program-specific detectors require frequent, sometimes costly updates as new viruses are discovered or as old strains are updated. After all, new viral signatures must be continually added to program-specific detectors' search code. Users not employing the latest versions of program-specific detectors are dangerously out of date.
- Program-specific detectors are rendered impotent by high-tech encryption viruses, which are designed specifically to defeat program-specific detection. Encrypted viruses mask their viral signatures by either enciphering them or by mutating infectious code on a case-by-case (per-infection) basis. The end result is that no telltale signs are available for program-specific detectors to latch onto and identify.

*Generic Detectors:* Well-designed generic detectors (and they are few and far between) are the most dependable type of antivirus software. Instead of attempting to identify and keep up with every computer virus known to man, instead of trying to plug every DOS interrupt and software hole possible, generic detectors target the single weakness, the Achilles heel, that all computer viruses share: *computer viruses must change normal executable files in order to survive.*

Executable program files should never change in either size or content, unless users physically update them, perhaps with manufacturers' software upgrades. Generic computer virus detectors operate under the basic assumption that unauthorized changes occurring in otherwise static program files are, in themselves, indications of viral activity. In fact, when program files do change without users knowingly causing the alterations, computer viruses are often at work.

Properly designed generic detectors catch all changes, no matter how small or insignificant, occurring in static executable files. As with all antivirus software programs, generic detectors are, unfortunately, not without their deficiencies. They may be complicated to use and take a long time to run, their output data files usually consume sizable amounts of valuable disk space, they sometimes generate false alarms, and, if they are stored on their host systems, they are themselves subject to viral illusions, viral infections, and viral alterations. Moreover, some generic detectors require time-consuming maintenance of their output data files, while others employ poorly conceived or outright harebrained detection schemes capable of being outwitted by sophisticated viral algorithms.

The ideal anti-rogue software safety net consists of an intelligent, well-tested, and well-balanced combination of safe-computing methods plus both rogue prevention and rogue detection software. Safe-computing methods work to protect most users from harm; prevention software erects barriers that stop common rogue programs from accessing user data, and detection software identifies advanced rogue programs after they've slipped through prevention shields. By adopting this three-pronged approach to computer defenses and by maintaining regular data

backups, users can be reasonably certain that their systems are sufficiently protected from the ravages of rogue software programs.

## The Trouble With Antivirus Software

Countless utilities for combating computer viruses emerged on the market when the first wave of viral paranoia struck in early 1988. Hot on the coattails of each new wave of virus hysteria are dozens of "new and improved" antivirus software measures. Of the antivirus software examined during my ongoing testing, all were overhyped, somewhat ineffective offerings or examples of engineering overkill—lots of features, no real benefits.

Practically all the antivirus software programs on the market today are promoted using "the fear factor"—the exploitation of undereducated users. Trusting, unknowing users are duped by vendor promises of total system security through a variety of means: disk write protection, viral signature scanning, metamorphosis monitoring, and so on. As previously stated, less-than-perfect antivirus schemes provide users with nothing more than a dangerously false sense of security.

Because it is of the utmost importance that users understand the failings inherent in most antivirus programming, some observations on the state of antivirus software measures follow. Some antivirus software vendors will cry "foul" when the deficiencies in their favored antivirus program types are pointed out. The real injustice, however, has been the ongoing misrepresentation of antivirus software effectiveness by many of those same antivirus software vendors. Starting today, let's place honesty and integrity first, hyperbole and profitability second.

## Vaccines

So-called "software vaccines" were among the first antivirus products to surface when the specter of computer viruses appeared. Initial end user response to these products was positive, but as users learned more about the underlying technology of these products (and the faults in that technology) sales fell off. Few such products remain on the market today, their developers out of business, their users out in the cold.

Developers of software vaccines will tell you that their programs provide software "antigens" that, when "injected" into executable files, "inoculate" them from viral infections. These implicit comparisons between antivirus programs and true biological vaccines are misleading—mere marketing rhetoric to exploit uninformed users. What software vaccines actually do is append small programs and checksum data to certain executable files. The targeted executables are then modified so that, when run, control is passed first to the appended antivirus programs. The antivirus programs compare the executable files' current checksums to their appended checksum data. When comparisons match, control is returned to the executable files, which continue their normal operations. When comparisons fail, users are alerted and appropriate actions can be taken.

It's important to be aware of the shortcomings and complications associated with software vaccines before putting them into service:

- "Vaccinated" programs take longer to load because appended antivirus code and data increase file sizes (the amount of data that must be loaded and positioned in computer memory) and because the prerun checksum-comparison process takes time.

- Large amounts of disk space can be consumed as appended antivirus code and data enlarge program files.
- Most vaccines can be used with either .COM or .EXE files. Device drivers, executable data, and overlay files cannot be protected.
- Some vaccines, which try to modify .EXE file headers, cannot protect .COM files because .COM files do not have .EXE file headers.
- Many vaccines cannot protect anything more than system start-up files (IO.SYS, MSDOS.SYS, and COMMAND.COM). All other files, including user data, remain unprotected.
- Many vaccines cannot protect *packed* .EXE files. Packed .EXE files have been compressed during programming's LINK process to conserve disk space; they expand in memory when run.
- Vaccines may not be able to act on executable files that have been processed through real-time data decompression utilities like System Enhancement Associate's AXE and PC Magazine's PCMANAGE programs. Such programs compress executable files and decompress them in-memory when run. While not unlike packed .EXE files, the superior data compression algorithms used by real-time data decompression utilities result in much smaller .EXE files and are applied after programs have been compiled and LINKed.
- False alarms are generated when self-modifying programs (like Borland's SideKick) update internal data areas or when programs are modified as part of an installation procedure. To prevent false alarms, users are forced to remove and reinstall vaccines before and after self-modifying programs are used or before installation procedures are run.
- For programmers source code modifications and recompiled executables are often misinterpreted as viral alterations.
- There are no guarantees that modifications to executable files (like the appending of antivirus code and data) will not adversely affect their operations.
- The virus-like behavior of vaccines may cause conflicts with other viral defense systems.
- Viruses can detect the presence of vaccine program code in target executable files and simply delete or modify the vaccine-related data, or they can patch vaccinated executables to bypass their load-time checksum process.
- Nondestructive file-checking utilities (in other words, generic antivirus detectors) provide a safer, easier way of conducting prerun file checksums and CRCs.

Vaccines operate in a manner fundamentally similar to computer viruses. They attach themselves to, and run in place of, executable files, although they do not reproduce without authorization, nor do they purposely damage files. Most users are uncomfortable with the concept of inserting a virus—antigen or otherwise—into executable files, especially when safer, less drastic alternatives are readily available.

### Antidotes

Viral *antidotes* (also known as *disinfectors* or *eradicators*) appeared on the market soon after the introduction of software vaccines. Even today, when a new viral strain emerges, a new viral antidote often follows close on its

heels. Originally, some viral antidotes were integrated into software vaccine programs; now most are sold as stand-alone, dedicated systems.

Most antivirus software systems offer subclassifications within their respective genres and viral antidotes are no different. The easiest way to categorize antivirus antidote programs is to separate them into disaster antidotes and infection antidotes.

*Disaster antidotes* (sometimes referred to as *format recovery programs*) are designed to restore systems to working order after destructive events have occurred. Vendors of such systems have been known to fool the computing public by boldly demonstrating their programs' "amazing" ability to restore "virus-deleted" data from reformatted, repartitioned hard disks. The truth is they are merely restoring backup copies of critical disk information: boot sectors, command processors, file allocation tables (a central disk management area known as *the FATs*), disk directories (as well as all root directory program files), and partition data. However, most users are completely unaware that reformatting and repartitioning hard disks does not actually delete all user data, but instead, just marks allocated disk space as "unused." When disk "road maps" are replaced with accurate backups, user data appears to be miraculously resurrected—and, in a way, it has been.

*Infection antidotes*, on the other hand, seek and remove known viruses on a file-by-file basis. These programs work reasonably well within the confines of their limited usability. You see, infection antidotes can remove only a limited set of known viruses from a narrow group of known program types. In fact, most infection antidotes are dedicated to removing a single computer virus strain. Because new viruses are being introduced all the time, infection antidotes quickly become outdated and ineffective. In contrast, the recommended method of removing computer viruses by overwriting infected files with certified backup or master copies is reliable regardless of the virus' date of origin or the type of infected files.

Like software vaccines, antidotes and format recovery programs are effective to a point but are not without their drawbacks:

- Format recovery programs cannot resurrect data on systems destroyed prior to their installation. A minimum of one up-to-date backup disk, created by an antidote program, is required.
- If the antidotes' backed-up data is not current, the information it replaces will be out of date, an inaccuracy that leads to further data loss. Most times, when antidotes' backup data is obsolete the restoration process fails completely and damaged disks remain unusable.
- Format recovery programs cannot reconstruct data erased by low-level reformatting or by destructive high-level reformatting. Low-level formatting is used by most hard disk controller manufacturers and can be activated by computer viruses (and by some DOS utilities); destructive high-level formatting occurs when users enter the FORMAT command using AT&T's, Compaq's, or Burrough's DOS and perhaps other OEM (Original Equipment Manufacturer) DOS versions. (Most DOS versions do not destructively format disks when the FORMAT command is used. Instead, disk sectors are marked as "unused," with the data they contain remaining intact.)
- If new data has been written to disks after the disks have been reformatted, deleted data cannot be reliably recovered.

- Many users already own data recovery programs (like the Mace Utilities, the Norton Utilities, and PC-Tools) that are capable of restoring even the most badly damaged disks to a fairly usable condition.
- Infection antidotes can search for, find, and remove only a small quantity of known viruses. In fact, most infection antidotes are designed to remove but a single computer virus strain. Other, uncataloged viruses remain active, undetected, and untouched.
- Users of infection antidotes must constantly update those programs in order to remain reasonably current. Even then, antidote programs are almost always one step behind their viral counterparts.
- The infection of normal program files with viral data is, in itself, severely corrupting because important program data is overwritten (replaced) with viral code. It is therefore likely that infection antidote programs, which in effect do the same thing, while attempting to remove viral code, will cause further, irreparable damage. Worse still, some disinfection programs mistakenly identify uninfected files as infected, thereby overwriting valid executable data and destroying the program file they are trying to fix.
- Every executable program has its own special characteristics, its own internal file formatting. It is physically impossible for any infection antidote to remove all known viruses from all known program types—no ifs, ands, or buts about it.

Once again, it is safer and far more reliable to recover damaged disks or eradicate infected files by restoring them from certified backup copies. There are no substitutes for regularly scheduled, conscientiously inventoried backups.

### File Comparison Utilities

Virtually every copy of DOS arrives with at least one file comparison utility, using program names like COMP, DISKCOMP, and FC. *File comparison programs*, be they provided with DOS or purchased independently, do one thing and do it well: they compare, letter for letter, byte for byte, two distinct copies of target file specifications.

Because good file comparison utilities detect even subtle changes between files, and because viruses must change files in order to infect them, antivirus software vendors have jumped on the file comparison bandwagon. It's easy to see why. File comparison utilities are simple to develop, a snap to document, and can be brought quickly to the ever-expanding antivirus software market.

File comparison utilities compare files in use against known good copies, thereby identifying viral infections. There are several problems, unfortunately, with this plain and simple technique:

- The same level of protection is achieved regardless of what file comparison utilities are used, including those provided with DOS at no extra charge.
- A duplicate copy of every compared file must be stored on another disk or directory, which is a waste of valuable disk space.
- File comparison utilities, even when designed specifically for the task of virus detection, generally do not support features essential for managing virus detection. Options like activity logs, alarms, data encryption, off-line storage, system locks, and wildcard (\* and ?) input file specifications are often lacking.

- Viruses can audit disk directories looking for file duplicates and can infect both copies of target file specifications. In such cases, future file comparison checks would not detect differences between the two infected files. This is not as much of a problem, however, when the duplicated files are stored on different disks.
- Most file comparison utilities designed specifically for virus detection prevent users from comparing anything more than system start-up files (IO.SYS, MSDOS.SYS, and COMMAND.COM). All other files, including user data, remain unprotected.

### Virus Scanners

Recently, a wave of programs that scan disks and files for known viral signatures has flooded the antivirus software market. Even conservative IBM Corporation has gotten into the act, with its fairly effective, easily updated program called "The IBM Virus Scanning Program." Virus scanners, as already noted, search only for the specific viral signatures they've been programmed to detect. The shortcomings of conducting searches for fixed sets of program types should be obvious to everyone. Users, however, appear to be undeterred, as virus scanning products continue to do well in the marketplace. Perhaps this is because searching for something we know—checking to see if there are any known viruses on your disks—makes sense to the average user. But in the netherworld of computer viruses, it's what we *don't* know that will hurt us.

- Virus scanners can detect only a limited number and fixed set of known computer viruses. This means that new or modified viruses can be active, spreading, and completely undetectable by out-of-date virus scanners.
- Virus scanners require frequent, sometimes costly updates as new viruses are discovered or as old strains are updated. New viral signatures must be forever added to the virus scanners' search source code. Users not using the latest virus scanner revisions are perilously out of date.
- Virus scanners are easily outwitted by modern computer viruses that employ data encryption techniques to mask their viral signatures. Such viruses encode their telltale signs by either enciphering or mutating them. Because their viral signatures change with every new infection, nothing tangible remains for virus scanners to seek out and identify.

### Disk Mappers

Disk mappers maintain centralized data files made up of coded disk images (*disk maps*). These coded disk images contain snapshots of their target disk status at any given point. With every run, disk mappers notify users of changes discovered between target disks and their coded disk images. Yet again the antivirus solution fails to provide an adequate defense against computer virus infections:

- Disk mappers' output files can occupy huge amounts of disk space; they increase in direct proportion to the number of files and the size of disks being tracked.
- Time-consuming maintenance of disk-mapper data files is generally required. As the condition of target disks change, entries must be sorted, updated, deleted, or purged.

- Disk mappers can be complex to operate because they must support many data-file maintenance options (like the sorting of the data base, the purging of obsolete entries, and so on).
- Disk mappers customarily update data files at system start-ups, increasing boot times.
- Viruses can *affect* nonexecutable files; they cannot *infect* them. Nevertheless, some disk-mapping programs monitor all files, regardless of whether they're executable. Nonexecutable files (like .DOC and .TXT files) are always changing (they're often created and maintained by users), and disk mappers frequently raise alarms when they encounter these changes. Better disk-mapping schemes allow users to specify precisely what file names and types are to be monitored.
- Viruses can detect, modify, and delete disk-mapper programs and data files.
- Some disk-mapping programs convert files (including user data) to read-only status (meaning that the files cannot be readily updated or deleted), thereby assuring conflicts when users want to use their applications or perform general disk maintenance.

### Memory-Resident Antivirus Programs

Several antivirus schemes rely on memory-resident modules to intercept DOS commands to provide last-minute checks of programs about to run. That presents another set of problems.

Antivirus programs employing TSR technology typically monitor disk writes directed to specific files or provide last-minute checks of files about to run. Some also provide a degree of password protection, while still others monitor usage as well as the installation of new software and data files. The complications presented by the use of antivirus TSRs are, regrettably, numerous and severe:

- Many computer configurations respond poorly to particular groupings of TSR programs—they crash, lock up, or simply behave abnormally. Let's face it, TSR technology attempts to perform something that MS-DOS was not designed to do—multitasking (running more than one program at a time). The methods used by TSR developers to coerce MS-DOS to multitask remain unstandardized and notoriously troublesome.
- Antivirus TSRs often consume considerable portions of limited available RAM space. This means there is less memory space available for normal programs and data.

- False alarms occur frequently, triggered by normal disk activity that is misinterpreted as viral activity.
- Unattended computer operations, such as e-mail transfers or file uploads and downloads, are subject to unanticipated—and usually fatal—interruptions when antivirus TSRs accidentally activate.
- Most antivirus TSRs allow only a limited number of files to be monitored, leaving other files, user data included, totally unprotected.
- System performance decreases as each BIOS- or DOS-driven task is intercepted for examination by antivirus TSRs. Computer operations can, in some cases, be slowed to a snail's pace, depending on the number and type of active antivirus TSR programs.
- Viruses can directly manipulate disk-controller hardware to bypass interception by antivirus TSRs.
- Viruses can easily detect antivirus TSRs. This is not surprising, because TSRs are always evident in RAM. Viruses can disable or remove TSRs, and for this reason alone, antivirus TSRs provide users with a false sense of security.

In summary, while the criteria for the selection of a word processor, a spreadsheet program, or an accounting application should be established in terms of the number of features available and their relative ease of use, there are a different set of priorities in play for the selection of antivirus software. All the nifty features in the world mean nothing if they do not add up to providing absolute, certain security.

It's of the utmost importance to remember that viruses have complete control of PC system resources at the moment of infection. Antivirus programs—even when renamed and stored in hidden subdirectories or on write-protected hard disks; even when they are hidden or read-only programs or embedded as vaccinated files; even if they are ruggedly reliable file comparison utilities, disk maps, TSR programs, or device drivers—all share one fatal flaw: they are subject to the scrutiny of computer viruses as they examine their hosts. Antivirus systems stored on nonremovable media, relying on support files stored on nonremovable media or residing in memory, are themselves subject to infection! ■

# LAN Security

## In this report:

Risk Analysis .....	3
Types of Threats .....	4
Protection Methods .....	5
OSI Security Considerations .....	6

## Datapro Summary

This report explores the risks and threats to network security and examines the methodologies for guarding against network intrusion. Protection should be evaluated for facilities (physical premises); the network itself including all hardware, software, and communications facilities; and the data. Some security threats are hackers, bandits, Trojan Horses, and viruses. Hackers are generally outsiders whose most common point of entry is a dial-up line. Bandits are typically insiders who gain unauthorized access to information. Trojan Horses, often put into applications during development, are triggered like time bombs to go off after a certain time period or when certain conditions have been met. Viruses enter a network in a variety of ways and, once in, spread throughout the applications, software, and network data, corrupting or destroying files. An analysis of security risk must also consider the consequences of security breaches. Compromised network security can result in lost business, damaged reputation, lost assets, lost trade secrets, and fiduciary losses—money, business, and confidence. Network risk analysis also considers the objects to be protected (such as data, computer processing time, equipment), the potential sources of security breaches, the likelihood of an intrusion, and the potential cost of an intrusion. Several methods of network security protection are available for the facility, the network, and the data.

## Introduction

Dateline—"November 20, 1991, Washington, DC—US Defense Department Computers Infiltrated—Foreign computer hackers successfully gained access to [DOD] computer systems at more than 30 sites. . ." (Prodigy Interactive Personal Service).

The fact is that illegitimately gaining access to the most "secure" systems either in our government or corporate offices is not nearly as difficult as it would seem. Most of the systems that we label "secure" are far from it. They are sensitive in the context of the information to which they are entrusted. But secure? Rarely! In fact, there is no such thing as a 100% secure system.

—By *Michael L. Rothberg*  
*President*  
*Applied Network Solutions, Inc.*

Given enough time and resources, someone will be able to compromise any system or network.

The question then, is "What can we do to make our most 'trusted' data resources secure?" The answer is far from simple and may include some surprise solutions that involve no more than common sense. In this report, we explore the very nature of network security as well as risk and threat analysis, and proven methodologies for different risk scenarios.

## Network Security Objectives

### What to Protect?

There are three general "frontiers" in the realm of network security. As the intruder becomes more resourceful (and intruders are always resourceful), our unwelcome guest breaks down each of the barriers until

he or she has arrived at the heart of our operating environment—the precious data itself!

The three frontiers are the facilities, the network itself, and the data resources.

- **Facilities**—the physical premises. These are usually protected with elaborate access control systems involving guards with guns, barrier technology, and sometimes elaborate perimeter defense systems. Unfortunately, all but the most inexperienced intruder will find these systems easy to compromise.
- **Network**—the hardware, software, and communications resources. Access to the network is usually restricted through use of passwords or tokens such as a “PIN” (personal identification number—used in ATM environments). While the most common technique, we will see that there are as many shortcomings to this approach as there are benefits, and password-based systems are relatively easy to compromise.
- **Data**—the information resources. The data is very often the most poorly protected resource because of the common misconception that a good “perimeter defense” is all that is necessary. We will see that protection of the data can be costly, but this cost must be evaluated against the cost of compromising the integrity of the data.

Another area that is frequently overlooked, at least in the context of “network security,” is the protection of equipment, personnel, and the theft of services (in a service product environment.)

In our analysis of network security, we will analyze each of these assets and attempt to formulate paradigmatic solutions.

### Profile of an Intruder

Intruders are generally perceived as sinister, masked characters attempting to break into our network for a variety of dark motives. The fact is that anyone might be considered an intruder in the right circumstances. Some examples of intruders follow:

1. **External intruders (hackers and other criminals)**—This group is composed of people who reside outside of the organization. They may have “inside connections” which makes their mission easier, but there are successful hackers who manage to scale all of the barriers in some innovative and resourceful ways.
2. **Servicers**—Servicers are one of the most dangerous groups of potential intruders. Servicers, due to their role in troubleshooting problems, often have access to resources and techniques that normally would not be permitted in a conventional operating environment. The servicer will likely be given license to bypass the normal configuration management and control procedures in the interest of getting the system or network “up” as quickly as possible. We may have to reconcile ourselves to our vulnerability in this area, but we will see there are numerous ways to reduce the risk, without reducing the servicers efficiency.
3. **Providers**—Providers are the systems development and programming personnel, including contractors, who have access to system resources that normally would not be available to users. The provider who operates in concert with a user opens up a wide range of possibilities for fraudulent use of the network for criminal ends.

4. **Consumers (or Unauthorized users)**—The “user gone astray” when in conspiracy with either the provider or the external intruder presents new areas of risk. The users would normally be the first to report unusual conditions in their operating environment. However, if a user is a conspirator, these incidents may go unreported until the next scheduled visit of the auditors. By that time, our conspirators have packed themselves up with their families and are living a luxurious life somewhere south of the border.

While we can develop specific security mechanisms to deal with each of these types of intruders, there is one invisible intruder that we cannot see or “put our finger on.” Human complacency! Humans tend to be lulled into a false sense of security, and unless rudely awakened by an incident or event, they commit serious breaches of security without ever realizing it.

An embarrassing example of this is the individual who suffers a virus—twice! After the first “hit,” precautions are taken in the form of virus-checking software. As the user becomes comfortable with the new found protection, the requirement to update the software with newly discovered “strains” is ignored. Then one day. . . .

### Impact of Security Breaches

It is imperative to recognize that breaches of security can impact our organization in some unusual and not so obvious ways. Aside from the lost assets or trade secrets, there are several other negative effects we might suffer.

1. **Lost business**—Access to confidential data can give competitors an edge in the form of marketing or sales leads. For instance, it would make sense for a “stock broker” to call prospects that already had invested positions in the markets. These potential customers are already savvy, and do not need a “hard sell.” Your stolen client list might be used to your competitor’s advantage.
2. **Damaged reputation**—If the breach of security is severe, or involves highly sensitive information, the reputation and image of the organization to whom the information had been entrusted would certainly tarnish. This is true whether the breach involves a financial services organization or a military intelligence agency.
3. **Fiduciary losses**—If your organization has fiduciary responsibility for assets, you are responsible not only for the value of the assets, but the confidentiality of the relationship as well. This type of loss can result in extraordinary losses in money, lost business, and damaged reputation.
4. **Lost assets**—Lost assets can take several forms. The obvious are money and equipment. Others include loss of services, loss of a salable information commodity, and loss of capacity in a public network.
5. **Lost trade secrets**—Compromised trade secrets can have a dramatic effect upon an organization’s profitability. In a pharmaceutical research environment for example, the loss might also include years of research resources, e.g., people, materials, capital resources, as well as future sales.

## Risk Analysis

Risk is a function of the object to be protected and the scenario within which it can be stolen or compromised. It is necessary to examine both issues but important to identify them separately. For example, a disk can be destroyed by a virus. The virus is the risk. The data is the object to be protected.

Risk analysis in a network environment involves four basic tasks:

1. Determining the "objects" to be protected;
2. Identifying sources of risk—the classes of intruders discussed earlier;
3. Estimating likelihood of risk;
4. Evaluating cost of an intrusion event.

### Objects to be Protected

The "objects" to be protected will vary with the nature of the organization. An intelligence agency may wish to protect a number of objects including personnel, information, as well as "sources and methods." On the other hand, a bank may be more interested in protecting physical resources (personnel, money) and information, but is less concerned with "sources and methods" of acquiring data.

The following list gives examples of objects to be protected:

- Information
- Equipment
- Personnel
- Personal possessions
- Consumables
- Money
- CPU processing time
- Network services (E-mail)

### Identifying Sources of Risk

Each of the classes of intruders (external intruder, servicer, provider, and the consumer) must be examined in the context of intentional versus accidental security breaches.

One approach to this analysis is to brainstorm with a group of interested parties who share compatible profiles. All ideas should be collected, and the overwhelming urge to reject the "ridiculous" should be suppressed until you have had a chance to review the ideas thoroughly. After all ideas have been collected, duplicates and dependent risks can be resolved.

If this exercise in risk analysis is to be successful, it is absolutely essential to "think like a criminal."

### Estimating Likelihood

Although this is not necessarily a "scientific" activity, we can make some reasonable estimates of the likelihood of an intrusion event occurring. We make these kinds of decisions every day, for instance, when we purchase insurance. How do we decide upon the maximum benefit? We examine the replacement value or the "reasonable" liability. If you were considering earthquake insurance, you would be more likely to purchase it in California than in New Jersey. In making these kinds of decisions, we typically evaluate available historical data. In the absence of clear statistics, we might use other people's experience as a benchmark.

One of the major problems in estimating the likelihood of a security breach is that it tends to be a moving target. A few short years ago, no one ever heard of computer viruses, but today they are spreading like computer "AIDS." This highlights the need for a flexible approach and a mechanism to re-evaluate the risks periodically.

### Evaluating Cost of an Intrusion

Getting a firm handle on costs varies in difficulty with the nature of the "object" that has been stolen or damaged. There are three basic cost elements to consider:

1. Primary cost—the replacement cost of an item. This will include those resources necessary to regenerate and re-enter data into the network databases as well as the cost of replacing any resources that must be repurchased.
2. Secondary cost—lost business. This item is less tangible in that lost business can only be quantified in terms of revenue projections before the incident. One way to evaluate this might be in the context of *business revenue* for the same period in the preceding year, adjusted by increases or decreases of actual business from one year to the next.
3. Proportional cost for statistical values—This type analysis is important in those environments where the loss of a network resource has a detrimental effect on your customer's business. Consider the scenario where you are a value-added-network provider, and you lose an E-mail database. Estimate the number of customers who will attempt to recover damages and the range of settlement costs. Thus if you ascertained that:
  - 1% of customers would seek damages,
  - \$50,000 average settlement,

Then: Cost = 1% of \$50,000 or \$500 per customer.

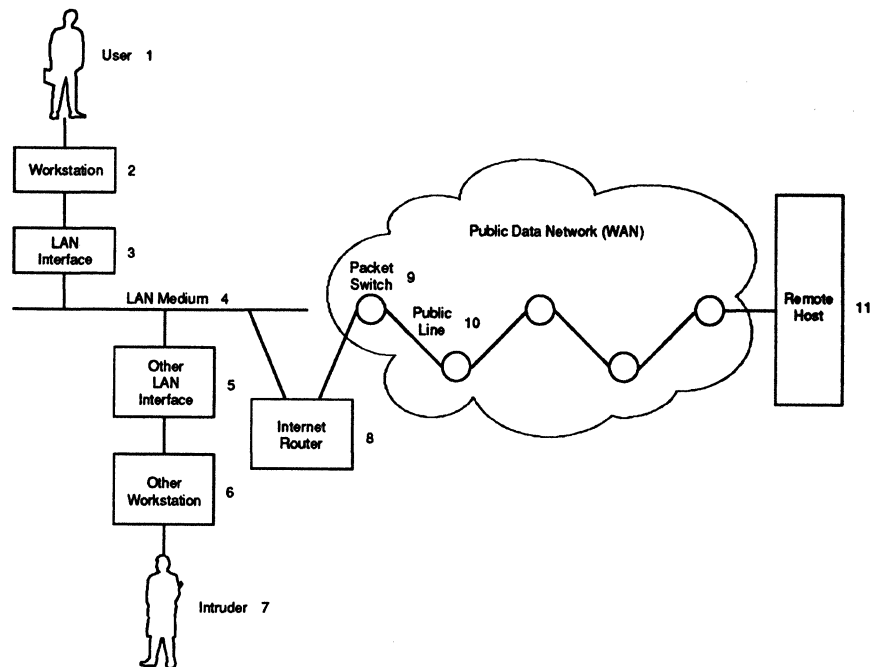
Proportional costs will also be a factor in estimating the cost of a "violated" fiduciary relationship.

### Network Vulnerability

Protecting a mainframe computer presents a set of problems different from protecting a network, particularly a network composed of both LAN and WAN domains. Mainframes operating in a standalone environment (admittedly rare these days) can be secured in vault environments. Networks, on the other hand, by their very nature, are accessible from the user's desk, wiring closets, floors and ceilings, and in the case of WANs, in the public domain.

Consider a LAN connected to a Public Data Network (PDN). Figure 1 illustrates a LAN/WAN connection scenario where there are no less than 11 points at which the security of the network can be conveniently compromised. It does not even consider the fact that each and every line and node within the WAN is subject to intrusion. Fortunately, there are tools and methodologies we can employ to minimize the risk at each of these points, but we should not operate under false illusions that any one of these techniques is foolproof.

Figure 1.  
Points of Vulnerability in  
LAN/WAN Interconnection



## Types of Threats

Threats to network resources come in several forms. The common denominator is that they all cause tremendous damage and render the network inoperable if not contained. Network security specialists often classify these threats as:

- Hackers,
- Bandits,
- Trojan horses, and
- Viruses.

### Hackers

Hackers are often young hobbyists who are motivated by the "thrill" of breaching a secure system. (This is not an "apology," and they should be prosecuted to the full extent of the law.) The most common entry point to a network or a system for a hacker is a dial-up circuit. This does not imply that a network without dial-up entry is immune to the hacker. The hacker may gain access to the premises through a variety of subterfuges. Most often, hackers treat the breach of security as an "intellectual challenge" with no intent to do damage, but, more often than not, damage does occur. An important facet of the psychological profile of hackers is that they need recognition and must leave evidence of their break-ins. This evidence is often the cause of the damage.

### Bandits

Bandits are typically "insiders" who gain access to unauthorized information. This presents one of the most dangerous security problems since mechanisms are rarely in place to provide notification that the breach has occurred. Network operating system security mechanisms, such as access rights, must be tightly controlled, and predictability in security techniques should be minimized.

### Trojan Horses

Trojan Horses are often placed in programs at development time. The code may be hidden in system or application software. Typically, these take the form of "time bombs" that are placed to explode when action is taken against an individual. For example, the programmer who writes the payroll system may place a Trojan Horse in the system to recognize an involuntary termination code in his or her payroll record and erase the entire database in response. This type of threat can be minimized by maintaining high professional standards and by performing periodic (and aperiodic) audits of software.

### Viruses

Viruses are the most recent and probably the most debilitating threat to network security that we have encountered. Unlike the hacker, the individual who spreads viruses does not have the overwhelming need for recognition. The "Typhoid Mary" of the computer network world is content to simply destroy as much software and data in the hands of anonymous innocent users as possible. Viruses are a form of Trojan Horse that spread to other programs and software throughout the network by quietly "infecting" them. Unlike the original Trojan Horse, the virus is contagious. It often attaches itself to executable program files and gradually eats away at code until the programs are inoperable. Some viruses actually reformat disks, destroy directories and wreak general havoc upon the system software. Unlike Trojan Horses, the virus usually does not result in an immediate total failure. Rather it gradually spreads as a malignancy through the system. Some of the signs of a virus include degradation of performance and erratic action with no apparent cause. When a user describes a problem by relating that "it worked 5 minutes ago, but the next try failed," this may indicate the presence of a virus. Viruses don't "get better." They only get worse until the system crashes in complete failure.



Techniques for controlling these phenomena are examined in the next section.

## Protection Methods

### Protecting the Facility

Earlier, we described three domains or "frontiers" for protection. The first was the physical facility. The most obvious defense for the facility is armed guards and access control systems. The extent to which the facility is secured will be a function of the objects inside that are to be protected. Be aware, though, that the elaborate defenses broadcast the fact that something of value lies within.

Another technique for protecting facilities is to shield the facility and prevent radio frequency emissions from emanating outside. This is a common technique in secure intelligence facilities, but it would hardly be cost effective in a commercial environment. In effect, it involves lining the exterior walls of the building with fine meshed shielding material (including windows). Naturally, this can only be implemented in new construction.

### Protecting the Network

The second frontier or line of defense is the network itself. Not all intrusions into the network will result in failure, but we have seen how lost data can have some very devastating secondary and proportional costs associated with it. The most common approach to protecting the network is through use of passwords and positive personal identification tools (e.g., hand geometry, fingerprint recognition, signature dynamics, speaker verification, etc.). All of these identification technologies operate by comparing an identity claim to a prestored reference using pattern recognition technologies. Some of the technologies, such as signature dynamics, actually measure the process of creating the identity claim as well as the resulting pattern. These can be powerful tools when coupled with passwords.

Passwords present some interesting problems. First is the need to maintain secrecy of the password itself. Second are the issues of convenience. Longer passwords are more secure, but they are more difficult to remember. Conversely, shorter passwords are easier to remember but easier to guess. Many people use the same password for everything to minimize the confusion (e.g., automatic teller machine, garage door opener codes, alarm codes, E-mail access, etc.). If their password is compromised, their entire lives can be turned topsy-turvy.

The following are some guidelines on choosing passwords:

- Five to seven characters in length
- Should contain one non-numeric character
- Should not be all alphabetic characters
- Should be random
- Do not change too often (hard to remember)
- Change often enough to preserve security.

Network security can also be compromised by capturing radio frequency emissions from the network media. This can be minimized by using optical fiber and shielded media as well as by installing the media in conduit and locations that are not easily accessible to an intruder.

### Protecting the Data

The third frontier is the data and software resources or any information that is transmitted and stored in the network. The most common approach to protecting the data is utilization of encryption technologies. Encryption technologies manipulate data to make it unrecognizable to an unauthorized observer. The result is an unintelligible stream of information. The intelligible information is referred to as *plaintext*, while the unintelligible information is referred to as *cyphertext*. The two primary encryption techniques are:

- Applied conventional cryptography and
- Public key cryptography.

Conventional cryptography involves the use of a secret key that is known only to the sender and the receiver. Public key cryptography involves using separate keys to cypher and decypher a message. In this case, neither the encyphering key nor the algorithm need be kept secret, thus reducing the exposure of the decyphering key and simplifying management.

The Data Encryption Standard (DES), which is not used in military and intelligence environments, is the former type—conventional cryptography. It involves a 16-stage encryption process where each stage expands and contracts the data using a 56-bit key. The process at each stage is similar, but different expansions and different elements of the key are employed.

DES has been criticized by some security professionals. If it took one microsecond per "guess," a two-key system would require 2285 years to break. However, if 1000 encryption chips were employed simultaneously in an attempt to break the encryption key, the time to break the algorithm could be dropped to 20 hours. Some implementors have overcome this criticism of DES by doubling the key length.

### Protection Against Viruses

The most important protection you can give yourself against viruses is frequent periodic backup of all server disks. It is possible that backups themselves will be corrupted, but most viruses infect only the executable programs. If you are fortunate, you can reinstall your software and reload your data files. The process may be tedious, but it beats the alternative. There are numerous backup programs on the market for PCs that are fast, efficient, and easy to use. Representative products include Norton Backup and Fastback 3.0.

Maintain several generations of backups so that if a backup becomes unusable, you will have another generation. This can be accomplished using either full backups that backup the entire system, or incremental backups that only backup files that have changed since the last backup. If a virus is introduced, an incremental backup will not damage earlier generations.

Other techniques used for virus protection are special scan programs. These programs operate in background mode and scan all files for "known" viruses before writing the files to the disk. The programs can also be incorporated in the boot sequence so that the important system files are scanned every time they are loaded.

If you suspect a virus, you can scan your entire disk, and the programs will report the viruses that have been found. Programs are also available to remove the viruses from the infected files. If the virus is caught early enough, use of

**Figure 2.**  
**OSI Reference Model Security**  
**Guidelines**

Service	OSI Layer						
	1	2	3	4	5	6	7
Peer Entity Authentication	N	N	Y	Y	N	N	Y
Access Control	N	N	Y	Y	N	N	Y
Data Confidentiality	S	N	S	S	N	N	S
Data Integrity	N	N	S	S	N	N	S
Data Origin Authentication	N	N	Y	Y	N	N	Y
Nonrepudiation	N	N	N	N	N	N	Y

N = No  
Y = Yes  
S = Sometimes

these programs will be sufficient. However, it is a good practice to reinstall any infected programs after the system has been "cleaned."

Since new strains of viruses are being discovered every day, it is essential to subscribe to an updating service. Major virus software providers such as McAfee and Norton provide bulletin board services for free updates.

One of the most frequent sources of viruses are shareware programs downloaded from bulletin boards. As a rule, one should always download software from bulletin boards to diskettes. Before transferring the files to hard disk, they should be scanned for viruses. They can then be loaded to the hard disk and executed from the appropriate directory. The same precaution applies to all diskettes that are being used for the first time. In some instances, reputable software firms have repackaged returned software, and reshipped it to new customers with virus infections. As viruses have become more common, this practice has stopped and the software firms do not reship the returned products.

Viruses are truly the "AIDS" of the computer industry. There is no panacea or 100% safe method for protection, but vigilance and careful procedures will minimize the risk.

### Security Audits

One of the most significant problems we face in network security is that as users become more complacent they also become more careless. We tend to relegate the responsibility for security to our technological implementations and never think about them again. Maintaining a secure network requires ongoing vigilance. Personnel have to be trained not to write passwords and encryption keys on their desk calendars. They must be admonished not to walk away from terminals that are logged on to secure applications.

It is essential to set up procedures that will give an audit trail of access to the network so that in the case of the server who has breached security, you can at least be alerted

to it after the fact. The network security system should maintain a database of user IDs, attempted accesses, violations of access privileges, and network monitoring data. These are all critical elements of network security, and their importance should not be underestimated.

### OSI Security Considerations

The International Standards Organization has identified several mechanisms to provide security services:

- Encryption
- Cryptographic error checks
- Source authentication
- Peer-to-peer authentication
- Access control—mandatory and discretionary.

The ISO Security Addendum to the OSI reference model provides guidance on which layers should or should not provide the different security services (see Figure 2).

### Summary

Network security is far from a trivial exercise. One of the first decisions that will be required is "Is it worth protecting?" In answering this question, you will find that protection is not justified for low-cost or low-value items. Effective security mechanisms will reduce network performance and may impose inconveniences within the work environment. Implementing security measures usually raises the perceived value of the items being protected. These disadvantages will have to be evaluated in the context of the potential costs incurred as a result of a loss. If you decide that security mechanisms are called for, then you will have to perform careful analysis of threats, risks, and protection methodologies. Alternatively, if you do not care to protect the resources using the techniques we have discussed, you can

- Watch it
- Screw it down
- Lock it up
- Hide it
- Make it unattractive,

or wait until it is stolen and then investigate and prosecute. ■

---

This report was prepared exclusively for Datapro by Michael L. Rothberg. Mr. Rothberg is president of Applied Network Solutions, Inc., a Somerset, NJ, firm specializing in designing, developing, and implementing local and wide area computer networks for government and commercial clients and provides market and product planning services for communications suppliers.

Prior to founding the company in 1981, Mr. Rothberg was a vice president of the Chase Manhattan Bank, N.A., where he pioneered the application of digitized speech and local networking technology to support banking applications. He is a frequent contributor to trade publications and is a member of the Program Advisory Board of the Interface Exhibitions and Conferences.



# Developing a LAN Virus Protection Strategy

## In this report:

Network Protection .....	2
Money Matters .....	2
Added Protection .....	4
Virus Scanners .....	4
What to Do .....	4

## Datapro Summary

Research conducted about computer viruses suggests that they will haunt the computer industry for a long time. Computer viruses have had devastating effects on networks in the past, but more recently, smart companies are implementing anti-virus network strategies. Appropriate funding for antivirus equipment and the acquisition of virus protection tools, are just two of the most effective strategies to combat computer viruses.

Viruses have crippled networks in the past, and they'll do it again. But you can reduce the risk of a virus incapacitating your network by taking preventive measures. In particular, you need to formulate an anti-virus network strategy.

The necessity for such a plan is rapidly becoming acute. Security analysts expect an enormous growth in viruses over the next few years. However, while the need for protection against viruses is widely recognized, few users have actually taken steps to safeguard their networks.

Two recent studies of network security show that only about 10% of computer sites nationwide have installed virus protection software or network security systems. These research studies were conducted by Market Intelligence, Research Corp. of Mountain View Calif., and Certus International Corp., a Cleveland-based antivirus consulting firm and software vendor.

Security analysts say it's puzzling why more users haven't protected themselves. Lack of knowledge is not an excuse. Most net managers feel threatened by viruses, but

the majority of managers *Network World* contacted for this report have no antivirus plan in place.

Some managers admit the plans they do have, such as a once-a-year check of local-area network workstations for virus infections, are inadequate. Among those managers who have antivirus strategies, the majority say they devised a plan only after their network was infected with a virus.

This lack of a coordinated response by net managers is more curious given that many users were hit by viruses last year. *Network World's* Critical Issues Survey found 32 of 100 network managers had a virus on their networks last year.

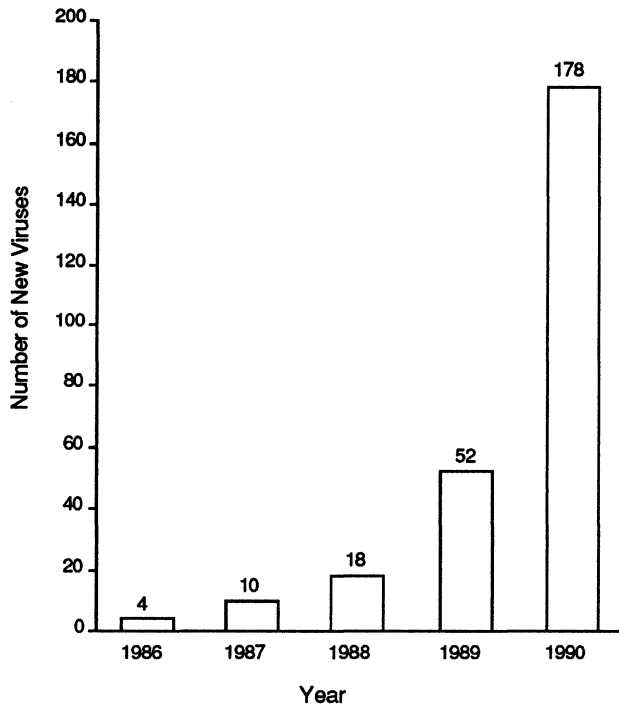
The Certus study, completed in March 1991, found that 26% of 2,400 surveyed sites with 400 or more microcomputers were infected by a virus in the first quarter of this year.

And the situation is likely to get worse. Virus makers are becoming more prolific, experts say. They produced an average of one new virus every other day last year (see Figure 1). The National Computer Security Association, based in Washington, D.C., predicts this rate will rise to six new viruses per day this year.

Virus creators are also getting more sophisticated. They've introduced stealth viruses that hide their existence by masking changes in file lengths listed in a directory or by disinfecting themselves when read into memory. This latter technique prevents many antivirus software packages

This Datapro report is a reprint of "How to Guard Nets Against the Growing Virus Plague" by Salvatore Salamone, pp. 1, 49-51, and 70, from *Network World*, July 15, 1991, Volume 8, Number 28. Copyright © 1991 by Network World/CW Communications. Reprinted with permission.

Figure 1.  
Explosive Growth Rate of New Viruses



Source: *Network World*

from detecting changes in a program that would indicate the presence of a virus.

## Network Protection

What can users do to protect their networks? Security analysts suggest drawing up a strategy based on three ideas.

- Realize the increasing value of data being stored on a LAN. Until recently, LANs didn't maintain any data of real value to a corporation. With companies downsizing, this has changed.
- Getting funding for antivirus equipment. Develop a plan to justify the expense of virus protection to upper management.
- Implement a wide variety of virus protection tools. No one product can keep all viruses off a network. Combating the problem requires a combination of products ranging from virus scanning software, which finds infected files by checking for characteristic changes in a file's structure caused by known viruses, to equipment, software or a combination of hardware and software, which runs on servers to prevent unauthorized software from running on a network.

Will following these guidelines truly safeguard networks against viruses? An examination of each area in detail reveals that virus-free networks can be a reality.

"When LANs were used for printer sharing, [security] wasn't an issue," says David Jackson, vice-president of technical services at Micro One, a Dallas-based systems integrator.

With companies now running mission-critical applications on LANs instead of mainframes or minicomputers, this perception has changed. Data on the LAN now becomes more valuable and thus needs protection. "Now that LANs are used for data processing, [antivirus security] becomes an issue," Jackson adds.

Others agree, "Because of the value of information, we've had strong security tools for mainframes and good tools for minicomputers," says Tom Patterson, technical director in the information security business until of Centel Federal Systems, Inc., a systems integrator in Reston, Va. "Information that companies use is just now moving to LANs. So LAN security becomes important."

Patterson explains that it's not necessarily more difficult to secure a LAN than mainframe- or minicomputer-based networks. "The security game hasn't changed a lot over the past 15 years," he says. "It's the same as it was with mainframes."

## LAN Intruders

Net managers agree that security measures that would keep viruses off LANs are similar to measures taken to secure mainframe and minicomputer networks, such as requiring passwords to gain access to the network and assigning users different levels of access to network resources depending on their needs.

However, one net manager, who requested anonymity, says there is a difference with LANs. "In your [traditional information systems] center, you wouldn't allow a stranger to walk in, mount a tape and load programs onto a mainframe," the manager says. "Yet everyday, people carry floppy disks into work and load software onto LAN workstations."

This manager says he often ignored this type of behavior until his LAN was hit with a virus that was introduced into the network via infected software loaded from a disk. Now he has a different view.

Security analysts echo this concern. "Sharing floppies and using bootlegged software in a [LAN environment] is about as safe as sharing needles," says Patrick Springer, a consultant with Computer Task Group, Inc., a Needham Heights, Mass., management consulting firm.

Indeed the sharing of floppy disks and the use of bootlegged software are the main sources of LAN virus infections, according to Centel Federal's Patterson.

## Money Matters

As the value of data stored on a LAN increases, many managers would like to see a corresponding increase in the amount spent on protecting it. But that has not happened.

"Many [net managers] are looking for ammunition to cost-justify [antivirus] security to upper management," Patterson says, "The only time they have good luck is when there's a publicized outbreak. Then they clip the stories and carry them to management."

Patterson and others note that for some industries, such as banking and government agencies, it's not a matter of money. These industries use the best precautions available because they are mandated to do so either by law, in case of banks, or by federal or Department of Defense regulations,

in the case of many government agencies. However, net managers in other industries are struggling with the cost-justification problem.

"When I ask for money, everyone seems to think viruses happen to other people's networks," says a net manager for a West Virginia chemical manufacturer who asked not to be identified. "Whenever there's a story on the front page of *The New York Times* [upper management] gets concerned. But their concern soon fades, and I don't get the money I need."

For many net managers, the unfortunate truth is that until their network gets hit with a virus, they won't get the support they need. "There is a very laissez-faire attitude among those who have not had a virus infection," says John McAfee, chairman of the Computer Virus Industry Association (CVIA), based in Santa Clara, Calif.

How can net managers cost-justify adding security measures to their networks to protect them from virus infections? Point out the amount of downtime such attacks create, says Micro One's Jackson. When a virus outbreak occurs, often the entire network must be brought down to identify, isolate and eradicate the virus.

"It's easy to justify the cost of security based on lost worker hours," he says, "It may be hard to estimate the value of lost data, but the downtime issue is easy to justify."

Consider the "worm" virus that struck the Internet in November 1988. The CVIA did a detailed breakdown of costs associated with the virus, which infected 7.3%, or 6,200 of the 85,200 machines on the network. Virtually the entire network had to be brought down, however, and examined for contamination.

The CVIA estimates end users lost over eight million hours of connection time to the network. And network operators and administrators spend about 1.13 million hours restoring the network to working condition.

The CVIA study conservatively estimates the total cost of lost access time and labor to restore the Internet at \$98 million.

When put in those terms, spending between a few hundred and several thousand dollars to equip a LAN with antivirus software seems like a bargain. While there is no clear formula for cost-justifying such expenditures, net managers say one network outage from a virus can cost a company as much in lost worker hours as the price of the antivirus software.

## Spending Money

Once the net managers get financial support, there is still no clear path to virus-free networks.

"It's very difficult to buy a box to solve all your [virus] security problems," Patterson says.

Since viruses can infect networks in several ways, a combination of approaches and products is required. For instance, in addition to sharing floppy disks or using unauthorized software, viruses can enter wherever users are dialing out.

"Today, almost every network has a dial-out communications server from which users can easily download files from bulletin boards," Micro One's Jackson says.

Several net managers say they've lessened the risk of virus infection from this source by setting up their communications server so files are transferred to a workstation that is not attached to the network. The files are scanned for viruses before being uploaded to the network.

This isolation approach works for dial-out, but allowing users to dial into a network represents a more difficult problem. One approach is to set up a security server—a device that requires password authentication before the caller gets onto the network.

Examples of products in this area include Burlington, Mass.-based Xylogics, Inc.'s Annex three and Cambridge, Mass.-based Security dynamics, Inc.'s ACE/Server.

Annex three is a UNIX communications server that restricts access to sensitive host computers on a network. Users dialing into the server must enter a password. Password assignments can be for specific Annex three ports, which provide connection to different network segments.

Prices for the Annex three range from \$3,995 for an 8-port system to \$6,995 for a 64-port system.

The ACE/Server controls access to networks via a gateway, remote dial-up or direct connection by using a credit card-sized device that displays a code, which automatically changes every 60 seconds.

The server maintains codes for each card issued to users. A user trying to access network resources enters the code number displayed at the time of the attempted access. If the number does not match the code on a centrally administered system, access is denied.

The ACE/Server operates in Transmission Control Protocol/Internet Protocol environments and supports UNIX servers, Sun Microsystems, Inc.'s SunOS, Digital Equipment Corp.'s Ultrix and other UNIX server platforms. The ACE/Server is priced under \$5,000.

The two security servers are examples of tools that prevent unauthorized users from gaining access. Since most viruses spread only when a program is run, blocking unauthorized programs from running on the network stops virus infections from spreading.

Net managers have a choice of products to keep such programs from running. There's a software-only approach offered by Tinton Falls, N.J.-based Brightwork Development, Inc.'s SiteLock program, and a software/hardware combination offered by Centel Federal's Net/Assure.

SiteLock designed for Novell, Inc. NetWare networks, runs as a NetWare Loadable Module on servers running NetWare 3.x and as a value-added process for NetWare 2.x. Using SiteLock, a net manager would construct a list of programs that can be transferred from a local disk to the LAN. Unauthorized programs, such as those brought in by users on floppy disks, cannot be loaded on the network.

If a user tries to run an unauthorized program from a local drive on the network, the executable file for that program is compared to a list of authorized files maintained on the server. For example, SiteLock compares the file name, byte size, owner name and creation date of a file prior to running a program.

An infected program, even an authorized one, is likely to have different characteristics than a clean copy. If the local drive version of the program does not match the characteristics of the clean file, the local program cannot be run on the network.

"Checking a file against a registered clean copy is the most reliable method of protection against any virus," says Lori Miano, a network administrator at Farmington, Conn.-based Otis Elevator, Inc.

Net/Assure takes a different approach by using a plug-in card along with software to control execution of unauthorized software. The plug-in card works in conjunction with a workstation's network interface card (NIC). Users must enter a password that is checked against valid passwords for the particular NIC address.

An advantage of Net/Assure is that laptop computers can be fitted with the plug-in card. If the laptop is connected to the network, it has the same level of security as a hard-wired workstation. That is, the user must enter a password associated with the NIC in the laptop.

### Added Protection

Even with such protections in place, net managers should be on the lookout for viruses. This doesn't require any special equipment, just awareness on the part of the net manager.

"Don't wait until someone says they're out of disk space," Micro One's Jackson says. "As viruses propagate, they consume disk space. So you should be checking disk structures regularly."

What should be checked? "Look for new directories that suddenly appear," he says. "Or for a disk that loses 20M bytes of storage capacity in one hour or 200M bytes in one day. "If any of those things happen, you may have a virus, he says.

### Virus Scanners

Virus scanning software can also help. Many products check for file size changes as well as program structure changes, such as a different cyclic redundancy check. There are many commercial virus scanning programs on the market as well as many shareware products available from several on-line sources (see Figure 2).

Virus scanning programs, which range in price from a dollar per machine for many shareware products to \$100 or less per machine for commercial products, offer a relatively inexpensive way to check for viruses. The problem with most scanning programs is that they are designed to identify existing viruses, not new strains.

Virus-scanning software vendors recognize the problem and have taken a variety of approaches. For example:

- Microcom, Inc. in Norwood, Mass., frequently issues upgrades to its VPScan product line. The company developed installer programs that make loading new versions of the software onto workstations easier and less time-consuming.
- McAfee Associates in Santa Clara, Calif., offers upgraded versions of its product line via the CVIA bulletin board.
- Central Point Software, Inc. of Beaverton, Ore., releases quarterly upgrades, and users can download characteristics of new viruses from a company bulletin board as well as from Central Point's CompuServ Forum.
- Torrence, Calif.-based Trend Micro Devices, Inc.'s product uses artificial intelligence to search for viruses, making upgrades unnecessary.

However, purchasing virus scanning software and other programs and equipment to keep viruses off networks is just one part of the antivirus strategy.

### What to Do

Industry analysts say net managers should develop a comprehensive plan to deal with the virus issue. The plan should include:

- Assessing the risks to the network from viruses.

Figure 2.  
On-Line Sources of Virus Information

#### Computer Virus Industry Association Virus Information Bulletin Board (408) 988-4004

- Downloadable shareware and public domain virus scanning software and virus fighting software.
- User forum for discussion of virus issues.
- Latest upgrades to several antivirus software programs.

#### The National Computer Security Association Bulletin Board (202) 364-1305

- Bulletin board service dedicated to microcomputer and LAN security.
- Downloadable shareware and public domain virus scanning software and virus fighting software.
- Product evaluations.
- Many tutorials on viruses and how to fight viruses.
- List of seminars and training programs on security issues.

#### National Institute of Standards and Technology Computer Security Bulletin Board (301) 948-5717

- Information about computer security and the National Computer Systems Laboratory.
- Many downloadable files on risk assessment, security and encryption.
- One section contains general virus information issued several times per week over Internet and Bitnet networks.
- List of NIST-sponsored and other government agency conferences on security.

Source: *Network World*

- Assessing the dollar damage that would be incurred if data is destroyed or networks are shut down by virus infections.
- Educating employees on the sources of virus infections and what constitutes safe computing practices.
- Deciding whether, and where, network access should be limited.
- Securing servers and workstations on LANs.
- Backing up important data.
- Protecting important program files by making them read-only.
- Keeping track of network resources, such as disk space on servers and workstations.
- Routinely scanning for viruses on servers and workstations.
- Checking all downloaded files for viruses.
- Scanning disks (particularly shared floppy disks or disks of borrowed software) for viruses before using them on a network.

Implementing an antivirus plan based on these points requires a wide variety of equipment to solve the network virus problem, funds to pay for the equipment and an appreciation by end users of the magnitude of the network virus problem.

All data about viruses suggests that they are going to be around for a long time. Users that are prepared with a proper preventive strategy will head them off and stay one step ahead of the competition. ■



# Application Layer Network Security

## In this report:

Encryption: The Keystone .....	2
Key Management Considerations .....	3
Picking Up the Keys.....	4

## Datapro Summary

A network that is not properly secured threatens the integrity of a company's most sensitive documents and critical financial data. Some early network security solutions have not proven to be highly effective and cost-efficient. Analysts are now pinning hopes on application-layer security (ISO Layer 7) as the most effective corporate network security solution. Encrypting at Layer 7 permits a secure course from the source to destination.

## Ineffective Security

Networking, once seen as a competitive advantage, is fast becoming a corporate necessity. That may be good news for vendors and suppliers, but for end-users there's a catch: As a network becomes increasingly indispensable, each decision not to use it exacts a price. And the cost of saying "No" will only get higher and more apparent with time.

But such refusals are understandable. Without adequate ways of securing corporate and public nets, users are leery about entrusting them with sensitive documents, engineering plans, and mission-critical financial data—all ripe for eavesdropping, alteration, and mishap. These concerns aren't limited to military and defense applications: Ensuring message privacy, guaranteeing message integrity, and authenticating data origin (among other examples) are now of paramount importance when it comes to business communications and transactions.

This Datapro report is a reprint of "Network Security: Just Say 'Know' at Layer 7" by Ken Rossen, pp. 103-106, from *Data Communications*, March 1991. Copyright © 1991 by McGraw-Hill, Inc. Reprinted with permission.

While we routinely estimate the cost of setting up and maintaining a network, it's difficult to pin a dollar value on opportunities missed because networks could not be used. Equally elusive are the financial risks involved when traffic is sent over an unsecured net. Thus far, however, most of the latter threats have remained more potential than realized.

Along with a newfound concern with network security comes the realization that many of the earlier approaches to the issue are neither sufficient nor cost-efficient. Deployment at ISO Layers 1-4, while providing comprehensive security for traffic within a LAN or WAN, cannot, by definition, deliver the necessary service for traffic that crosses network boundaries, such as e-mail, EDI, and file transfers between customers and suppliers. Further, within an organization's own network, it is likely that some traffic is oversecured (usually at a prohibitive cost) while some may not be fully covered.

Defense organizations have long recognized security needs and have had the mandate, resources, and control to implement them. Typically, however, they have attacked the problem at Layer 3, throwing a security cordon around transmission facilities between hosts, or at Layer 2, by protecting individual network links. Both are inappropriate to most corporate requirements (and budgets).

## Layer 7 Security

Application-layer security (ISO Layer 7) offers one solution to the corporate network-security gap. A number of recent technical, political, and commercial developments have helped to make it feasible for corporations to implement their own secured applications. Among them are:

- the CCITT's 1988 passage of X.509, a standard for the distribution of public keys that encrypt and decrypt data;
- readily available directory server implementations and installations;
- software tool kits for public-key security features;
- mechanisms that generate public keys for users; and
- the emergence of secure applications, such as Internet Privacy Enhanced Mail and the Massachusetts Institute of Technology's Kerberos project, that protect transactions in distributed computing environments.

The advantages of application-level security can be seen clearly in light of the limits of security at the lower levels.

Nearly all security uses encryption in one form or another, protecting anything from one message field to an entire message packet. Secure LAN products running at Layer 2 (the media access control layer), can protect traffic on a given LAN effectively, scrambling packet contents while leaving MAC-layer information readable. But not all traffic from a particular node may need this much protection. Here, the danger is one of overkill: misapplied fiscal resources and unnecessary performance hits.

What's more, the traffic that does need to be secured probably will be traveling beyond the originating LAN, even hopping out to other organizations via routers or application gateways (after all, internetworking is the name of the game). But once data is encrypted at Layer 2, Layer 3 addresses can't be read by routers.

Going up a layer or two to Layers 3 (network) or 4 (transport) makes internetworking possible, but security still isn't guaranteed all the way to the user, since lower-layer services often terminate at a file server or mail relay. Layer 6 (presentation) is often cited as the place for encryption, but when the Layer 6 endpoint is not the same as the secure destination, the argument in favor of Layer 7 becomes very good indeed.

In fact, only by encrypting at Layer 7, where the application itself hands off traffic for transmission to the network, is it possible to establish a secure path all the way from source to destination. Further, application-level security need not be concerned with lower-layer protocols, media, interconnections, and the like, making it feasible to "ignore the network," at least in some senses. Since enterprise networks typically rely on heterogeneous technologies (e.g., Ethernet, token ring, FDDI, unshielded twisted-pair, coax, Sonet, and SMDS), finding a lower-level product that works with all can be a problem. A Layer 7 solution, by comparison, will be transparent to network technologies and configurations.

As organizations move toward enterprise-wide networking, network-independent security will become increasingly desirable for several technical and tactical reasons.

For instance, when linking file servers to clients across a backbone, application-level security is easy to implement and can be fielded on a per-application/per-communication basis. Again, when multiple administrative domains are involved, how likely is it that each will

choose the same network-wide security product—or even interoperable ones? And how likely is it that all domains will even want to spend money on security?

With a Layer 7 solution, applications can be secured individually, and only end-users need agree on products and approaches. (Note: On a per-host basis, encryption-based solutions can be applied as far down as Layer 4.)

Equally important, security schemes above the network layer (i.e., Layers 5 through 7) don't rely on the individual security of any particular segment or network. Thus, data is protected end-to-end, no matter what security—if any—is in place. Finally, when there's not enough demand for an organization or network provider to justify network-layer security, individuals and groups can set up cost-effective Layer 7 security for themselves.

## Encryption: The Keystone

While other technologies play a part in network security, encryption is the linchpin of a variety of services. It is crucial to authenticating data origins via digital signatures; establishing message integrity with on-the-spot evidence of tampering through cryptographic calculations based on transmitted text; ensuring confidentiality and privacy; and determining beyond a doubt who is the author of a message (known as authentication of data origin).

As far as application-layer security is concerned, encryption is the only mature technology that can be readily and confidently deployed. Many algorithms for encrypting data have been devised, the most common of which is the Data Encryption Standard (DES). The better algorithms (DES included) let users employ a unique "key" to encrypt and decrypt data. Thus, simply knowing what algorithm has been used is not enough. The key itself is needed.

But that approach means that a key must be kept secret, shared only among authorized users. Keys therefore must be distributed, stored, and accessed in a highly secure fashion. And keep in mind that when secure applications are used, the availability of the secure servers is essential.

MIT's Kerberos is one such secret-key security system. It is intended to provide authenticated, integrity-verified, optionally private transactions for distributed computing. So that each user need only be responsible for a minimum number of secret keys, a Kerberos server formulates all the keys used to protect traffic and acts as a secure clearinghouse for network transactions through a sequential exchange of secure handshake messages. Kerberos has been adopted by Digital Equipment Corp. (Maynard, Mass.) and the Open Software Foundation (Cambridge, Mass.) among others, and versions are available at no cost.

More recently, encryption algorithms that work with "public" keys have been developed, the best-known of these being the RSA encryption algorithm (named for its originators, Ron Rivest, Adi Shamir, and Len Adelman).

These algorithms substitute a matching pair of keys for the single secret key; one is used for encryption and the other for decryption. But only one key must be kept secret; the other can be made public. In fact, the beauty of the scheme is that "public" means just what it says: The relationship between the two is such that private component cannot be deduced from the public one and vice versa.

Public keys are very useful in maintaining acceptable levels of security without having to meet the same rigorous operational concerns that accompany single-key algorithms. For instance, it's possible to use the designated private key to encrypt message verification data; the fact that

the matching public key decrypts it establishes a so-called digital signature, an electronic seal of authenticity regarding origin and integrity.

On the other hand, if a message has been encrypted with a secret-key algorithm, like DES, public-key cryptography can be used to achieve a second layer of protection. Encrypting the DES key with a public key ensures that only the holder of the corresponding private key can retrieve the message. And the accompanying digital signature, signed using the sender's private key, authenticates the message's origin. This, in fact, is the basic mechanism for the Privacy Enhanced Mail system being deployed on the Internet.

### Key Management Considerations

The hardware or software that performs the actual encryption and the implementations of various algorithms (RSA, DES, etc.) are increasingly to be found in off-the-shell products. That sort of availability, though, means that security-conscious network managers have to set up effective ways to keep track of keys, particularly difficult when it comes to multi-organizational configurations. Among the questions to be addressed are: Who will issue keys? How will they be distributed securely? How will they be kept? How will legitimate users access them? How will users know when keys have been compromised, expired, or canceled?

In secret-key systems, such as Kerberos, central servers transform keys into session-specific "tickets," suitable for near-real-time applications like remote log-in. This means that secret-key servers, like many other network resources, must be highly accessible. But they also remain completely secure themselves, a design challenge in itself.

For non-real-time applications, such as messaging and disk storage of sensitive data, the enterprise networking scenario is even more complicated, and in fact introduces problems that key-based systems were never intended to solve. At present, Kerberos doesn't support long-term storage of encrypted data: Files and e-mail become unreachable after a password change.

Further, it is important to understand that with any secret-key system, it is always possible to subvert the system by taking control of the key server—by licit or illicit processes.

Public keys, as noted, don't suffer from the same problem. It's possible to design public-key systems in which keys cannot be compromised without the cooperation (willing or unwilling) of end-users. There need not be a superuser who has physical access to all secrets. In fact, as the name itself suggests, public keys are intended to be made widely, easily available. The challenge is how.

The answer can be found in X.500, the CCITT standard for the OSI directory issued in 1988, which defines a database repository for a plethora of organization-wide network data. X.509, the authentication framework for X.500, defines a format for distributing keys in public-key "certificates."

The signature on certificates can be verified using the public key of the issuer, in turn obtained from another certificate. This hierarchy can be rooted at a single, common top-level certifier. Alternatively, issuers of certificates can bilaterally cross-certify one another to establish mutual trust.

The merit of public keys is that the servers need be neither secure nor real-time.

## Buying Network Support: A Shopping List

1. Do you want to buy fixed-rate contracts or pay as you go?
2. Is the service provider flexible with issues such as pricing?
3. Are networks the service provider's primary business?
4. How do you want to divide support tasks between in-house staff and outside service providers?
5. What hardware and software components do you want to cover?
6. Does the service contract cover the file server and the network operating system?
7. Do you want help with software repairs and upgrades or with learning and using the applications?
8. Does your contract cover preventative maintenance?
9. How are services priced?
10. What response time do you need, and how does the service provider define response time?
11. Do you need service during business hours or around-the-clock?
12. Does the service provider use remote access software or monitoring tools?
13. What spare parts does the service provider keep in stock, and how fast can it get parts?
14. Will the service provider make a replacement unit available to you during repairs?
15. Does the contract include a document of understanding, a scope of work document, an escalation policy, and a "no excuses" clause?
16. How does the service provider document your network and any changes to it?
17. What do current clients and vendors say about the provider?
18. How long has the service provider been in business, and what is its financial history?
19. How many technicians work for the service provider, and are they certified by the maker of your network operating system?
20. What is the ratio of technicians to sales personnel?
21. What is the ratio of service contracts to technical network engineers; is it 10-to-1 or better?
22. What is the staff turnover rate?
23. What are the service provider's biases?
24. Do you feel comfortable with the service provider's staff?
25. Can the service provider support sites in several cities?

DEC is implementing X.509 within its Network Application Support (NAS) framework. At least one free copyrighted X.500 implementation already is publicly available, QUIPU, which is distributed with the ISO Development Environment (ISODE). Software to retrieve and cache public keys also is available.

It's worth noting that certificates themselves are useful even without directory servers—they stand on their own. In the absence of an X.500 server, it's possible to establish an identity by sending all the certificates (probably only two or three) between sender and the root certifier, inclusive, in the message.

## Picking Up the Keys

This leaves the question of a mechanism for getting the keys. In secret-key systems, distribution and update tends to be done "out of band," meaning not via the network being secured.

In public-key systems, the technology itself can be used to construct a verifiable "hierarchy of trust," so that keys can be distributed securely even via the network, in the form of certificates "signed" by an issuing authority and accompanied by an authenticating digital signature. The procurement of a private key is a one-time, non-real-time action.

When it comes to Privacy Enhanced Mail (PEM), the necessary RSA technology-licensing is done by charging for each certificate, rather than the implementation and use of the RSA algorithm. There may be separate charges for the certificate itself and for the overhead of maintaining the hierarchy of trust and its associated facilities. Both are highly affordable. With PEM, for example, a certificate can cost as little as \$2.50 and is good for two years. If the generation and maintenance is done for a user, the price, including administrative fees, is \$25 for the same period.

PEM's infrastructure also allows for a mechanism that facilitates issuing RSA certificates. This so-called Certificate Postage Meter, consisting of trusted hardware and software, stores assignable certificate numbers. Under the direction of an authorized user, the meter can be directed to issue certificates. One meter can easily issue thousands of certificates a year, or an organization may want several, geographically distributed meters. BBN Communications's version of the meter costs about \$5,000, not including the charge for certificates. If an organization participates in PEM, the meter will have paid for itself as soon as it has issued 223 certificates.

Even maintenance of the meter itself is accomplished with an RSA-secured exchange: E-mail is sent directly from the supplier to the meter to maintain configuration, return billing data for certificates issued, and authorize new ranges of assignable certificates.

There may or may not be a price for using a particular encryption technology. There are, for instance, commercial and public-domain implementations of DES. Specifications and implementations for Kerberos are available free through MIT. For RSA public key technology, software that makes use of public-key certificates and verifies "digital signatures" can be found in applications and tool kits for UNIX, MS-DOS, and Macintosh environments.

For e-mail systems that use the simple mail transfer protocol (SMTP), and that means most UNIX-based systems, PEM user and administrative software will be available this year from Trusted Information Systems (Glenwood, Md.) and others. And anyone running SMTP can use PEM to add value to it. It runs comfortably with any implementation of SMTP, since it doesn't alter the protocol at all. Similarly, Kerberos is readily available for use in distributed network applications.

For many network users, application-level security will be a necessity, possibly complementing security at other layers on some parts or throughout an entire network. True, the technology is still maturing, but it's no longer out of reach for any organization.

The question is: How much can an organization rely on its networks without it? ■

# Microcomputer Encryption and Access Control

## In this report:

Products .....	4
Selection Guidelines.....	4

## Datapro Summary

Microcomputer encryption and access control products limit access to the computer and the information stored on it. Although there is no way to completely prevent unauthorized disclosure of sensitive or confidential information, these products provide a wide variety of measures to prevent such disclosure while making the information readily available to persons with a "need to know." These products implement security precautions with two separate but related technologies—encryption and access control. Encryption is the application of mathematical algorithms to change plain, readable text into an unintelligible form. It is primarily used to prevent unauthorized disclosure of sensitive information stored on a computer or a diskette. Access control involves defining procedures to restrict access to the computer itself. This is most commonly done by requiring authentication and identification of all persons attempting to use the computer, and then choosing which files, documents, or applications are available for each authorized user. Many of these products can also keep track of computer use by creating audit files.

## Technology Basics

### The Need for Encryption and Access Control Products

Within recent years, the role of the business computer has been radically altered. Instead of being occasional aids to productivity, these machines have become integral parts of virtually every business function. The dedicated computer or data processing facility is now practically nonexistent; microcomputers now sit atop individual desks, where it is more difficult to monitor their use. As computing has become decentralized, the number of people who know how to use the machines has increased.

Computers now process information that used to be stored on secured host systems, hidden behind locked doors, or secreted in secure filing cabinets. As both access and knowledge have become easier to

obtain, computer crime has also increased. This can include such actions as political or corporate espionage, attacks by terrorist groups, retaliation by disgruntled employees, deliberate mischief by hackers, and inadvertent virus and worm transmission via modems or diskettes.

Microcomputer encryption and access control products prevent unauthorized individuals from gaining access to information, despite having physical access to the computer itself, while insuring that the information is easily accessed by authorized users.

The Computer Security Act of 1987 requires all U.S. government agencies and contractors using sensitive or classified information to appropriately secure any computers that handle this information. This mandate is especially applicable in today's turbulent political climate, and microcomputer encryption and access control products can provide one element of this protection.

—By *D. A. Hess*  
Associate Editor

### Access Control

Microcomputer access control products regulate which users can access the system, the resources to which they have access, and what they can do with these resources. They are intended to prevent users from deliberately or accidentally obtaining files or data without prior permission. Access control comes in two forms: *system access control*, which prevents unauthorized users from getting access to the computer itself, and *resource access control*, which prevents authorized users from obtaining specific files, directories, or other resources without prior permission. Normally one person—the security manager or “super-user”—is responsible for deciding who can use the computer, the resources to which they can gain access, and the type of access (read, write, or none) permitted with each resource. In some cases the product can support two security managers, each with different tasks and responsibilities, for added security.

### Passwords and IDs

System access control—known as the *logon*—is normally implemented with user identification schemes (user IDs) and passwords; some products allow system managers to specify additional forms of identification, such as project or billing IDs, before access is granted. In many cases, the user ID is readily available—usually some form of the user's name—but the password is known only to the user and, possibly, the security manager.

Some products support several types of passwords, such as onetime passwords for guests and a primary-alternate password scheme in which regular users are given two different passwords. A number of products have begun to include hardware keys that use an algorithm to generate a new password each time that the user enters the system. In any case, the access control scheme should require users to change all of their passwords at least twice a year.

Some packages limit the number of unsuccessful logons—for example, the computer may crash after five unsuccessful password guesses—while others shut the system down if successful logon does not take place within a specified time period. One product increases the amount of time that must elapse between logon attempts as the number of unsuccessful tries increases. Many products track both successful and unsuccessful logon attempts and place them in a special security manager file.

### Resource Access Control

Once a user gains access to the computer, the product should have some means of tracking those pieces of information to which she or he has access, and the access level permitted for each. A number of different mechanisms can be used for this tracking. These include *user profiles*, which define information about authorized users; *resource profiles*, which define the files, directories, applications, system utilities, and input/output devices that are protected by the package; and *access control lists*, for specifying which users have read, write, execute, create, and/or delete access to the information in the resource profile.

Within these tracking mechanisms, a variety of protection schemes exist. Some products restrict information access on a file-by-file basis, while others protect information at the directory level; several make protected files or directories invisible to users who do not need to use them. With the recent increase in viruses and worms, security products are now offering even more protection, such as prohibiting

computer startups from a diskette or restricting input and output from external devices such as modems, CD-ROMs, or printers.

A number of products automatically run predefined batch files, sometimes called *execs*, when an authorized user gains access. These can usually be tailored for individual users, and are quite effective for keeping users out of the operating system or other sensitive resources.

### Audit Trails

Although microcomputer products do offer a high level of protection against unwanted intrusion by either authorized or unauthorized users, most security managers will want some process by which computer activity can be tracked and monitored. Such information is normally available in a special log called an *audit trail*.

Audit trails can respond to virtually anything that the computer recognizes as an event—logon/logoff, opening or closing applications, working with files, transmitting on a modem, or printing—as well as the responsible user, date, and time at which the event occurred. Many products also include features for tracking unauthorized access attempts, system manager functions, or project billing information. Audit trails can be read on-screen or printed in hard copy as audit reports; a number of vendors provide filters for sending audit trail information to popular database managers or word processors.

### Encryption

Encryption is the application of one or more mathematical algorithms to transform readable text—called *cleartext* or *plaintext*—into an unintelligible form known as *ciphertext* to protect against unauthorized disclosure or modification. *Decryption* is the inverse operation of encryption; when the information is retrieved from storage, the encryption algorithm is used to decrypt the data from ciphertext into cleartext. Both encryption and decryption are controlled by *keys*; these are specialized characters or numeric formulas that must be applied to the information before the appropriate algorithmic transformation can take place. Encryption is actually a sophisticated access control strategy; even if the logon or password protection is bypassed, the information itself is far less likely to be compromised if it is stored in encrypted form.

### Software or Hardware?

Microcomputer products use either hardware, software, or some combination to implement encryption schemes. Although software-only products are much easier to install and are better suited to smaller computers lacking expansion card slots, the encryption/decryption process is slower and the software itself can become corrupted or damaged, possibly causing irretrievable data loss. A number of products use expansion cards (in either half-card or full-card format) to store the algorithm, encryption keys, and, in some cases, the information itself; the algorithm is most often implemented on a specially designed microprocessor. Processor-based encryption tends to be much faster and more secure than software encryption, although the size of the computer's power supply and the availability of expansion slots may preclude the use of cards. Several well-known vendors offer optional hardware encryption boards for their software products, providing a choice for customers.

## NCSC Certification: What Is It, and Why Should It Matter?

The National Computer Security Center (NCSC) was established in 1978 as part of a Department of Defense initiative to encourage commercial development of trusted computer systems. First known as the Department of Defense Computer Security Evaluation Center, it changed its name to NCSC in 1984. By establishing the NCSC, the federal government was able to persuade commercial vendors to develop and provide off-the-shelf products for handling classified and sensitive information. By motivating commercial vendors to absorb the cost of developing secure products, the government saved federal funds from being needlessly spent on expensive, custom-designed security products, and benefited private industry as well. The "Trusted Computer System Evaluation Criteria" (known as the Orange Book) was issued in 1985. The role of the NCSC changed with the passage of The Computer Security Act of 1987, which changed the name of the National Bureau of Standards to National Institute of Standards and

Technology (NIST) and assigned responsibility for the security of unclassified information running on federal systems to that agency.

The Computer Security Subsystem Interpretation of the Department of Defense Trusted Computer Subsystem Evaluation Criteria, issued in September 1988, is an interpretation of the Orange Book criteria as they apply to subsystems. These criteria are used by the NCSC to evaluate the types of microcomputer security products covered in this report. Since such products are added to a computer after it has been installed, the products are judged by somewhat less stringent standards than, for example, a workstation with built-in security features.

The products discussed in this report are evaluated against four criteria—*Discretionary Access Control (DAC)*, *Object Reuse (OR)*, *Identification and Authentication (I&A)*, and *Audit (AUD)*. *DAC* evaluates how the product controls access by groups of users. *OR* evaluates how the product erases or overwrites

information to prevent it from being accessed by hackers or other unauthorized persons. *I&A* rates how the product implements an authenticated user-ID system to provide accountability for and control access to a computer. *AUD* evaluates how data from security-related events are captured and recorded for use in detecting security breaches and tracing the responsible party.

Evaluation results under the Trusted Subsystem Interpretation fall into four classes: *D*, *D1*, *D2*, or *D3*, with *D3* the highest. The evaluation process itself can generate several different results: the entire product may fall into a single class, it may have different rating levels for different features, or only some of its features may be certified. If a product fails the certification process, the vendor may redesign the product and submit it when its flaws are corrected.

Only the most dedicated microcomputer security vendors subject their products to this type of evaluation. The evaluation process itself is lengthy—it can take two years to perform a full product evaluation—and the entire cost is borne by the vendor. Successfully certified products qualify for lucrative contracts to government agencies and contractors.

Since microcomputer encryption and access control products are frequently updated to reflect changes in hardware and operating systems,

the NCSC developed the Rating Maintenance Phase (RAMP) to keep its evaluated products listing current. RAMP allows a vendor to maintain the security rating of an evaluated product on subsequent versions as long as the product adheres to the appropriate NCSC criteria. This means that a product need not be re-evaluated unless it appears to qualify for a higher rating or it undergoes a radical design change.

Although many businesses probably do not require the same level of trust as the U.S. Department of Defense, NCSC evaluation can be taken as a "seal of approval" to indicate the quality and effort that went into developing and maintaining a product. A successful evaluation adds to the validity and reputation of the vendor, and also indicates a commitment to providing quality security products. On the other hand, these products can be considerably more expensive than noncertified products with similar features; so the cost must be balanced against the need for certified security.

### Encryption Algorithms

There are nearly as many encryption algorithms—also known as ciphers—as there are encryption products, with many vendors offering at least two per product. Of all possible ciphers, two well-designed examples dominate the market—DES and RSA.

The *DES (Data Encryption Standard)* algorithm is probably the most widely used cipher on the market. Originally developed by IBM and the National Bureau of Standards in 1975, it was first endorsed by the U.S. government in 1977 and several times thereafter. DES is a *symmetric* cipher; the same key is used for both encryption and decryption. The algorithm itself is public; a key, known only to the user, encrypts information in blocks of 64 bits (or 8 characters). Before the encryption itself takes

place, the key bits are shuffled and manipulated by 16 consecutive linear operations, thus producing 16 different keys. During encryption, each key is applied in a single round, so that DES actually consists of 16 different encryptions. Decryption simply reverses this process. Although critics claim that it may be possible to "crack" DES-encrypted information (if someone has thousands of years and thousands of parallel computers), it is acknowledged as the most secure cypher on the market and is offered by the majority of encryption product vendors.

The *RSA* encryption algorithm, first published in 1978, is named after its inventors—Rivest, Shamir, and Adleman. RSA is an *asymmetric* cipher; one key is used for encryption, and another for decryption. Under RSA, text is converted into numbers and grouped into blocks consisting of a specified number of digits. These blocks are

raised to a numerical power  $E$ ; the exponent  $E$  is the encrypting key. The result is divided by another number  $N$ , and the remainder is the cyphertext. Decryption takes place by raising the encrypted text blocks to the exponent  $D$ , which is the arithmetic inverse of  $E$  and also the decrypting key. Again, the result is divided by  $N$  and the information reappears in plaintext form. While it is considered mathematically impossible to break RSA encryption, it has not been as widely adopted as the DES.

A great many vendors offer some sort of proprietary encryption algorithm, either by itself or as an alternative to DES and RSA. Using a proprietary algorithm offers a slight advantage: since so many are available, anyone attempting to compromise information must know which one is in use and then determine how to break it. On the other hand, no standard verification process exists for these algorithms, and the only guide may be the reputation of the vendor.

### Encryption Keys

Although the algorithm itself may be public, the keys used to encrypt and/or decrypt data need to be carefully managed. Two overall management schemes exist—*private keys* and *public keys*. The DES is an example of a private-key system; it requires only one key for both encryption and decryption, but this key is available only by those who have access to DES-encrypted files. On the other hand, RSA uses a public-key system; the encryption key is published in a public directory, whereas the decrypting key is restricted to persons who need to decrypt RSA-encrypted files.

Key selection, storage, and transmission is extremely important if the information is to remain safe from compromise or loss. The keys themselves can be selected by the security manager, the user, or the product itself. A number of products automatically manage encryption/decryption keys, whereas others require users or security administrators to maintain a list of keys; in the latter instance, valuable information can be lost if a key is forgotten. To minimize this information loss, some vendors provide special key management and recovery utilities in addition to the main security product.

### Products

Our survey represents three types of products: products offering only encryption, products offering only access control, and products that offer both encryption and access control features.

#### Encryption and Access Control Products

Products that provide encryption and access control, sometimes called *environmental control* products, operate as miniature mainframe security systems. The functions of these products often include encryption of files or file groups using DES, public-key, or proprietary algorithms; password access to the hard disk with additional identification for opening designated directories or files; and an audit trail capability.

Other features provided by environmental control products include an accelerated hard disk transfer rate for better communication between the computer and the disk, improved file handling utilities, a user interface that mimics or even simplifies interaction with the operating system, and a messaging facility which allows authorized users to post secure messages.

These products are available in hardware-only, software-only, or combination software/hardware packages; the combination packages generally use the hardware for encryption and put access control functions in the software. In most cases, the program must reside on a required hard disk, and a designated system manager must enter passwords and system specifications. The product controls the entire system operation, from logon to logoff. These products are generally designed to support from 2 to 20 users, although some products support many more.

A wide variety of encryption and access control products are available. These include Fischer International's Watchdog; Sophco's Protec; Magna's Empower I, II, and III; Micronyx' Tri-Span, Trimph! and Triumph/DES!; and Computer Associates' CA-ACF2/PC, CA-Cortana, and CA-Top Secret/PC.

#### Encryption-Only Products

Encryption-only products can be either software-based, hardware-based, or use a combination of hardware and software. Most encryption-only products feature the DES algorithm, although some vendors also offer public-key or proprietary algorithms, which may be faster than the DES algorithm. A few vendors provide one or more proprietary encryption algorithms without offering either DES or a public-key system.

Encryption-only products may offer a wider choice of algorithms, faster encryption speed, or convenience features not available in products offering encryption and access control. On the other hand, this speed and convenience comes at the expense of user identification and authorization features. Encryption-only products are best suited for single-user computers, diskless workstations, and environments (such as a small business or a government office) where access to the computer itself need not be restricted.

Relatively few microcomputer products provide encryption only. Some examples are Glenco Engineering's HardLock; Prime Factors, Inc.'s Decrypt/MS and U-Psypher; and Security Microsystems, Inc.'s DesMaster Hardware and DesMaster Software.

#### Access Control-Only Products

Access control-only products provide many of the same user authorization and secure-document grouping features found in encryption and access control products. Some products in this category are strictly designed to prevent reading from or writing to the hard disk without user distinction. Of course, these products do not include encryption capabilities.

The majority of access control-only applications are implemented entirely in software. They are inexpensive—often under \$100—and are appropriate for small organizations where access to computers must be restricted, but the information itself is not sensitive.

Access control-only products include Kent-Marsh's QuickLock, FolderBolt, and GuardCard; Key Concepts, Inc.'s SureKey/2; Kinetic Software's Kinetic MicroLok; Magna's Empower Screenlock, and Security Microsystems, Inc.'s Lockit I Extended Edition.

### Selection Guidelines

Controlling access to and securing information on microcomputers can be a costly proposition, especially when more than one machine is involved. In nearly every case, purchasing a security system involves a trade-off between



how much security is actually needed and the costs involved—not only in money, but in setup and training time, as well as impact on the users. The security manager, along with the accounting manager, MIS director, and workgroup supervisors, must determine how much microcomputer security is affordable. The following questions should be asked before actually purchasing security products?

*What level of security is required?* Organizations with predominantly single-user machines may need little security, if any. Computers located in private offices may only need a basic access control product to discourage use by unauthorized persons; in some cases, a simple screen darkener/keylock utility is all that is needed to quickly remove sensitive information from unattended computer screens.

For organizations who need only prevent disclosure of moderately sensitive files on shared microcomputers, encryption-only products may suffice. These products prevent unauthorized users from reading confidential files without the administrative overhead incurred with elaborate access control procedures.

If the information to be protected is highly sensitive, such as classified government information, confidential proposals, or restricted files, an encryption-only or an encryption and access control product offering multiple levels of encryption may be needed.

If the computers are connected through a LAN, a network security system residing on a server is often far more economical than purchasing separate packages for each computer.

*What is the security budget?* The price ranges of the products in all three categories vary considerably. The most expensive products in each category are usually hardware based; if they include both encryption and access control capabilities they may cost \$500 or more per computer. Many vendors offer site licensing agreements or volume pricing, which may reduce a large installation's cost. Nonetheless, you get what you pay for: the more expensive products usually provide the greatest flexibility.

*Is maintaining a record of system usage important?* If tracking system usage is a management priority, select a product that offers audit trails with varying levels of detail. Most encryption and combination products include some provision for audit reports.

*How many users share each microcomputer?* Most access control-only and combination products impose limits on the maximum number of users; some support as few as five, while others have no such limits. An access control-only or encryption-only product will usually offer satisfactory protection for computers that are normally used by only one person.

*Does the product support the computers and operating systems in use?* It is understood that a product intended for an IBM PC-compatible computer will not work on a Macintosh, and vice versa. On the other hand, IBM PC products may not work with every PC; some packages will not support EISA or MCA buses (MCA is found on IBM PS/2-compatible computers), while others are too large for a small IBM PC/XT compatible. Even if the software and the computer are compatible, there may be a conflict with

the operating system version in use. Most Macintosh packages require the most recent System software, while PCs using Microsoft's OS/2 may require a separate version of the security program.

Check with the vendor if any doubt exists about whether the package can be used on a specific machine or operating system.

*Can the product restrict the use of input and output devices?* Most access control and combination products can prevent booting the computer from one or both diskette drives. A number of products can also restrict or monitor incoming and outgoing modem transmissions, which provides a barrier against accidentally importing viruses or worms into a system. In environments that use sensitive or classified information, protection against accidental or unauthorized printing is also needed.

*Are expansion slots available in all the computers to be protected?* Almost all hardware-based and hardware/software products require an expansion slot. If expansion slots are unavailable, a software-based product may be the only alternative.

*Is the interface sufficiently friendly and convenient?* While important, the answer to this question depends upon the sophistication and patience of the users, and is more relevant to products with encryption features. A menu-based interface is more appropriate to unsophisticated users, while experienced users may be happier with the speed of keyboard commands; most of the better packages offer both types.

Although many products offer transparent encryption, they are frequently hardware-based and very expensive. Other products can automatically encrypt files inside the primary application, so that users need not remember to encrypt the files or waste time closing and opening the application during the encryption process.

Although a number of packages can be opened through Microsoft Windows—an increasingly popular utility in the business world—they may not be available except in full-screen mode, which prevents the users from working with multiple applications.

*Do hardware-based products have advantages over software-based products?* Hardware-based products—or products that include both hardware and software components—are generally considered more secure than software-based products, and they also provide faster encryption. One weakness in some hardware-based security solutions is that the board itself may be stolen. Some vendors have addressed this problem by issuing a software command that searches for the board before allowing the computer to start. Other vendors recommend that the user purchase a separate locking device to keep the board in place.

Software-based products may be less expensive than hardware-based products and the only solution if expansion slots are unavailable.

*Are there certain features that provide optimal security?* Many access control-only and combination products permit the system manager to restrict full or partial operating system access to specified users. This is an important consideration if maximum security is to be maintained.

Another important factor is the freedom with which the users can change passwords. This is critical in large installations where user patterns may vary. Products which let the system manager assign several password aging limits will maintain both user convenience and system security, especially if the product automatically prompts the users to change the passwords.

A product offering encryption should be capable of recovering an encrypted file even if the encryption key is forgotten. Without such a provision, files can and will be lost unnecessarily. Some products provide automatic encryption key management, eliminating the risk of forgotten encryption keys. For products which offer multiple levels of encryption, an indication of the state of a file can prevent accidental multiple encryption. ■

# An Overview of Local Area Networks

## In this report:

Products .....	9
Selection Guidelines .....	12

## Datapro Summary

The terminology of LANs has escalated from connectivity to interconnectivity, from networking to internetworking, and from operability to interoperability. All kinds of diverse equipment and software are talking to each other, and all kinds of vendors are working together to accommodate interconnected, global communications. From its simple beginnings as a departmental network, the LAN has graduated to a position of prominence in communications, cutting across office, campus, city, and country boundaries. Integrating the components in networks and the networks with other networks remains the greatest of challenges. Vendors have made peace with the existence of communications equipment from competitors and are actively promoting compatibility. "Vendor friendliness" can characterize the market's approach to the '90s. Vendors that formerly would not mention competitors' names are now boasting of their alliances and interoperability with the products of those very same competitors.

## Technology Basics

### What Is a Local Area Network?

In addition to the independent computer networking vendors, most major computer and data communications companies also are active in the LAN marketplace. They each have branded at least one of their offerings a local area network. Although in a broad, functional sense, most of them may be right, consensus holds that the term refers quite specifically to a certain class of products. For this report's purposes, we have selected the following definition:

A local area network is a system for the interconnection of two or more communicating devices that are:

*Intracompany, Privately Owned, User Administrated, and Not Subject to Regulation by the FCC:* This excludes from our definition traditional local connections over common carrier facilities, such as tie lines,

and public local networks, such as Digital Termination Services and local cable television networks.

*Structured:* Local area networks are integrated into a discrete, physical entity, with devices interconnected by a continuous structural medium. In a local area network, many types of equipment and applications, such as data processing, word processing, electronic mail, video, and voice, can operate over a single cable plant.

*Limited in Geographical Scope, with Devices Physically Separated But Not Mobile:* Devices can be on different floors of a building, on the same industrial or university campus, or in several buildings in the same city. The maximum distance, depending on the technology, is about 50 miles. Our definition excludes co-located computer systems interconnected by a high-speed parallel bus, global network systems designed primarily for use as long-haul networks, and mobile radio networks.

*Supportive of Full Connectivity:* Every user device on the network must be potentially

—By Barbara Callahan  
Associate Editor

capable of communicating with every other user device. This characteristic excludes traditional local environments that support only hardwired, point-to-point connections between a host computer and its attached terminals.

**High Speed:** Since LANs are not subject to the speed limitations imposed by traditional common carrier facilities, they usually support operations in the 1M-to-16M bps range. Minimum and maximum throughput generally ranges from 500K bps for low-speed LANs based on twisted-pair wiring up to over 1 billion bps for fiber optic LANs.

**Commercially Available:** Although in this report we examine future trends and some technologies under development, our primary concern is to provide information on the current commercial environment and its capabilities.

In local area networking, commercial availability is a matter of degree. Only the simplest LANs are true turnkey products. Most local area networks require a great deal of on-site engineering to ensure the efficient location of stations, ease of reconfiguration and expansion, accessibility for testing and repair, and compliance with building and fire codes. To ensure proper design and installation, users may have to contract with a number of secondary suppliers in addition to the primary vendor of LAN equipment. For example, although some LAN vendors provide complete configuration and installation services, others require the user to purchase and install all but the intelligent components of the network.

One final note regarding our definition of LANs. Although the purpose of any definition is to facilitate unambiguous communication, it cannot be cast in stone. For example, the average geographical scope is not the same for fiber optic LANs as it is for twisted-pair wire-based LANs. With the increasing use of repeaters and bridges within a single LAN installation, the general span length of the communications media constantly expands. Within a highly integrated multilevel network, the delimitation between a local area network and a wide area network is blurring.

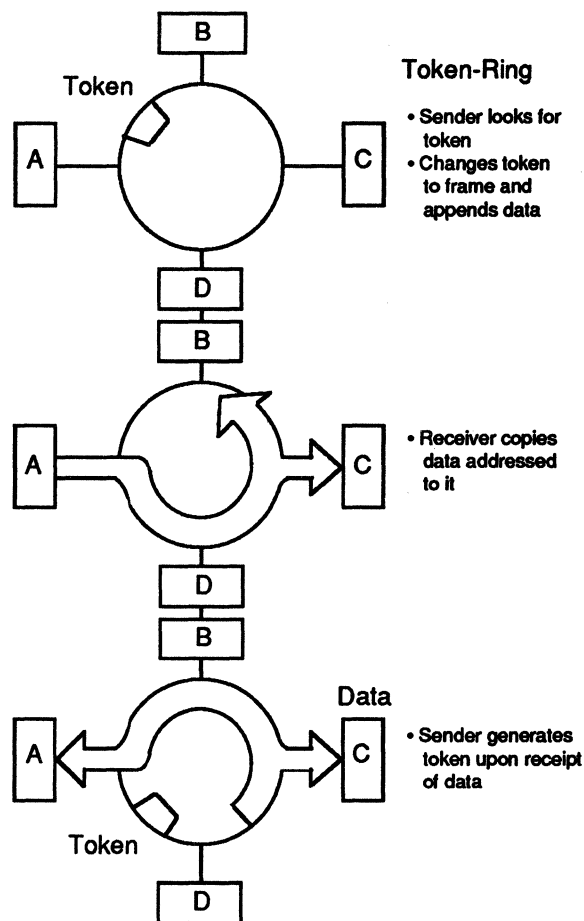
### LAN Technology

Initially, it seems that a bewildering number of arcane technologies compete for attention in the LAN market, but the number of distinct techniques used in local area networking is actually quite small. Like most modern methods of data communications, local area networks employ a multilayered model. Only a few technologies, usually two or three, compete to solve the problems of each layer.

In today's market, however, a great deal of mixing and matching among techniques at different layers creates the illusion of a complex array of technologies. The reality is rather simple when taken a layer at a time. When choosing a local area network, users face the network's:

- physical medium and transmission technique;
- topology; the logical arrangement of its stations;
- access method, the way it arbitrates among its stations for use of the shared medium; and
- higher level services the LAN offers, such as protocol or file format conversion, data encryption, or network management.

Figure 1.  
Token-Ring Operation



IBM's Token-Ring Network conforms to the IEEE 802.5 standard for token-ring, baseband LANs. The operation of a single token-passing ring is illustrated here.

Today's market offers three basic choices of media, three of topology, and two of access method. The issue of higher level services is somewhat more complex, since a vendor can offer such services on top of any practical combination of the other three factors. Presently, however, such services are only beginning to emerge as commercial offerings; they depend heavily on specific applications. Another issue is the increasingly important aspect of bridges and gateways to other networks.

This report discusses each of these issues separately, although we must state from the outset that one cannot mix and match LAN technologies at will. Some combinations are currently impractical, while others are simply impossible. Still, one should never underestimate the power of technology. Using a contention access method such as CSMA/CD over fiber optic media was once thought impossible, but LANs using such technology are now available. Ethernet over twisted-pair wire was once thought to be impractical, but today products based on the 10BASE-T subsection of the IEEE 802.3 standard are available from nearly every vendor of Ethernet solutions.

### Baseband vs Broadband

Most networking schemes offered today are baseband networks. Baseband networks carry one signal at a time at rates from 1 to 16 million bits per second. Baseband signals are always digital, with the presence of a specified voltage representing the "on" condition and the absence of that voltage representing the "off" condition.

Baseband cable cannot extend beyond a few thousand feet without expensive digital repeaters, but it is easy to install and requires almost no maintenance. Baseband networks work well within a single building, though a single baseband network can span a small campus. With current technology, baseband networks can handle only data traffic.

Broadband networks were once used as the primary means of implementing metropolitan area network backbones that might connect several buildings on a campus and stretch for a few kilometers. Signals on a broadband network are analog; bits of information are represented by variations in the strength or frequency of a carrier signal. Broadband networks can carry many signals at a time, each signal occupying a different frequency band on the cable. Broadband networks can also carry voice and video traffic as well as data. While data rates practical on any one channel of a broadband network are somewhat lower than those available with baseband transmission, between 1 and 5 million bits per second, the availability of up to 20 or 30 such channels on a single cable greatly increases the amount of data that the medium can carry.

The attractiveness of broadband networks has faded as higher transmission speeds on baseband cable and fiber optic media have become widely available. Where a broadband network would have been installed as a backbone in the mid-eighties, a 16M bps token-ring on optical fiber is much more likely to be in evidence today, with a clear upgrade path available to 100M bps FDDI when it becomes available. The several clear disadvantages of broadband make it impractical when other means to achieve the same ends are available. A broadband data network requires a careful physical design process, and its components must be tuned carefully to handle specific ranges of frequencies. Broadband networking requires a staff of trained RF technicians, for both design and everyday maintenance.

### Transmission Media

Three media types are currently practical for local area networking: twisted pairs of copper wire, coaxial cable, and optical fiber. Each type serves some applications better than others, and each supports certain transmission techniques and has its own price/performance benefits.

*Twisted Copper Wire:* Twisted copper wire can also be called twisted-pair wire or common telephone wire. Ranging in price from \$0.05 to \$0.25 per foot, it is the least expensive medium available for LAN installations. It is also the easiest to install; a user can string it along a baseboard in minutes. Twisted-pair wire is also the most readily available of all LAN media, since it is in constant, high-volume production for voice telephone use. Twisted-pair wire is best for low-cost, short-distance local area networks, especially for small networks linking personal computers. It can effectively carry data at rates up to 10 million bits per second (bps) over distances up to several hundred feet without repeaters.

Standard twisted copper wire has, however, several significant disadvantages for data transmission. It is extremely susceptible to electrical interference (noise) from

outside sources (e.g., typewriters and air conditioners). Such noise interferes minimally, if at all, with an analog voice signal, but causes two interrelated problems for data transmission. First, it limits the speeds at which data can travel, since a burst of noise that would garble only a few bits of low-speed data will destroy many bits of high-speed data. A "line hit" of a given duration garbles a number of bits that increases in proportion to the rate of transmission. Second, it limits the distance that a data signal can travel. A signal grows weaker, or "attenuates," as it travels farther from its source. Signals attenuate on all media, but twisted-pair wire's vulnerability to noise adds another factor. A length of twisted-pair wire acts as an antenna: The longer it gets, the more noise it gathers. After a given distance, the increased noise obliterates the attenuated signal.

Two techniques reduce vulnerability: shielding and repeating. Shielding makes the medium less vulnerable to electrical noise but adds significantly to the wire's cost. Active repeaters, devices that receive a signal and retransmit it to another length of wire or cable, increase the distance a signal can travel. Repeaters are expensive and add to the cost of running twisted-pair wire.

*Coaxial Cable:* Coaxial cable comes in several forms, each suited to a different kind of application. All forms of coaxial cable comprise a central conductor, the part of the cable carrying the signal, that is surrounded by a dielectric, or nonconducting, insulator; a solid or woven metal shielding layer; and finally, a protective plastic outer coating. All these layers are concentric around a common axis, thus the term "coaxial." Coaxial cable is largely immune to electrical noise and can carry data at higher rates over longer distances than twisted copper wire.

*Optical Fiber:* Optical fiber is the newest medium in the commercial LAN market. There is no doubt that, in the long run, fiber optic technology has the greatest potential as a transmission medium. Many types of fiber optic cable are available (e.g., single mode, multimode), providing varying bandwidths and transmission speeds. Presently, fiber optic hardware, including optoelectronics and connectors, is the most expensive medium for LANs, but recent developments promise to change that quickly. Netronix introduced a Plastic Optical Fiber (POF) LAN in 1989. AT&T and a partnership of Codenoll Technology Corp. and General Motor's Packard Electric division both announced Plastic Optical Fiber products during the last quarter of 1990. Cheap and much easier to install and maintain than glass fiber optical media, POF promises to bring fiber, along with the high transmission speeds possible with it, to the desktop. Glass optical fiber cable was once extremely difficult to terminate, since each fiber had to be precisely aligned to ensure a continuous connection. Recent developments in the technology have made glass fiber media much easier to work with. Codenoll's plastic fiber media can be terminated in seconds with simple-to-use hand tools.

The principal advantages of fiber optics with present-day transmission technology are sturdiness and security. Optical fiber is immune to both physical and electrical influence from the environment. Copper corrodes; glass and plastic do not. Copper conducts electricity; glass and plastic do not. Fiber optic cable is difficult to tap surreptitiously; with current military technology, operators can isolate a break, or even significant movement, in a fiber optic cable to within a single inch over a mile or more of cable.

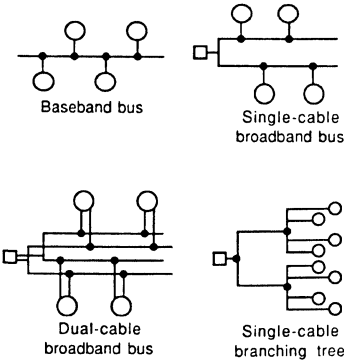

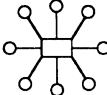
**Table 1. Comparison of Transmission Media**

	Twisted-Pair Wire	Baseband Coaxial Cable	Broadband Coaxial Cable	Fiber Optic Cable
Topologies supported	Ring, star, bus, tree	Bus, tree, ring	Bus, tree	Ring, star, tree
Maximum number of nodes per network	Generally, up to 1,024	Generally, up to 1,024	Up to about 25,000	Generally, up to 1,024
Type of signal	Single channel, unidirectional; analog or digital, depending on type of modulation used; half or full duplex	Single channel, bidirectional, digital, half duplex	Multichannel, unidirectional, RF analog, half-duplex (full duplex can be achieved by using two channels)	One single channel, unidirectional, or bidirectional simultaneously over a single wavelength half or full duplex, signal-encoded light-beam per fiber; single-encoded lightbeam per fiber; single fiber per cable
Maximum bandwidth	Generally, up to 16M bps (or higher)	Generally, up to 10M bps	Up to 400MHz (aggregate total)	Up to 200M bps in 10-kilometer range; up to 1G bps in experimental tests
Major advantages	<p>Low cost</p> <p>May be in existing plant; no rewiring needed; very easy to install; easy to support</p>	<p>Low maintenance cost</p> <p>Simple to install and tap</p>	<p>Supports voice, data, and video applications simultaneously</p> <p>Better immunity to noise and interference than baseband</p> <p>More flexible topology (branching tree)</p> <p>Rugged, durable equipment; needs no conduit</p> <p>Tolerates 100% bandwidth loading</p> <p>Uses off-the-shelf, industry-standard CATV components</p>	<p>Supports voice, data and video applications simultaneously</p> <p>Immunity to noise, cross talk, and electrical interference</p> <p>Very high bandwidth</p> <p>Highly secure</p> <p>Low signal loss</p> <p>Low weight/diameter; extremely flexible, pliable; can be installed in small spaces</p> <p>Durable under adverse temperature, chemical, and radiation conditions</p>
Major disadvantages	<p>Low immunity to noise and cross talk</p> <p>Lacks physical ruggedness; requires conduits, trenches, or ducts</p> <p>Speed and distance limitations</p> <p>Existing plant may be unsuited to data transmission (i.e., wire pairs may not be twisted; grade and quality may vary; accurate cable records may not be available)</p>	<p>Lower noise immunity than broadband (can be improved by the use of filters, special cable, and other means)</p> <p>Bandwidth can carry only about 40% load to remain stable</p> <p>Limited distance and topology</p> <p>Conduit required for hostile environments</p> <p>Not highly secure</p> <p>Rigid and bulky, difficult to install</p> <p>More expensive than twisted-pair</p>	<p>High maintenance cost</p> <p>More difficult to install and tap than baseband</p> <p>RF modems required at each user station; modems are expensive and limit the user device's transmission rate</p> <p>Rigid and bulky, difficult to install</p> <p>More expensive than twisted-pair</p>	<p>Higher cost, but declining</p> <p>Requires skilled installation and maintenance personnel</p> <p>Taps not perfected</p> <p>Currently limited to point-to-point connections</p>

*Cabling Schemes:* When discussing transmission media for local area networking, it is important to touch on two major vendors' cabling schemes: the IBM Cabling System and AT&T's Premises Distribution System (PDS). The

IBM Cabling System is a star-wired system that can connect either computer devices or telephones to wall outlets; the wall outlets, in turn, connect to central wiring closets located on each floor of a building. Transmission media

**Table 2. Comparison of Basic Topologies**

Typical Schematics	Performance Considerations	Constraint Considerations
<p data-bbox="154 348 358 369">Linear Bus Topology</p> 	<p data-bbox="586 348 1006 493"><b>Delay</b>—in token bus networks, waiting time is a fixed function dependent on number of nodes in network; in contention bus networks, delay is a variable dependent on current traffic; delay distortion ("jitter") is possible</p> <p data-bbox="586 514 1006 661"><b>Throughput</b>—in token bus networks, throughput decreases with each node added; in contention networks, throughput is best in light, bursty traffic conditions and decreases in high-volume steady-traffic environments</p> <p data-bbox="586 682 1006 756"><b>Reliability</b>—failure of one station will not affect the rest of the network; break in cable may affect only part of the network</p> <p data-bbox="586 777 1006 892"><b>Robustness</b>—relationship between stations is peer-to-peer; network is difficult to monitor; in contention networks, the difference between noise and collisions may be difficult to distinguish</p>	<p data-bbox="1016 348 1453 378"><b>Circuit speed</b>—varies up to 50M bps</p> <p data-bbox="1016 399 1453 430"><b>Distance</b>—generally unlimited by topology</p> <p data-bbox="1016 451 1453 567"><b>Maximum number of nodes</b>—user stations may be added or deleted without reconfiguring the networks; in token bus networks, addition of each station directly affects performance</p> <p data-bbox="1016 588 1453 682"><b>Error rate</b>—bit errors are lowest when fiber optic cable is transmission medium, low when coax cable is used, high with twisted-pair wire</p> <p data-bbox="1016 703 1453 777"><b>Cost</b>—generally, lower cost per user station than star networks and higher than ring networks</p>
<p data-bbox="154 913 300 934">Ring Topology</p> 	<p data-bbox="586 913 1006 966"><b>Delay</b>—waiting time is fixed function dependent on number of nodes in network</p> <p data-bbox="586 987 1006 1039"><b>Throughput</b>—decreases with each added node</p> <p data-bbox="586 1060 1006 1249"><b>Reliability</b>—if one station fails, whole network fails unless bypass circuitry has been implemented in each interface or node; if loop is severed, the whole network fails, unless redundancy features have been implemented; potentially low reliability can be compensated for by high-quality engineering design</p> <p data-bbox="586 1270 1006 1438"><b>Robustness</b>—Nodes are easy to understand, construct, and maintain; may require custom-designed, device-dependent interface; communications control overhead is generally high; if network fails, recovery may be difficult, and may require complex logic and processing</p>	<p data-bbox="1016 913 1453 945"><b>Circuit speed</b>—varies up to 80M bps</p> <p data-bbox="1016 966 1453 1018"><b>Distance</b>—limitations are imposed both on total distance and distance between nodes</p> <p data-bbox="1016 1039 1453 1134"><b>Maximum number of nodes</b>—may be fixed parameter dependent on command station capacity; addition of each station directly affects performance</p> <p data-bbox="1016 1155 1453 1228"><b>Error rate</b>—twisted-pair wire is vulnerable to transient errors; fiber optics has very low error rate</p> <p data-bbox="1016 1249 1453 1291"><b>Cost</b>—generally, lower cost per station than other topologies</p>
<p data-bbox="154 1449 300 1470">Star Topology</p> 	<p data-bbox="586 1449 1006 1522"><b>Delay</b>—in heavy traffic conditions, requests for service may be blocked at the switch in a PBX</p> <p data-bbox="586 1543 1006 1596"><b>Throughput</b>—dependent on internal bus capacity of central node</p> <p data-bbox="586 1617 1006 1690"><b>Reliability</b>—failure of one station does not affect the rest of the network; if central node fails, the whole network fails</p> <p data-bbox="586 1711 1006 1827"><b>Robustness</b>—Ready availability of network monitoring and control software; high overhead for communications control; corresponds well to applications in hierarchical (master/slave) networks</p>	<p data-bbox="1016 1449 1453 1522"><b>Circuit speed</b>—varies considerably depending on medium, to a maximum of 10M bps</p> <p data-bbox="1016 1543 1453 1617"><b>Distance</b>—limitations are imposed on distance between central node and any user station</p> <p data-bbox="1016 1638 1453 1711"><b>Maximum number of nodes</b>—expansion limitations are dependent on capacity of central node; difficult to reconfigure</p> <p data-bbox="1016 1732 1453 1785"><b>Error rate</b>—twisted pair wire is vulnerable to transient errors</p> <p data-bbox="1016 1806 1453 1848"><b>Cost</b>—high initial cost, but low incremental costs thereafter</p>

**SCHEMATIC SYMBOLS**  
 — Transmission medium  
 ○ User station  
 • Connection device (network interface unit, RF modem, transceiver, etc.)  
 □ Command station (central host, PBX switch, etc.) or cable head-end

for the cabling system consists of Type 1, Type 2, Type 6, Type 8, and Type 9 data grade shielded twisted pair; Type 5 optical fiber; and Type 3 voice grade unshielded twisted pair. AT&T's PDS is a distribution plan based on fiber optic and twisted-pair technology supporting voice, data, text, and video communications for various environments (multitenant and high-rise buildings, campuses). PDS subsystems comprise various parts of PDS including the fiber optic or twisted-pair media, cross-connect and interconnect hardware, connectors, plugs, jacks, and adapters.

The anticipated widespread acceptance of the IBM Cabling System and the AT&T PDS may signal a decline in the use of coaxial cable as the prime transmission medium for local area networks, a position that coax has held since the birth of the industry.

Table 1 compares the transmission media available for local area networking.

**Wireless LANs:** One of the most interesting new developments in LAN technology in recent years has been the wireless LAN. First introduced by small start-up companies such as O'Neill Communications and Arlan, wireless is now attracting the attention of several larger firms, most notably Motorola and NCR.

Most wireless LANs are based on low-power radio transmission, though there are also some line-of-sight systems based on laser or infrared transmission. Wireless LANs of all types are expected to find acceptance in businesses where frequent floor plan changes make LAN cabling impractical, and in more or less temporary installations, such as large construction sites. In many cases, wireless LANs will supplement traditional cabled LANs, providing quick connection to the main network wherever temporary workers or teams of specialists are deployed for a limited period.

### Topology

A network's topology is the physical and logical arrangement of its stations in relation to one another. For local area networking, we use the term "stations" rather than the more traditional "nodes." A node in a traditional data communications network sits at the intersection of two or more transmission paths and switches traffic among those paths. On most local area networks, a station attaches to a single transmission trunk at one point, catching signals addressed to it and transmitting signals along the single path to other, similar stations.

There are three basic LAN topologies: linear bus, ring, and star. In a linear bus topology, stations are arranged along a single length of cable that can be extended at one of the ends. A tree is a complex linear bus in which the cable branches at either or both ends, but which offers only one transmission path between any two stations. All broadband networks and many baseband networks use a bus or tree topology.

In a ring topology, stations are arranged along the transmission path so that a signal passes through one station at a time before returning to its originating station; the stations form a closed circle. A loop network is a ring network in which one master station controls transmissions. A star network has a central node that connects to each station by a single, point-to-point link. Any communication between one station and another must pass through the central node.

In the United States, bus and tree topologies are presently the most common, due largely to the efforts of the

Ethernet community to establish that particular bus architecture as a standard. In Europe, ring architectures are more common because much of the pioneering work in ring networking occurred at European universities. The introduction of the IBM Token-Ring Network has begun the proliferation of ring-type networks in the U.S.

In bus and ring networks, all transmissions are broadcast. Any signal transmitted on the network passes all the network's stations. The receiving intelligence in each station recognizes its address on a given signal and copies only such signals. In star networks, signals sent through the central node are circuit switched to the proper receiving station over a permanently or temporarily dedicated physical path.

Each topology has particular strengths and weaknesses. In choosing a topology, one must examine performance issues including delay, throughput, reliability, and robustness—the network's capability to continue through, or to recover after, the failure of one or more of its stations. Users must also consider such physical constraints as circuit speed (or raw data rate), maximum operating distance, maximum number of stations, channel error rate, and overall system costs. Table 2 compares the three basic topologies, with graphic representations of a number of their variations.

### Access Method

Using a network access method, the network distributes the right to transmit among its participating stations. The right to transmit is an issue only in broadcast topologies, where stations share a single, main data channel on which all stations receive and on which any station can transmit. The access method is the network's way of controlling traffic.

In general, access control can be centralized or distributed. Most conventional networks of computer terminals use central access control: A mainframe or its front-end processor polls the terminals in sequence for their transmissions. Most LANs use distributed access methods: Each station participates equally in controlling the network. There are two general classes of distributed access: random, or "contention," and deterministic. With a random access method, any station can initiate a transmission at any time. With a deterministic access method, each station must wait its turn to transmit.

**Carrier Sense Multiple Access (CSMA):** The most common random access method on today's LAN market is carrier sense multiple access (CSMA). In a CSMA network, all stations can sense traffic on the network. When a station wishes to transmit, it "listens" on the main data channel for the sort of electrical activity it recognizes as traffic—it "senses carrier." If the station senses traffic, it defers its transmission for a random interval and then resumes listening. When the station senses no traffic on the channel, it transmits.

One weakness in CSMA is that two stations may sense a clear channel at the same time and transmit simultaneously. A "collision" results—the signals from simultaneously transmitting stations interfere with one another. Many CSMA networks implement a mechanism for collision detection (CSMA/CD), which allows stations to recognize a collision, stop transmitting immediately, and resume transmission after a random wait (reducing the probability that any two stations will again transmit at the same time). Another mechanism implemented is collision avoidance (CSMA/CA). Using CSMA/CA, a transmitting



station first “senses” whether the line is free. If other traffic is already on the line, the device waits until the line is free before transmitting. If it senses that the line is free, the device waits a predetermined period of time before reserving the line via a “handshake” process.

**Token Passing:** The most widely used deterministic access method is token passing. In a token-passing network, stations distribute the right to transmit on the channel by circulating a token, a special bit pattern that assigns the right to transmit to the station that receives it. A transmitting station waits until it receives the token from the previous station in the token-passing order. When the station receives the token, it transmits its data, then passes the token to the next station. Token passing is a form of distributed polling; each station on the network polls the next station in line for its transmission. Token-passing networks require a slightly greater effort to configure than do contention networks, because each station must have not only a logical address, but also a logical place in the token-passing sequence. CSMA, CSMA/CD, and CSMA/CA are most often found in bus and tree networks; token passing is most often seen in bus and ring networks.

A network’s access method is the most important factor in determining its performance. Each access method functions differently under different kinds of traffic and on networks of different sizes. CSMA networks perform better with sporadic, or “bursty,” traffic patterns in which some stations transmit a great deal of data at a time or transmit very often, while others transmit a smaller amount of data less frequently. Performance on a CSMA network degrades as the likelihood of a collision increases. The probability of a collision increases with the number of stations likely to transmit and with the physical length of the network’s main cable (since the time a signal takes to reach the station farthest from the transmitting station affects the likelihood of that station’s sensing carrier in time to withhold its transmission). Another factor in CSMA networks is the length of the individual transmissions. A CSMA network operates more efficiently when stations transmit long individual messages rather than a large number of short messages.

Deterministic access methods perform better under uniform, heavy traffic than do CSMA networks. The number of participating stations is the most important factor affecting performance in token-passing networks, since the right to transmit must circulate through every other station before a given station may transmit again. Under any loading conditions, performance is more predictable for deterministic access methods than for random access methods.

### High-Level Network Services

The design of most LANs is based on the reference model for OSI proposed by the ISO. All the characteristics we have discussed concern the two lowest levels of network services: those that can be classed as Layer 1 (Physical layer) or Layer 2 (Data Link layer) services in the OSI reference model. These layers pertain to the physical and electrical characteristics of the transmission medium and to link access management (e.g., transmission setup, address recognition, message acknowledgment, and basic error checking).

A few vendors provide Layer 3 (Network layer) and Layer 4 (Transport layer) services. Layer 3 services involve network control, management, and maintenance. This type of control is generally software based and resident in a

network processor, controller, or master unit. Layer 3 services include call establishment, maintenance, and disconnection; end-to-end traffic routing and flow control; management of buffering between end-user devices and the network; packet assembly/disassembly; end-to-end error checking and recovery procedures; network monitoring and diagnostic services; dynamic network reconfiguration; priority and security management; and status and statistics reporting.

Layer 4 functions pertain to “internetworking,” the interconnection of one LAN to another or to a public or private long-haul network. Internetworking is generally performed through a software package called a gateway. Gateway functions include store-and-forward operations, protocol/code/interface conversion, and security procedures. Generally, the gateway resides on a single network node, and all traffic traveling between the two networks is funneled through a single port on the node.

TCP/IP is currently the de facto standard for internetworking in the United States. Originally developed for the U.S. military’s Arpanet network, the Department of Defense requires TCP/IP for all government and military contracts. TCP/IP software is now used frequently in the private sector for communications between personal computers and other supporting processors linked to mainframes and departmental systems. Until other standards allowing companies to implement an ISO-standard network supplant it, TCP/IP will continue as the de facto standard for high-level applications.

The physical location of network services within a network varies greatly from one LAN product to another. The placement of network intelligence can generally be classified on one of four levels:

1. The end-user device performs all station functions and drives the communications medium; the network provides Layer 1 services only.
2. The end-user device connects to a separate network interface unit that provides Layer 1 and Layer 2 services and a small amount of buffering.
3. The network interface unit provides end-to-end Layer 3 reliability services, plus increased provision for buffering.
4. The network interface unit is a full-fledged micro- or minicomputer that provides all interconnection functions.

Whenever a function resides in the end-user device, it is considered to be outside the network. The user is usually responsible for developing and maintaining this software, though some LAN vendors have begun to provide high-level software for certain applications. Applications-related functions—program-to-program communications, distributed database management, file transfers, peripheral sharing, network management, and applications switching—often identified as the primary purpose for a LAN, are, in most cases, performed outside of the network.

Services at the Session (Layer 5), Presentation (Layer 6), and Application (Layer 7) layers depend heavily on the specific purpose of the individual network. Different types of services are best for different applications. Users can find such services in single-vendor, comprehensive LANs such as Digital Equipment Corporation’s adaptation of Ethernet to its DECnet architecture.

**Table 3. IEEE 802.3 Specifications**

	10BASE-5 (Standard Ethernet)	10BASE-2 ("cheapernet")	10BROAD-36 (Broadband Ethernet)	1BASE-5 (1M bps Starlan)	10BASE-T (10M bps Starlan)
<b>Bandwidth</b>	10M bps	10M bps	10M bps	1M bps	10M bps
<b>Media</b>	"Thick" Coaxial Cable	"Thin" Coaxial Cable	CATV Coaxial Cable	Twisted-Pair Wire	Twisted-Pair Wire
<b>Distance</b>	500 Meters	200 Meters	3.6 Kilometers	500 Meters	100 Meters
<b>Topology</b>	Bus	Bus	Bus	Star	Star

**Bridges, Routers, and Gateways**

As the number of separate, different types of LANs in an organization grows, interest in linking them increases. Bridges, routers, and gateways all perform this function, each at a different layer of the network and each in a different manner.

**Bridges:** A bridge connects two or more networks at the Media Access Control (MAC) portion of the Data Link layer, where differences in the high-level protocols (such as TCP/IP or OSI) used on the two networks to be linked are not a factor. A bridge will pass packets of any protocol, but the station receiving the packet must employ that protocol to read the packet. Bridges, therefore, generally connect networks with common architectures and protocols. Some bridges can translate packets from networks with differing MAC-layer characteristics, such as Ethernet and tokenring.

Bridges can improve the performance of a large network by splitting it into smaller segments, thus reducing traffic on each of the resulting subnets. Bridges can also connect networks using different types of transmission media, such as coax and twisted pair, or to connect a network to a backbone running between floors or buildings.

Remote bridges may link LANs in widely separated geographical locations over telecommunications lines. Most bridges today are referred to as learning bridges, because they keep tables of network addresses. Each time the bridge reads an address it has never seen before, it broadcasts the packet in question. When the receiver acknowledges, the bridge notes the location in its table, so that when it sees the address again it will know how to get to it.

**Routers:** Routers, which operate at the Network level of the OSI model, feature more sophisticated addressing software than bridges. Where bridges simply pass along everything that comes to them, routers can determine preferred paths to a final destination. Routers can be employed in complex internetworks and can be programmed to route packets according to various criteria. They can select the cheapest or fastest route, depending on the needs of the network and its users. This additional intelligence, however, makes them slower and more expensive than bridges. In addition, a particular router only works with one protocol. In internetworks with segments that operate under different protocols, a separate router is necessary for each one, but several router devices can reside in one chassis.

**Gateways:** Gateways operate at the OSI Transport layer or above and link LANs to networks which employ different protocols, such as TCP/IP, IBM SNA, DECnet, and X.25. Packets received by a gateway must be restructured into a format understandable by the destination network. This restructuring means delays in transmission.

**LAN Standards**

The number of technologies available for local area networking, and the number of practical combinations of those technologies, creates both opportunity and confusion. The opportunity comes to the sophisticated purchaser who can pick and choose carefully among the range of available options to create a specific solution for specific needs. The confusion comes from the lack of compatibility among different commercial solutions and the end-user equipment that eventually must communicate over the local area network. The local area network market has seen two major efforts to establish standards: one launched by industry organizations attempting to legislate standards in advance of the market, and one by individual vendors and groups of vendors attempting to establish de facto standards by making their interfaces widely available at low cost. Committee 802 of the Institute for Electrical and Electronics Engineers (IEEE) has led the legislative effort; the Ethernet vendors, spearheaded by Xerox Corp., led the initial market effort.

**IEEE Committee 802 Standards:** Late in 1982, Committee 802 published draft standards for two types of local area networks. The first, Standard 802.3, was published in 1983 and has been adopted by the International Organization for Standardization (ISO) as its 8802-3 standard. It describes a baseband, CSMA/CD network—similar to Ethernet—and includes several addenda adopted since its publication. 10BASE-2 deals with 10M bps baseband networks running on thin coaxial cable. 1BASE-5, similar to AT&T Starlan, is a 1M bps, twisted-pair configuration. 10BROAD-36 is a broadband 10M bps network running over thick coaxial cable. 10BASE-T is a 10M bps network operating on unshielded twisted-pair wire. It requires two separate twisted-pair lines—one for transmit and the other for receive. Table 3 summarizes the 802.3 specifications.

The second standard, Standard 802.4, describes a token-passing, baseband or broadband, bus network, similar to the Manufacturing Automation Protocol (MAP) standard.

IBM presented specifications for its Token-Ring Network to both IEEE Committee 802 and the engineering and trade press. The result is Standard 802.5 for tokenring, baseband local area networks, published in 1985. 802.5 panels are still at work on addenda on Early Token Release and Counter Rotating Rings. Early token release is a method of enhancing token-ring network performance, and counterrotating rings provide fault tolerance through redundant data paths.

Standard 802.1 describes network architecture concepts applicable to all networks. Subcommittees are still developing addenda on bridging and network management under this standard. The Committee has also released specifications for Logical Link Control, the protocol

to be used with the two networks, in Standard 802.2. Other 802 committee work includes Standards 802.6 on Metropolitan Area Networks (MANs), 802.7 for broadband LANs, 802.8 on fiber optic media, 802.9 on LANs and Integrated Services Digital Network (ISDN), 802.10 on network security issues, and 802.11 on wireless LANs.

### LAN Applications

A local area network can support almost any application now served by conventional point-to-point communications. The implementation of a local area network can, however, be a radical and expensive step—and hard to justify if its sole purpose is simply to replace an existing cable plant for tried-and-true applications. A radical innovation must offer radical benefits. A local area network can simplify and streamline current procedures, of course; but in addition, it can offer benefits not available, or simply too expensive, with conventional local communications. These benefits vary for different applications in different environments. In the following, we list the best capabilities of LANs in several broad areas of application.

#### General Business Data Processing

Simply running computer programs has never been the job of the corporate data processing department. Data processing must also manage the data that goes into a computer and the information that comes out, design and implement software for new applications, adapt current software to changing needs, maintain current programs and hardware, plan for the expansion of existing facilities and for the replacement of obsolete components, and ensure that such expansions and replacements remain compatible with current programs and procedures. The local area network offers DP managers a chance to restore order from the chaos that was slowly creeping in via cheap personal computers, compromising the DP manager's management and planning functions.

Using a LAN, the DP manager can centralize control of the company's newly distributed computing resources, ensuring, at minimum, that each department's new computers are compatible with the network and, ideally, that they are compatible with every other department's machines. The DP manager can also ensure that all the company's decisions are based on the same data and not on each planner's custom-tailored collection of numbers. Used properly, a LAN can provide a common interface for a diversity of otherwise incompatible equipment, serving as the backbone of an orderly hierarchy of computing functions extending from the mainframe to the desktop.

In the computer room, the local area network can relieve the mainframe of the job of arbitrating among users needing access to storage and communications facilities, releasing a greater percentage of its resources for actual computing. The LAN can also streamline users' access to remote communications facilities, eliminating redundant hardware and further easing the mainframe's burden. In large DP shops, it can provide a high-throughput path for computers sharing common databases.

#### Office Automation

In a data processing shop, integrity and compatibility are big issues. The automated office requires timeliness and friendliness. In the office, the local area network can give users fast and efficient access to a common pool of information including customer lists, supplier lists, schedules, and document formats. In many cases, it allows a company to establish standard formats for files and documents for

the first time and to guarantee a minimum of deviation from those standards. A LAN can allow an entire office to pool expensive resources such as printers and duplicators, streamlining the production and distribution of paper documents.

Ultimately, the office local area network can eliminate the need to circulate paper documents by distributing schedules and memorandums almost instantly to each worker's workstation. Workstations have yet to appear on every desk in the office, but in many companies they are now more common than typewriters. In the completely electronic office, a LAN can provide nearly instantaneous desk-to-desk communications, allowing the workstation to function as typewriter, copier, and telephone for internal use. The local area network also offers management a more direct way to monitor staff performance and to control the quality of information handling throughout the office.

#### Industrial and Laboratory Automation

Efficient process control requires rapid feedback from monitoring devices to the central site. A good feedback loop requires very fast, very reliable, two-way communications. An automated factory or laboratory, with a proliferation of intelligent robots, sensors, and measuring instruments, is a natural environment for the local area network. In the factory, a LAN can simplify "retooling" by allowing the user to download new software to a number of programmable devices simultaneously from a central site. It can allow the near-instantaneous isolation of failures and bottlenecks in plant operation. By permitting feedback among a number of intelligent machines, a LAN enables managers to automate the minute-by-minute decision-making process to a degree not possible with point-to-point communications through a central mainframe or minicomputer. A LAN also simplifies the gathering of performance information, allowing designers to optimize plant operations and to plan for future growth.

Eventually, laboratory and factory process control networks will hook up directly to the engineers' CAD/CAM workstations and to the data processing department's inventory and distribution records, realizing automation's potential for the entire operation over a single, integrated system.

### Products

#### Local Area Networks for PCs

The most active and lucrative subsection of the LAN marketplace is currently the PC LAN market. Today's microcomputer LANs are complex combinations of communications hardware and software parts—some common between vendors, and some very specialized. The specialized approaches are, of course, what determine how well a given vendor's approach suits the LAN buyer's needs. The fundamental building blocks common to most microcomputer LANs can, however, be used as a basis for evaluation.

#### Hardware

In addition to cabling to physically connect microcomputers, a variety of other hardware components is involved in the construction of a microcomputer LAN. The first consideration is the choice of microcomputers for user workstations. Next, some type of interface board or transceiver is usually necessary to complete the connections from servers and workstations to the transmission medium. Finally, most LANs have at least one server station for access by

the other PCs. This must be chosen and configured to handle the anticipated demands of the network.

**Workstations:** Individual workstations on the LAN are usually microcomputers from the same or compatible manufacturers, particularly if full compatibility is required. If network software must be resident in each workstation PC, a minimum amount of memory may be a prerequisite. Standardization to a given release of the PC operating system is generally required as well. If the user stations are incompatible types, special provisions must be made for connecting them to a PC LAN. 3Com's Ethernet LANs are a good example. Using special interface cards, Apple Macintosh users can access the LAN, originally designed for the IBM PC and compatibles. Unrestricted file sharing, however, is still limited to compatible devices. Macintosh users can share files and programs (on any network server) with other Macintosh users and can receive electronic mail or straight ASCII files from any other station. Apple users cannot, however, share IBM binary files and programs, or vice versa. Their programs and file formats are entirely different, and the 3Com network, like any other microcomputer LAN, cannot bridge that kind of compatibility gap. Usually, vendors of smaller LANs do not provide any provision for mixing incompatible PCs on their networks.

**Interface Cards:** The most common interface between a PC and the network's transmission cable is a network interface card. These printed circuit boards fit inside a PC cabinet, generally into an accessory expansion slot on the motherboard. The transmission cable will usually attach directly to the card, or a short drop to the main cable might be used. The major consideration when installing an interface card is whether the PC power supply can handle the extra load. With the IBM PC and compatible PCs, the card's interrupt address and the number of Direct Memory Access (DMA) channels already in use are also important. Conflicts between addresses of the interface card, disk controllers, modem cards, etc., must be resolved. Shuffling of expansion cards may also be necessary if all available DMA channels are already in use on the intended server PC. IBM's PS/2 personal computers based on the Micro Channel Architecture, however, eliminate the problem of conflicts.

**Transceivers:** Standalone network transceivers may be required in addition to network interface cards. For example, "thick" Ethernet LANs that use heavily shielded coaxial cable generally require a special transceiver to link a PC to the LAN. Thin cable interfaces are usually made directly to the interface card, using "T" and/or barrel coax connectors. Transceivers are also available to connect interface cards designed for one type of media to a different type. For example, a transceiver can convert the signal from a card intended for coaxial cable to one that can be transmitted over unshielded twisted pair.

**Servers:** The server station provides a central repository for programs, text, and data available to LAN users. The server is usually a suitably configured microcomputer. For use as a server, a PC must have a hard disk for storing shared information and loading appropriate server control software. The microcomputer must also have enough memory to handle the overhead of the network software. Some networking systems allow a hard disk-equipped PC to also be used as a workstation while network functions

are handled in the background. This will usually slow down overall network response to requests to the server, however, since the PC's I/O capability is limited to one request at a time. For this reason, most IBM PC-oriented LAN users find it practical to dedicate a PC to server use. With the wide availability of microcomputers based on 80286 or 80386 chips and with the next generation of networking software based on the Microsoft OS/2 LAN Manager, some hardware characteristics will change and some limitations eventually disappear.

Not all PC LANs require servers. For example, certain network operating systems provide a configuration that effectively distributes the accessible database throughout the workstations. Implemented at every node, the network software is actually an operating system that converts the LAN into a continuous, multiuser information processing system. In this instance a workstation may request to share information directly with other workstations. The requester must then decide how to store the information. Options are local storage or copying back to the source.

**Superservers:** A recent development in server technology is the so-called "superserver," which is based on multiple processors and has an I/O architecture similar to that of a minicomputer, with many ports for peripheral attachment. The multiple CPUs of such a machine, generally from two to eight Intel or Motorola processors, can divide tasks among themselves to speed network transactions. Newly formed companies such as NetFrame, Tricord, and Parallan offer the most technically interesting designs so far, while traditional computer manufacturers have been quick to adopt the "superserver" label for some of their especially high-performance machines. Compaq has brought out a dual-processor model of its SystemPro microcomputer.

**Device Sharing:** Not all servers provide only disk and file service. As discussed previously, a major feature of a PC LAN is provision for sharing expensive peripherals. While this capability may be an option, most LANs have some provision for parallel printer sharing, including print spooling routines. Some allow sharing of the server's serial ports, permitting the connection of modems and plotters. If the need is great enough, a dedicated printer server may be justified. Networks needing connection to the outside world via dial-up communications can now purchase a modem server, which gives any network station access to a modem and communications software, as well as allowing the sharing of outside lines.

**Repeaters:** To overcome distance limitations of the basic network, vendors also may supply repeater or line driver devices that effectively boost the signal strength on the transmission medium, allowing the connection of two or more basic networks. The repeaters enable communications for stations beyond the limitations of a basic network. Repeaters do, however, have their own restrictions—signal propagation delays between extreme ends of a network will generally limit the number of repeaters that can practically be used together. With CSMA/CD networks, unusually long signal delays can subvert the collision detection mechanism, forcing careful adherence to the vendor's published distance limitations.

## Software

Software considerations for PC LANs are very important and can be analyzed from two perspectives: network operating system software and applications software. All microcomputer LAN vendors provide some software to facilitate basic network management, file transfer, and connectivity functions. Network services such as electronic messaging, mail, remote access, and print spooling functions may also be available as either part of the basic network software or as add-on modules. The most important consideration in choosing network software is how it interacts with the microcomputer's operating system software and with the applications software to be networked.

*Applications:* Spreadsheets, word processing, graphics, and other miscellaneous business software packages are the most common applications run on microcomputer LANs. These are traditionally oriented toward the stand-alone user, and they require no special considerations to operate in a networked environment. Users must merely have some sort of agreement in place regarding access to the data files each creates and maintains. For example, a manager may elect to make some departmental information, such as employee work schedules, addresses, meeting schedules, open memos, etc., available as public data. Other information such as spreadsheet-created budgets might necessarily be shared only with other managers, and a third type of information might be private and for the manager's review and edit only. The network software can easily handle these classifications.

Multiple access levels can be implemented by combinations of password security and the creation of public, shared, or private information volumes (storage areas). Public volumes are available for reading or copying only; private and shared volumes allow reviewing and editing by privileged users who know the password to access a volume. Initial setup of user volumes is done via administrative routines in the network system software. These provisions are usually implemented and maintained by a designated system administrator and can be changed as applicable.

*Integrity of Data:* Databases and accounting software present unique problems in a network environment, since access to data files must be controlled to ensure the integrity of the database. If one user of a networked accounting system attempts to generate monthly invoices and statements while another user is posting all applicable receivables, an unfortunate customer might be billed twice for the same purchase. While not a problem with the network per se, it is a typical consideration when migrating applications from standalone PCs to a PC LAN.

A more serious task is the implementation of database software on a LAN. Here, some provision must be made for file or record protection to avoid a situation in which two users are simultaneously updating data in the same record. When each record is stored, only the last "copy" of the record written to disk will be available for the next query. In this scenario of simultaneous attempts to update the same record, the last record written will have "old" data in the field that was just updated by the first user, and this old data will corrupt the record. Database users planning to migrate from single PCs to a LAN must keep this in mind when choosing their system, so that some data protection scheme is used. This situation can usually be addressed by updating to a "network" version of the software, by provisions in the LAN's system software, or by

the LAN's server configuration itself. Network versions of applications software have either file- or record-locking provisions to prevent users from simultaneously accessing the same record. They usually depend on some interaction with the LAN's operating system for this protection.

*Record and File Locking:* Most LAN vendors provide some file or record locking as part of their system software, and as long as users access the files in a way the network expects them to, these provisions will eliminate nearly all data corruption problems. Another type of corruption protection is provided by network servers that qualify as true file servers, as opposed to disk servers that merely respond to workstation requests for information copies. (Today, however, LAN server software has reached such a sophisticated level that users would have a hard time running across a plain disk server program.) File servers have the capability to intercept requests for access to a given file and interleave them as the file is available for edit. This protection may extend to the record level, allowing simultaneous updates of files from multiple workstations without exposure to corruption problems. Some file server programs, such as Novell's NetWare, not only provide concurrency checking but transaction management as well. NetWare's Transaction Tracking System (TTS) provides a rollback capability to transactions that did not complete due to power failures or other incidents that interrupt the processing of a transaction.

*Future File Systems:* Most of the file integrity issues raised here will, we hope, be nonissues within the next few years. This is because the next generation of personal computer operating systems (e.g., OS/2) and network operating systems (e.g., OS/2 LAN Manager based) will have a consistent filing system. Application programs will have a consistent set of compatible programming interfaces that would support their running on a network without the need for tricky, and most often inconsistent, file- and record-locking facilities. Network programs, such as 3Com's 3+Open and the IBM OS/2 LAN Server program, which are both based on Microsoft LAN Manager, are designed to support back-end processing of a global database application. They have an integral SQL database engine (IBM OS/2 LAN Server) or are compatible with such a database engine (Microsoft OS/2 LAN Manager and Novell NetWare).

Finally, if the network is installed by a systems house offering the microcomputer LAN as part of a solution, a special adaptation of applications or system software may be provided as part of the installation. In this case it is best to check with the systems house regarding the specifics of file security before implementing the LAN.

---

## Future Trends

*Token-ring:* The IBM Token-Ring Network is expected to become the dominant networking scheme in the early 1990s, overtaking Ethernet as the leader in number of installed systems. Several clear advantages of token-ring make it an attractive choice. The deterministic token-passing access method makes performance under load and slow to degrade as traffic increases. This predictability and consistency also facilitate expansion, as administrators can judge the effect additional stations will have on the network before actually installing them.

Token-ring is also more fault tolerant, since workstations are attached in a physical star topology to a hub called a multistation access unit (MAU), and the MAUs are linked to form the ring. This arrangement of hardware makes faults easier to isolate and facilitates quick repair.

IBM's range of systems, from PCs to the largest mainframes, can attach to its Token-Ring Network without special devices such as bridges. Keeping in mind that IBM still holds the largest share by far of the business computing market, it should be obvious that this kind of connectivity makes token-ring the network of choice for many potential users. IBM's 1988 introduction of 16M-bit-per-second token-ring presently makes it one of the fastest networks available.

While IBM dominates the token-ring market with a 90 percent share, other vendors have their own ideas about issues in token-ring implementation. The Open Token Foundation (OTF) was formed in December 1988 to help vendors adhere to IEEE standards for token-ring. OTF's 10 full members and 16 associate members include 3Com, Proteon, Gateway, Digital Equipment, Texas Instruments, and National Semiconductor. Thus far, IBM has declined OTF membership but joined OTF members in a multivendor interoperability demonstration at the 1989 NetWorld show in Dallas.

Several vendors have begun to market 16M token-ring networks that run on economical unshielded twisted-pair wiring. IBM initially opposed this trend on the grounds that unshielded cabling is prone to noise interference, particularly at high speeds, as in the 16M bps version. IBM has, however, joined with other token-ring vendors in an IEEE 802.5 Unshielded Twisted Pair study group. Other vendors participating in this group are AT&T, NCR, Proteon, Ungermann-Bass, David Systems, SynOptics, Western Digital, and Cabletron.

**FDDI:** The Fiber Distributed Data Interface (FDDI) is an ANSI standard for a high-speed fiber optic network. Running at 100M bits per second, FDDI is expected to be used as a high-speed backbone connecting multiple LANs in a single building or as a Metropolitan Area Network (MAN) connecting LANs dispersed over several buildings. Other areas where FDDI is likely to be used are in process control and other realtime applications, where high-speed response is critical to performance, and in applications such as medical imaging that involve the transfer of graphics.

FDDI employs a token-passing ring topology with two separate rings, a primary and a secondary. Under normal conditions all traffic travels on the primary ring. If a cable fault interrupts data flow on the primary ring, the network stations involved can automatically reconfigure the path to use the secondary ring. Any station attached to the physical ring must be a dual attachment station; that is, one that connects to both primary and secondary rings. Another attachment method allows many single attachment stations to connect to concentrators on the physical ring, but in the event of a cable break between the concentrator and the single attachment station, automatic reconfiguration is impossible. Concentrators will, however, provide an economical method of attaching several stations to the ring.

**Distributed Processing:** Fully distributed application processing is perhaps the most important development on the local area network horizon. With the full implementation of IBM's OS/2 LAN Server under the Systems Application Architecture (SAA) umbrella, and with network implementations based on the Microsoft OS/2 LAN Manager

and complying with the IBM SAA specification, all connected nodes, from the desktop workstation to the mainframe, would become peers. These developments portend that applications will communicate not only with each other peer to peer, but that processing loads can be dynamically distributed for efficient use of CPU resources. For the PC LAN market, distributed application processing offers many advantages.

The major drawback of current PC LAN systems is the inefficient use of CPU power across the network. File servers do nothing but push blocks of data between their local disk storage and client stations' memory and/or disk storage. Because all processing is done at client stations, even a simple query in a database application creates much data traffic over the cable. Bottlenecks can then ensue, negatively affecting response time when there are many such activities running simultaneously. Security is also compromised, because the client station gains access to the entire database instead of receiving extracted information.

**SQL Databases:** The biggest promise of OS/2 networks is support of SQL-compatible database engines on server machines. This capability allows database applications to execute in two modes—front-end processing of user programs takes place at a user workstation, and back-end processing of core database functions takes place on the database server machine. The potential for improvements in security and reductions in unnecessary network traffic is substantial. LANs for interconnecting OS/2-based systems are also said to provide basic network management functions, including audit trails, network use statistics and report generation, and realtime network administration. Because the industry still awaits a standard for network management, some of the network management software that would be available cannot be expected to be full functioned.

Vendors are wary of incompatibility problems if and when a network management standard is adopted. In the meantime, some vendors feel that offering network management products that are partially compatible with IBM's NetView/PC network management program is a safe bet in the short term. The local area network was never designed to facilitate distributed applications processing, but to provide ease of communications among interconnected computers. Incompatible hardware, operating systems, and communications protocols that many organizations must contend with, however, necessitated some level of integration in which LANs would play a key role. LANs alone cannot resolve the incompatibility problems in today's computing environments, but they have been the catalyst for impelling standardization efforts, and they surely will continue to do so.

## Selection Guidelines

### Advantages and Restrictions of LANs

Today's local area network is a wonderfully sophisticated engine for moving several streams of data concurrently, rapidly, reliably, and inexpensively from one physical interface to another. This is, of course, rapidly changing; LANs are slowly becoming the vehicles for true distributed data processing. Every major advantage now offered by a local area network, however, is balanced by one or more restrictions. Some restrictions are built into the technology; others will fall by the wayside as vendors and standards bodies advance the technology.

### Resource Sharing

A local area network allows a large number of intelligent devices to share resources, including storage devices, program loads, and data files. Sharing of hardware such as disks, printers, and connections to outside communications distributes the cost of that hardware among all participating devices and offers large savings compared to the installation of individual disk drives, printers, and modems at each station on the network. Sharing software enhances security, since all attached devices use not only the same version but the same master copy of a given program, and further reduces the need for separate storage hardware. Sharing data increases the reliability of a database, ensuring that changes made by one user are immediately available to all other users.

Resource sharing is perhaps the greatest advantage currently offered by local area networking. Unfortunately, current commercial technology limits resource sharing to mutually compatible devices. Incompatible computers or workstations can share the same disk drive but cannot read or update each other's files. Some users can implement LANs in which one vendor provides all-inclusive solutions: the network, the computers, the storage media, and the software. In such systems, all devices are mutually compatible by definition. Other users, especially those using the local area network to integrate an existing array of incompatible devices, cannot accept a single-vendor solution. For new users, or those who wish to replace an entire existing facility with new equipment, the single-vendor option presents a different set of disadvantages: It precludes mixing, matching, and price shopping.

Some users are large enough or sophisticated enough to design their own answers to the compatibility problem, but even for these users, the effort can be difficult and costly. Most users must rely on a vendor to ensure compatibility, and many vendors know that compatibility sells products.

Many local area network vendors offer aids to compatibility including protocol conversion and file format conversion. Vendors specializing in networks for personal computers are most likely to offer format conversion, especially among file formats for popular models such as the Apple Macintosh and the IBM Personal Computer. Vendors of large-scale, general-purpose networks offer protocol conversion, usually allowing asynchronous ASCII terminals to emulate IBM 2780/3780 BSC or IBM 3270 BSC or SDLC terminals. For many local area networks, protocol conversion is an everyday task; such networks must convert any end-user signal to an internal network protocol for transmission.

Although many commercially available LANs still do not offer protocol or file format conversion, their numbers are shrinking rapidly. The movement toward standardization offers some hope, but more for communications protocols than for file formats. Even for protocols, it may be too early to adopt a universal standard for local area networking. Just as the potential for resource sharing is the major current advantage of local area networking, the incompatibilities blocking that potential's realization form the major current restriction.

### Integration of Functions

The capability to integrate a wide range of functions into a single, harmonious system is another potential advantage of LANs. A local area network can provide a rational framework around which management can build everything from office procedures to strategies for planning, purchasing, and growth. By focusing on the LAN, creative

managers can establish an orderly hierarchy of job functions and of hardware, facilitating the flow of responsibility and information in their organizations.

Implementing a management system in hardware is, however, restrictive. A large investment in a given organizational plan generates a proportional amount of inertia against which efforts to change that plan must struggle. Increasing the efficiency of a good system makes it better; increasing the efficiency of a bad system makes it worse.

No hardware system is a panacea, however attractive it may be. The greater a technology's potential to affect an organization, the more carefully managers must plan its implementation. A local area network is only a tool. Creative management can make it a powerful and effective tool.

### Higher Channel Speed

A high data transfer rate is inherent in our definition of a local area network. Most LANs transfer data at rates ranging from 1M bps to 16M bps, and the Fiber Distributed Data Interface (FDDI) will reach 100M bps, a rate many times faster than those available over conventional switched facilities. High throughput rates are indispensable for such applications as high-resolution, movable color graphics, which need megabits of information to paint a single screen, and bulk data transfer among mainframe computers. Users must realize that these high data rates apply to throughput over a multiplexed facility, the network's shared main data channel, and are difficult to translate into turnaround and response times applicable to end users.

Turnaround time on a local area network depends as much on the kinds of applications sharing the network as on the total throughput. Three or four high-resolution CAD/CAM stations can generate as much network traffic as 30 or 40 word processors. Fully interactive applications can generate more than twice as much traffic as simple data entry.

Potential LAN users must avoid infatuation with data transfer numbers and look carefully at the size of the proposed network and the nature of the applications to be installed. The network's access method also plays a large part in determining throughput and turnaround time under different loads.

### Simplicity and Flexibility

Most local area networks use a simple and elegant architecture with control distributed among the participating stations. Since the entire network does not depend on a single polling or switching device, such networks tolerate isolated failures quite well. A hardware or software failure in one station usually affects only that station. Distributed control also eases reconfiguration and expansion; participating devices in most LAN architectures need not be aware of the precise number or arrangement of the other stations. Users can move or add stations on such networks with relative ease. A LAN's simplicity and flexibility are among its most notable selling points. Again, however, poor initial planning can negate this advantage. Users must plan for both device failures and growth. A faulty network attachment unit is easy to replace, but only if a spare is on hand. An inflexible cable layout might have to be replaced completely in order to expand the network.

### Security

By design, most local area networks are easy to tap. This makes networks easy to expand and reconfigure but makes

it virtually impossible to prevent simple physical intrusion. At the current state of the art, lack of data security is arguably the biggest disadvantage of a local area network. New, simpler tapping mechanisms exacerbate the problem. Some vendors have addressed the security problem by implementing data encryption as an add-on feature, but encryption can only prevent the use of intercepted data. A relatively unsophisticated vandal can still easily jam or destroy data.

Users should never allow plant security applications, such as card-access locks or security video, to share the same cable plant as everyday data applications. Separate networks should be established for secure and open facilities; if necessary, such networks can be interconnected through a secure bridge or gateway that can block unwanted signals. If possible, redundant cabling should be installed, so that an intentional or unintentional break in the cable will not bring down the network.

The increased use of optical fiber in LANs will alleviate many of these security problems. Fiber optic cable is virtually impossible to tap and is immune to electromagnetic interference (EMI) and radio frequency interference (RFI). See the section on Transmission Media under LAN Technology for more information on optical fiber.

### Alternatives to LANs

For some applications, LANs are either too costly or too unsophisticated technologically. Other technologies are available that can do the job better.

For simple port selection or port contention among one or more computers and a network of terminals, a local area network is simply overkill. In such situations, users wish only to gain access to host-resident applications; the host computers handle requests for storage and peripheral service. Their terminals are usually unintelligent, asynchronous devices with no capability to share software with the host or with one another; these terminals also need to communicate with only one computer. Applications requiring simple port selection and contention are most common on university campuses or in their industrial counterparts' research and development labs.

For port selection among a small number of similar computers with similar operating systems, a shared front-end processor is more effective than a local area network. When mutually incompatible computers are involved, a port selection switch or a data PBX is the best choice. Both front-end processors and data switches have been available for years and offer all the benefits of proven technologies: stable interfaces; time-tested maintenance and control procedures; established, reliable vendors; and a history of satisfied customers. Users should not risk using an infant technology for simple applications when less risky alternatives are available.

Large-scale, fully integrated voice and data networks lie at the other end of the spectrum. No existing LAN can handle a full load of voice telephone traffic along with a full load of data. Local area networks are a creation of the data processing industry, and their technology has bypassed, not solved, the problems of voice communications that the telephone industry has been addressing for decades.

From the telephone industry comes the voice/data PBX, a circuit switch built on a digital matrix and designed to handle both voice and data traffic. In data handling applications, such systems represent a technology

even younger than the LAN, and their data transmission capabilities are still somewhat narrow. Nonetheless, they do offer a number of advantages for large offices in which management would like at least the capability to place a terminal on every desk. Any large office must have a telephone network and, in a large plant, the costs saved by having data applications share that network rather than installing a separate cable plant for data can more than make up for the high initial cost of a voice/data PBX. The big switches' present data handling limitations are not really an issue; vendors will have corrected any bugs long before current LAN vendors teach their networks to switch voice calls.

Both data switch and voice/data PBX vendors are heating up the competition with LAN technology. The resulting products are hybrids of local area networks and traditional circuit switches; individual switches handle local communications in their own domains while participating in a network of similar switches based on a LAN technology such as the token-passing ring. Several vendors of data-only switches now offer such hybrids.

### Pricing Considerations

Costs are one of the hardest elements of LAN design to analyze. Comparison is difficult because, in many cases, users are trying to compare disparate systems and because costs are declining rapidly as the market develops. In considering price, many of the rules of thumb that apply to long-haul networks and to data processing technologies in general are applicable to LAN networks:

- Older, established technologies are less expensive than state-of-the-art and experimental technologies.
- Equipment available off the shelf is less expensive than customized systems.
- Prices increase as the level of network services increases.
- Price increases as speed and performance increase.

One frequently used leveler is "average price per workstation connection." The price per connection for a LAN may, however, vary widely, depending on what the price includes. When we asked vendors to approximate the price per connection for their LAN products, the prices they quoted ranged from \$50 to nearly \$40,000, depending on whether the price was simply the cost of tapping a user station into a preexisting LAN, or whether the total cost of the whole network system—including cabling, transmission components, network management hardware and software, connection hardware, installation, and technical support—was averaged over a typical number of workstations.

When requesting preliminary price estimates from several vendors, we suggest users describe a configuration that approximates their needs and specify which components should be included in the price, giving true grounds for comparison. Right now, the typical price to tap a user terminal into a LAN (\$200 to \$1,000) may be more than the cost of the terminal itself. As with other new technologies, however, prices are dropping as the number of users increases and volume production methods are implemented. When LAN standards are accepted, and the interface-on-a-chip becomes available, the connection price is expected gradually to become negligible, comparable to the cost of installing an extra telephone connection on a PBX system. ■



# An Overview of Bridges, Routers, and Gateways

## In this report:

A Guide to Selecting Bridges .....	4
Linking Ethernet and Token Ring Networks .....	5
Local Routers and Brouters .....	7
A Checklist for IBM Mainframe Gateways .....	8
Bridges vs. Routers .....	11

## Datapro Summary

The process of linking together two or more local area networks is becoming increasingly popular in organizations. Local area network components such as bridges, routers, and gateways play a major role in the efficiency and security of companywide networks. While it is sometimes difficult for network managers to determine the need for a bridge from a router, the hybrid combination of these components, called a "brouter," is ideal under certain circumstances. The technical merits of bridges, routers, brouters, and gateways are examined.

## The Internet or Enterprise Network

Anyone who has followed the local area network industry over the past five years knows that large companies no longer concern themselves with linking two small LANs together. Today, companies want to know how to link together several LANs that might or might not be located at the same physical site. While the term *enterprise network* is starting to become popular, I use the traditional term internet to describe several networks linked together. Each individual network is known as a *sub-network* because it is part of the larger internet.

## What Is a Local LAN Bridge?

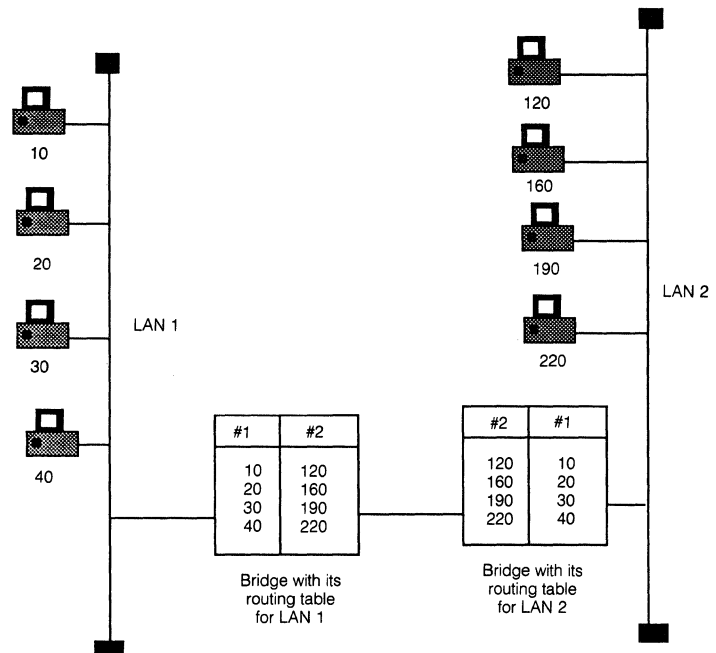
A local LAN bridge consists of the hardware and software required to link together two different LANs or subnetworks physically located at the same site into one internet. The simplest type of bridge examines a packet's 48-bit destination address field and compares this address with a table that lists the addresses of workstations on its

network. If the address does not match any of its workstations, it forwards the packet across the bridge to the next network. These simple bridges keep forwarding packets hop by hop until they reach a network containing a workstation with the desired destination address. This process of examining address tables and forwarding packets is referred to as *Transparent Bridging*; it is a technique used by all Ethernet bridges and by some Token Ring bridges. Figure 1 illustrates how this type of bridge operates.

Some bridges create their own network address tables. These bridges examine the source and destination addresses in every packet transmitted onto the LANs to which it is connected. These bridges then build their own address tables that list workstation source addresses found on packets they have seen on their network with this network's corresponding number. These bridges then try to match the destination addresses of packets with one of these source addresses. When a bridge matches an address, it *filters* the packet and sends it on its way along the network where the destination workstation will recognize its own address and copy the packet to the RAM in its RAM. If no match is found, then the packet is *forwarded* and permitted to travel across the bridge to the next network. Broadcast and multicast packets are always forwarded because their destination address fields are never used as source addresses.

Reprinted, with permission, from book #3577 "Linking LANs: A Micro Manager's Guide", by Stan Schatt. Copyright © 1991 by Windcrest, an imprint of TAB Books, a division of McGraw-Hill, Blue Ridge Summit, PA 17294. (1-800-233-1128 or 1-717-794-2191.)

Figure 1.  
A Simple Bridge With  
Transparent Bridging



Bridges do not understand or concern themselves with higher level protocols. They function at the Media Access (MAC) sublayer of the OSI model's Data Link layer, far removed from the upper level protocols such as XNS or TCP/IP. As long as networks on both sides adhere to IEEE 802.2 Logical Link Control (LLC) standards, a bridge can span them regardless of differences in their media or network access method. As you will discover by reading this report, this means that it is possible for corporations to bridge their Ethernet and Token Ring networks, as well as their 802.3 LANs that might include a StarLAN, a 10BaseT Ethernet, and a thin coaxial cabled "cheapernet" network.

## Why Use Bridges

Of a number of network design reasons for using bridges, a few include increased efficiency, security, and distance. Efficiency is usually the most often cited reasons why bridges are installed on a network. Because bridges are capable of filtering packets according to programmable criteria, a network manager could use a bridge to reduce traffic congestion and improve speed by dividing up a large network and then bridging the resulting subnets. The two smaller networks would run faster because they had less traffic.

Because larger Ethernet networks are slowed down by collisions, it makes sense to create smaller Ethernet subnetworks and use a bridge to provide such services as E-mail. Ethernet has a maximum length limitation of 2.5 KM and also restrict the number of network segments to three to avoid exceeding the 9.6 microsecond propagation delay. Network managers and systems integrators can overcome both Ethernet limitations by using bridges.

The 4 Mbs version of Token Ring Network limits the number of network workstations to 72 with unshielded twisted pair and 270 workstations with IBM's own shielded cabling. Network managers can overcome these limitations by using smaller subnetworks and then bridging them. The smaller subnetworks operate more efficiently and are easier to manage and maintain.

Another reason why bridges make a network more efficient is that a network designer can use different topologies and media wherever appropriate and then link these different networks via bridges. Offices within a department might be linked via twisted pair wire. A bridge could connect this network to the corporate fiber-optic backbone. Because twisted pair wire is so much less expensive than fiber-optic cabling, the network design saves money and increases efficiency by using the high bandwidth medium on a backbone where the most traffic is carried.

Bridges can link two similar networks that have different transmission speeds. As an example, a company might be perfectly happy with an 802.3 1 Mbs StarLAN network using unshielded twisted pair for one department but require a 10Base5 Mbs 802.3 network using thick coaxial cabling and transmitting at 10Mbs for its manufacturing plant. A bridge buffers packets, so it is no problem for it to span packets from LANs with different transmission speeds.

Because the 802 committee developed a common Logical Link Control layer for its various network topologies, it is possible, for example, to link two Token Ring networks that are separated by an Ethernet LAN. The Ethernet LAN can forward the packets just as a mail carrier can deliver a letter written in a different language as long as the envelope (packet) follows the rules and regulations established by the IEEE 802.2 LLC standard. While added efficiency is a major reason for using bridges, bridges also can increase security. They can be programmed to forward only those packets that contain certain source or destination addresses so that only certain workstations can send information to or receive information from another subnetwork. The accounting subnetwork, for example, can have a bridge that permits only certain workstations outside this network to receive information. In addition to providing a security barrier that filters out unwarranted access, bridges also add a measure of system fault tolerance. If a single file server on a large network fails, the entire network fails. If, however, internal bridges are used so that two file servers back up each other continuously, then traffic is reduced

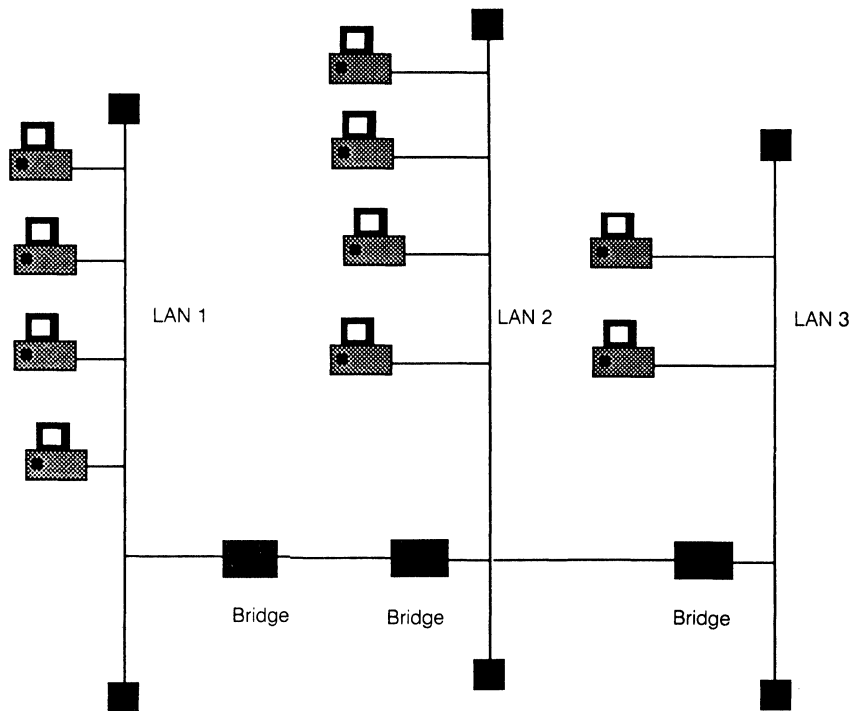


Figure 2.  
Cascaded Bridges

while providing an additional measure of security in the form of a "backup" file server.

Finally, bridges can also increase the distance that a network can span. Because a bridge rebroadcasts a packet to the workstations on the receiving side, it functions like a repeater to increase the distance a packet can travel without its signal attenuating. Often bridges are *cascaded* to connect LANS sequentially as seen in Figure 2.

### Intelligent Bridges

All bridges have the ability to update their routing tables; that is how they keep track of which workstations have been added to the network. Intelligent bridges differ from the "simple" bridges I have been describing by offering additional capabilities. They can be programmed to filter packets based on desired criteria. I referred to an application appropriate for this type of bridge when you examined how a bridge can enhance network security by restricting traffic to and from specified subnetworks. An Ethernet segment for the accounting department might be connected to the rest of the network via a programmable bridge that only permits workstations on the corporate subnetwork to access it.

Intelligent bridges also might offer *Source Explicit Forwarding* (SEF). This feature enables network supervisors to assign internetwork access privileges by labeling specified addresses in a routing table as either accessible or inaccessible to specific users and groups.

### The Spanning Tree Bridge

The *Spanning Tree algorithm* (STA) was developed by DEC and Vitalink and later adopted as a standard by the IEEE 802.1 committee. Spanning Tree is an approach toward bridging multiple networks where more than one loop might exist. Figure 3 illustrates how multiple paths might connect one network with another. Data can be transmitted between LAN 1 and LAN 2 a number of different ways.

Under STA, each bridge has an identifier that consists of a priority field and a globally administered station address. Bridges negotiate with each other to determine the route data should take. A *root bridge* is selected during this negotiation process on the basis of having the highest priority value. If two bridges have the same priority value, the one with the higher station address is selected as the root bridge. While this process proceeds automatically, the network manager can "fix" the results by giving a particular bridge a higher priority value.

After the root bridge has been selected, each bridge determines which of its ports points in the direction of the root bridge and designates it as the root port. If more than one bridge is attached to the same LAN, a single bridge is selected based on the one that offers the least "cost" based on criteria established by the network manager. "Costs" could include such elements as line speed and buffer capacity. If all costs are set as equal, STA will produce a tree-like structure with bridge ports selected that result in the least number of hops from bridge to bridge for a packet to be transmitted from one LAN to another LAN.

Now that all these negotiations have taken place, each bridge sets its root port in a forwarding state to move data toward the root bridge. It also sets its port pointing away from the root bridge in a forwarding state. Other bridge ports are blocked so that packets cannot travel through them. As Figure 4 illustrates, STA ensures that only one bridge port in each direction on a LAN is operating and that only the most efficient path is available.

What happens under STA if a bridge port goes out of service? A port that has been blocked is placed in a learning mode so that it can examine packets flowing on the network and update its database of network station addresses. The port then changes to a forwarding mode and forwards a notification of its change to the root bridge. The root bridge notifies all network bridges to update their databases to include this new bridge port.

A bridge's address database includes information on the direction to forward data for a particular workstation as well as a timer. If the timer for a particular workstation

## A Guide to Selecting Bridges

Bridge vendors are having a hard time differentiating their products from each other. The noise level and confusing jargon associated with bridges have grown to the point that selecting a bridge is a confusing and frustrating task even for knowledgeable systems integrators. In this section, I examine several features to look for in bridges. While you might not need specific features, this section will help you make intelligent decisions.

### Packet Filtering and Forwarding Rates

Some bridge vendors like to boast about their products' packet filtering and forwarding rates. The filtering rate in packets/second (pps) measures how quickly a bridge examines a frame, matches

its address with its address table, and then decides whether to filter or forward it.

Frame sizes can vary as can network traffic. Vendors often provide information on the number of frames/second required for various frame sizes at 50% and 100% loads. With Ethernet, 100% load conditions are unrealistic because of the number of collisions under Ethernet and the subsequent dead time on the cable. Check to see if the statistics are based on the same size frames.

A network manager or systems integrator must look at how a network will be used on a day-to-day basis in order to make sense out of

these statistics. Say that network will be utilizing the TCP/IP transport protocol heavily. This protocol provides unacknowledged delivery service, which means that the network could become flooded with TCP/IP datagrams because source workstations do not have to wait for acknowledgements before sending large files.

Also, bridge vendors rarely specify the conditions under which their statistics were gathered. Just as the manufacturers of dot-matrix printers publish statistics on character/second that fail to take into account the use of any special printing features, filtering and forwards rates for bridges rarely indicate the traffic conditions on both sides of the bridge.

Statistics on bridges for Ethernet-like half-duplex networks rarely point out that these statistics hold only for the condition in which the bridge is forwarding packets to a network with no current

traffic to slow up the process. That is simply not realistic. Similarly, bridge statistics are usually based on the very simplest of bridges and ignore any programmed filtering even though that might be the precise reason why the network manager selects a specific type of bridge.

Statistics indicate that the key to a bridge's success could be the way it handles bursty traffic. The size of a bridge's queue or holding area for frames is critical. The queue should be large enough to handle reasonable temporary overloads, but not too large to introduce excessive delays. If a queue is too small, the frames will be lost when they back up in the queue. If the queue is too large, frames will remain in the queue past their "time-out" and still be transmitted by the bridge that is too simple to consider such circumstances.

expires, then the information regarding which direction to forward data for this workstation might no longer be valid. The bridge monitors the source addresses on packets it receives and updates its address table. If a new network topology is introduced, the result might be to require a change in the direction to forward data information in a bridge's address database.

### Source Routing Bridges

In 1985 IBM introduced Source Routing with its Token Ring network. In fact, IBM's PC LAN program and its OS/2 Extended Edition version 1.1 are both designed to work only with Source Routing bridges. *Source Routing* is a bridge that actually performs routing duties at the Network layer of the OSI model. Each device on a LAN that uses Source Routing must have a unique six byte address. The address field's first high order bit, called the I/G bit, indicates whether there is an individual or group address. The second high order bit, the U/L bit, indicates whether the address is universally (IEEE addresses assigned to manufacturers) or locally administered.

Source Routing takes the I/G bit in the source address only and uses it as the Routing Information Indicator (RI) bit. When this bit is set to 1, it indicates the presence of additional routing information in the frame header. This additional routing information (up to 18 bytes in length) specifies the frame's complete path from source workstation to destination workstation. Figure 5 illustrates the location within a Token Ring frame of this critical routing information.

Each LAN ring is assigned a unique number just as an office routing slip might indicate the order in which a memo should be circulated before it reaches the file clerk for filing. If two bridges on the same LAN are parallel and can lead to the same destination, Source Routing will arbitrarily assign each of them a different number to keep the routing directions from becoming confusing. Figure 6 illustrates what this routing information would look like if it were translated into English.

The length of this information field limits routing to eight ring numbers, which means a maximum of seven bridges; Source Routing bridges refer to this limitation as a seven-hop limit.

### How a Workstation Gathers Source Routing Information

A workstation gathers Source Routing information by transmitting an *all-routes broadcast frame* to all rings connected on an internet. This frame contains control information but a blank buffer that can be filled in by other workstations. Bridges fill in the numbers for the two rings they connect and their own bridge number. The destination workstation receives this broadcast frame and returns it to the source station, which now has a road map of the route that the frame took.

It is possible to use a spanning tree topology with Source Routing bridges. A special *single-route broadcast frame* is circulated once. It ensures that only certain bridges in the network are configured to pass single route type frames. Because no loops are permitted in a spanning tree topology, between any two rings only a single path can exist. Bridges will not pass a frame onto a ring if it already

**Filtering on the Basis of Packet Length**

Some bridges have the ability to filter packets based on the actual packet length. If a network has a lot of interactive traffic so that response time will be faster. By programming a bridge to block longer packets during heavy traffic periods, a network manager can keep response times bearable.

**"Learning" bridges**

Some network managers must deal with networks where some users are frequently moving from area to area while other users are being added or deleted on a daily basis. Some bridges require the network manager to modify the bridge's network address table each time there is a change. Other bridges are able to learn the locations of devices by examining the source address fields of packets they handle and then modify their own tables. These "learning" bridges are worth the additional expense when network managers spend inordinate amounts of time manually modifying bridge address tables.

**Link Ports**

The ports on some bridges can be individually configured. In the case of the Retix Model 2265 local LAN bridge, for example, this means that one port could be configured for a standard 802.3 network (10Base5) while the other port could be configured for thin Ethernet (10Base2). This particular bridge also has a StarLAN option to link a StarLAN network with Ethernet. In the case of the Hughes 8050 Broadband/Ethernet Bridge, it is possible to link a baseband Ethernet LAN with a broadband 2 Mbs Ethernet network.

**The Ability to Filter Broadcast and Multicast Packets**

Some bridges have the ability to filter broadcast and multicast packets. Broadcast storms consist of a packet that is broadcast and then endlessly replicated until it creates so much traffic that it brings a network to its knees. By being able to filter and restrict broadcast and multicast packets, bridges can reduce broadcast storms.

**Load Balancing**

Load balancing makes it possible for multiple ports to carry information to the same destination. By balancing the data traffic on two 56 kbs lines, the effect is to widen the total bandwidth transmission to 112 kbs. Different bridge vendors have their own proprietary methods of implementing load balancing. Some bridges simply divide up all traffic evenly using a first-come-first-served approach. Others handle a specific queue first before handling a second queue. Sequencing protocols used by these bridge vendors can be important because packets might arrive out of order.

Load balancing also provides a system fault tolerance feature because the built-in redundancy means that some level of communication is maintained even if one line goes down.

**Bridge Management and Statistical Software**

Some bridges come with network software that provides network management features and the ability to generate statistical reports on the

bridge's activity. Most software provides information on the number of packets filtered, forwarded, refused, and rejected.

DEC's LAN Bridge 100 offers some very sophisticated statistical reports on its Ethernet bridge including network utilization and throughput, the top ten protocols used, and the top ten transmitting stations. It also provides information on which workstations transmit multicast packets. Because DEC's Local Area Transport (LAT) protocol can only be bridged and not routed, DEC network managers need this protocol information because they cannot obtain it by using routers.

has circulated on that ring. A workstation uses the information it receives from its two types of broadcast frames to determine the optimal route containing the least number of hops for the frame it transmits.

**The Sole Routing Bridge in Operation**

A Source Routing bridge examines every frame on each of the Token Ring Networks it links. If it sees the RI bit set to one, it then examines the routing information field to see if the two ring numbers match the two rings it connects. Assuming the routing information matches, the bridge forwards the frame across the bridge. Frames that do not have matching ring numbers are filtered.

**NetWare and Source Routing**

NetWare has always used a proprietary distributed routing algorithm at each of its bridges and in its file server. Novell responded to its customers who have Token Ring networks by providing a Source Routing algorithm so that NetWare bridges can communicate with Token Ring bridges on other LANs.

**Linking Ethernet and Token Ring Networks**

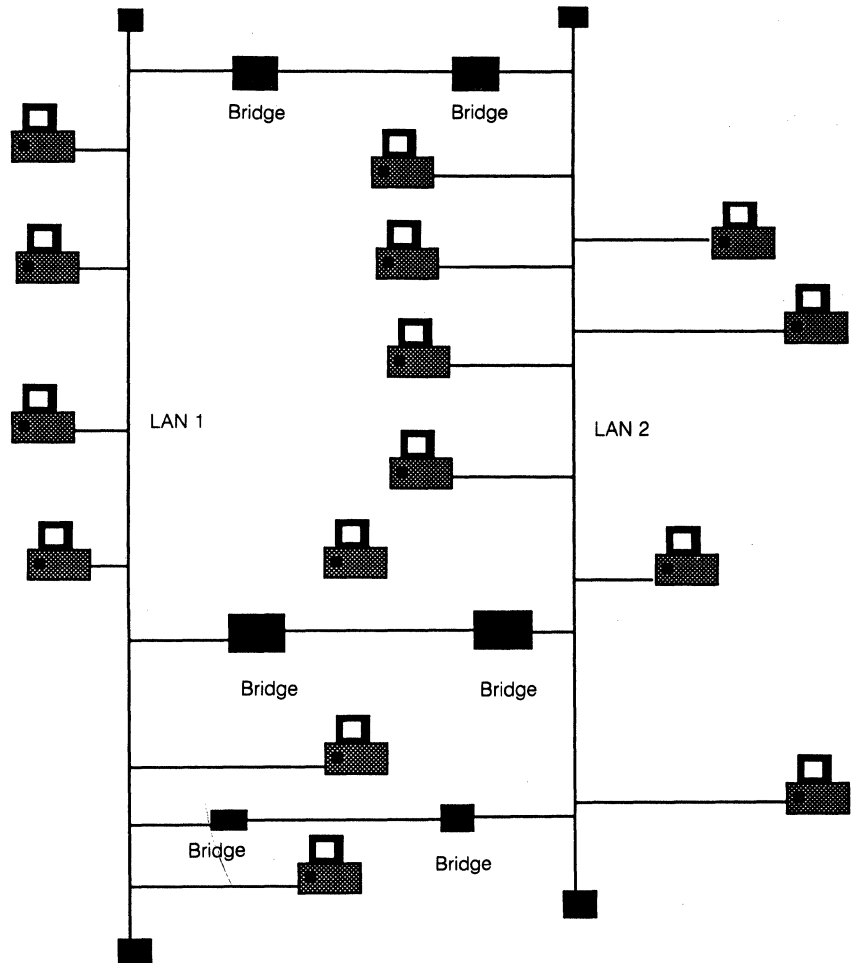
Some industry surveys estimate as many as 75% of Fortune 500 companies have both Ethernet and Token Ring

LANs. The accounting department might be using Ethernet to connect PCs running productivity software such as WordPerfect and Lotus 1-2-3 to their VAX computer running DEC accounting software. Other areas might be using IBM's Token Ring Network. What is a network manager or systems integrator to do in such a situation? You have already examined the differences in frame structure between the two networks and also seen that the significant differences between Spanning Tree and Source Routing approaches to routing frames.

It is absolutely critical to remember that there is a significant difference between connectivity and interoperability. *Connectivity* refers to being able to link together the two networks and transmit data while *interoperability* refers to the ability for each network to use the data transmitted to it.

Sometimes connectivity is all that is needed. Say that you have several Ethernet networks in an enterprise internet along with a 16 Mbs Token Ring network that serves primarily as a backbone, a gigantic switching station. While the 802.3 and 802.5 frames differ, they do have a common MAC layer. The Token Ring network can forward Ethernet frames through its ring and onto a bridge connected to another Ethernet network. The Token Ring Network cannot "open" the frame and understand the data contained within it, but it can understand the Source

**Figure 3.**  
**Bridging Offers Multiple**  
**Paths Between Two LANs**



and Destination address fields. What a Token-Ring-to-Ethernet bridge does is to support Source Routing on the Token Ring side and transparent bridging on the Ethernet side.

Bridges are available today that can perform the changes in the frame required to convert an Ethernet frame to a Token Ring frame. With such bridges, workstations on the Token Ring side view the Token Ring bridge as just another bridge. Workstations on the Ethernet side, however, view the bridge as just another Ethernet workstation. Frames generated from the Token Ring side addressed to an Ethernet workstation to the bridge, where they are stripped of the Logical Link Control (LLC) protocol. They are converted into Ethernet frames and transmitted over to the Ethernet network. Frames sent from an Ethernet workstation to a Token Ring workstation must go through an additional step. The bridge must search its own address database to learn the additional routing information required for Source Routing over Token Ring networks.

The CrossComm Token Ring to Ethernet bridge family is one of the first bridges to perform this critical task. It supports higher protocols including NetWare, TCP/IP, and the 802.3 LLC protocol. As far as media, it supports thick and thin coaxial cabling, twisted pair Ethernet and StarLAN, and fiber-optic Ethernet and Token Ring. The bridge is designed to detect Ethernet packets that do not have a Source Routing information field and insert this field so that they can travel on the Token Ring side of the

bridge. The actual protocol conversion that takes place is handled by CrossComm's proprietary Dynamic Conversion mode technology.

IBM's 8209 LAN Bridge can also handle the Ethernet to Token Ring protocol conversion. Because there is a significant difference in maximum frame size between Ethernet (1500 bytes) and Token Ring (approximately 5000 bytes), the 8209 bridge uses part of the Token Ring protocol to indicate to the source workstation that the maximum frame size it can use is 1500 bytes. The smaller frame sizes add overhead to the file transfer because more frames are required.

The 8209 bridge looks like a Source Routing bridge to Token Ring workstations while Ethernet workstations see all Token Ring workstations as workstations on the same Ethernet segment. Because Source Routing uses redundant parallel bridge connections while Spanning Tree permits only a single path, the 8209 bridge permits multiple connections but only one path can be active at any given time. The 8209 bridge operates in three different modes: Token Ring to Ethernet version 2, Token Ring to 802.3 LANs, and a mode in which the bridge detects the type of LAN and then switches to mode 1 or mode 2.

### Source Routing Transparent Bridges

Does a *Source Routing Transparent Bridge* sound like a contradiction? IBM has proposed an addition to the IEEE 802.1D standard for transparent bridges that will define a

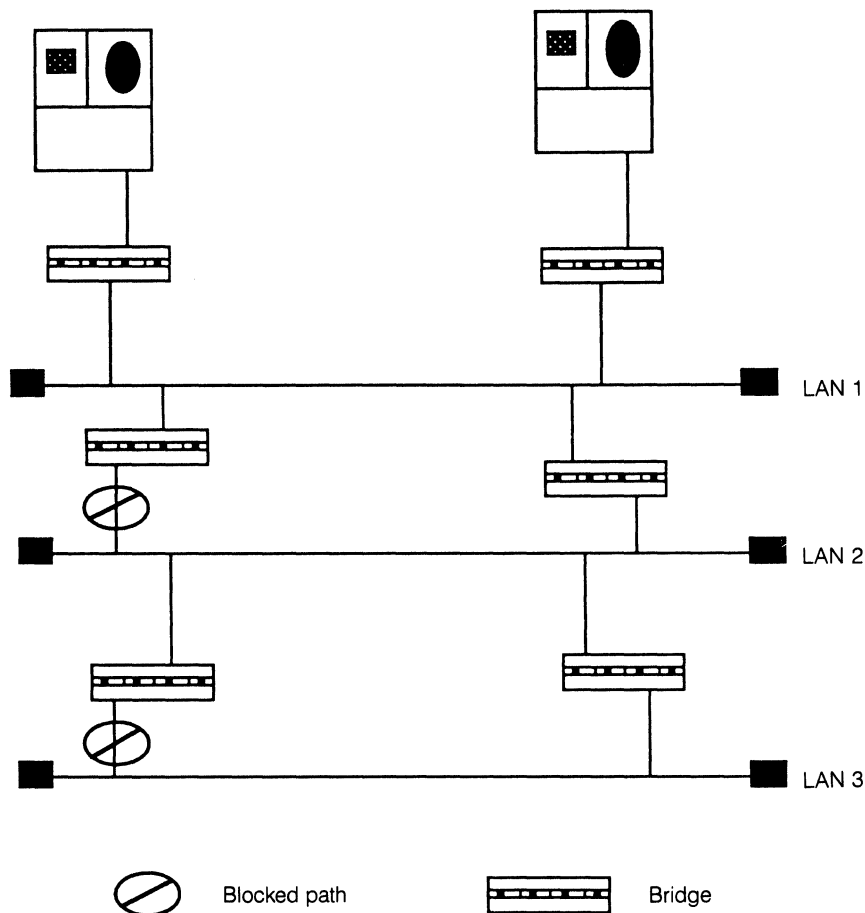


Figure 4. Spanning Tree Algorithm in Action

Source Routing Transparent bridge (SRT). This bridge will be able to forward both Spanning Tree and Source Routing frames. IBM's proposal would seem to benefit virtually everyone in the interest of interoperability except for those customers who have already purchased IBM's own Source Routing bridges; they alone will be cut off from enterprise networking.

The Source Routing Transparent (SRT) bridge uses the routing information indicator (RI) to distinguish frames using Source Routing from those using transparent bridging. Source Routing frames set their RI indicators to one, making it easy to distinguish this group of frames.

The movement to SRT bridges will not be painless. Many industry experts point to the hardware modifications to current Source Routing bridges that will be expensive. A casualty of this new type of bridge will probably be the current solution for bridging Ethernet and Token Ring networks by performing what amounts to a protocol conversion, a transformation from one frame format to another. The future 802.1D bridges will not require anything more from the workstations on the Ethernet side because Ethernet bridges handle most of the work associated with the Spanning Tree algorithm. Workstations on the Token Ring side, however, will have to construct their routing tables and build their frames to accommodate the workstations on the Ethernet side.

### Summary of Bridge Concepts

Bridges link subnetworks into an enterprise-wide internet. By dividing large networks into smaller subnetworks and

bridging them, network managers gain greater efficiency because there is less traffic congestion. The subnets provide greater security because of their redundancy than is possible with a single network. The Spanning Tree algorithm is used in Ethernet bridges. It requires a dingle path with no loops. Token Ring networks use Source Routing, an approach that places responsibility upon the source workstation to develop the complete routing path for a frame. While current Ethernet to Token Ring bridges perform protocol conversion and transform a frame from one format to the other depending upon the direction of the transmission, future bridges will be able to read a frame and transmit it in a transparent manner.

### Local Routers and Brouters

#### Routers Create "Fire Walls"

One major advantage that a router has over a bridge is that a router does not automatically replicate all broadcast messages. This means that if a device begins to flood a network with copies of a single packet, the routers are able to keep the problem local by presenting a "fire wall" that prevents the storm from engulfing the entire network.

#### Easier Management of a Large Internet

Routers can take advantage of addressing schemes such as the one used by Internet Protocol (IP) to create subnetworks. An IP address includes a network number, subnetwork number, host number, The XYZ Corporation might have a corporate headquarters as well as three local plants, each of which has its own subnetwork. The "host" number

## A Checklist for IBM Mainframe Gateways

This section briefly describes major gateway features. From this shopping list, you should be able to formulate the questions you need to consider as a network manager or ask as a systems integrator in order to evaluate the wide range of gateways now available.

### Number of Simultaneous Terminal Sessions on a PC

Gateways provide varying numbers of simultaneous (concurrent) terminal sessions on a single PC. Is a DOS session supported separately, or is it part of the total sessions permitted a terminal in the vendor's literature? The number of terminal sessions available to a PC range from 1 to 32 among the more popular gateway products. Is there a hot key that makes it easy to switch back and forth from DOS to mainframe terminal emulation?

A gateway that emulates an IBM 3299 looks like 8 terminals to a cluster controller. If each terminal can support 5 sessions, then the gateway provides a total of 40 sessions for distribution among network workstations. The total number of concurrent sessions supported by a gateway can vary widely from less than 16 at the low end to 254 at the high end.

### Terminals Supported

Some gateway software packages can emulate all members of the 3178/9 and 3278/9 families of terminals; others are limited to the more common models. Increasing numbers of LANs will require graphics terminal emulation in the near future because virtually all software is becoming more graphics-oriented. Does the software support IBM 3179 G or 3279 G graphics terminals?

### File Transfer Capabilities

While network managers have been able to transfer binary files back and forth between mainframes and micros using a tedious record-by-record approach, they have pushed for more efficient software that would let them encode, compress and block data, calculate check sums, and then decompress, unblock and decode the files at both ends. IBM's IND\$FILE protocol now has become an industry standard. Some vendors such as Data Interface Systems Corporation provide their own proprietary file transfer software as well as IND\$FILE.

Some vendors improve the speed of IND\$FILE by increasing the sixth of the PC gateway's buffer. Another key file transfer feature for gateways is to permit file transfer to take place in host session. This ability enables multiple file transfers to take place simultaneously at each workstation. These file transfers might be taking place between a workstation and different hosts or on the same host. The gateways's ability to perform these file

transfers in background mode is important for a LAN's busy gateway. If host to LAN file transfers is a major function for your proposed gateway, then it is important that you determine the file transfer protocols supported by the gateway.

### Protocols Supported

Virtually all gateways support IBM's Synchronous Data Link Control (SDLC) protocol. A number of gateways support Binary Synchronous Control (BSC) protocol, an older half-duplex approach still found on a significant number of mainframes. Many remote sites still use Remote Job Entry (RJE), a method of submitting work to a mainframe in batch format. Other key protocols that you might need supported include X.25, VT100 for communication with DEC systems, and Burroughs for communication with Burroughs mainframes.

### Support of Programming Interfaces

In enterprise networks of the near future, companies will want far more connectivity than simple terminal emulation. They will want to automate logon procedures, create custom screens, automate many of the LAN to

refers to any IP device that has an IP address; these devices can take the form of bridges, personal computers, mainframe hosts, etc.

Figure 7 illustrates a large company that has four major networks. The VAX is used to handle accounting, while other departments use a Banyan server running VINES, a 3Com server running 3+ Open, and a Novell file server running NetWare. Rather than concern itself with the incompatibilities of the various network operating systems, the company opted to run TCP/IP on all four networks and then link them together via routers.

Routers can be programmed to be very selective as to the class of service they provide. This enables the company to permit electronic mail to flow and to restrict certain types of accounting information. Notice also that the routers offer a redundancy that means that if one path is blocked, packets can be routed via an alternate path.

Finally, routers make network management easier by offering network management software that monitors and controls network operations. The TCP/IP routers used in this example provide Simple Network Management Protocol (SNMP).

### Point-to-Point Protocol (PPP) for Multivendor Router Communications

Many examples in data communications of standards are not really very standard. The RS-232C standard is a good example; printer manufacturers never have agreed on which pins should be used for specific handshaking functions. As a result, you cannot simply substitute one printer's serial cable for a different vendor's printer cable and expect it to work.

The same sort of problem exists with routers. While TCP/IP routers use the Internet Protocol (IP) for routing their datagrams, vendors used Serial Line Internet Protocol (SLIP) as a basis for developing different ways of encapsulating IP datagrams. This has made it difficult or even impossible to mix and match TCP/IP routers on the same network. Lack of router compatibility has been a particular problem for enterprise networks in which one plant might be using a low-end router to satisfy its simple needs while another facility might be using a sophisticated router from another vendor to meet its complex needs. Recently the Internet Engineering Task Force (IETF) completed work on its Point-to-Point Protocol (PPP).



host file transfer tasks, and develop communication links between mainframe and LAN programs. Does your gateway support the necessary Application Programming Interfaces (API's) to enable your programmers to write the necessary code? IBM's high Level Language Applications programming Interface (HLLAPI) enables mouse support.

A gateway's support for IBM's Advanced Program-to-Program Communications (APPC) protocol for distributed processing could also be important in the future. APPC is already being used on IBM's SNA Distribution Services (SNADS) to send documents and files back and forth between two different systems. IBM's Distributed Data Management (DDM) uses APPC to transmit data between a client system and a database server. As a key component of Systems Application Architecture (SAA), APPC will be a tool for corporate programmers who want to establish links between programs running on different IBM systems.

APPC incorporates two relatively new SNA protocols. PU 2.1 permits two processors to communicate on a peer-to-peer basis, while LU 6.2 permits two application programs to have a peer-to-peer conversation. In the future, it is likely that programmers will use APPC and LU 6.2 to write code so that two programs running on different machines such as a host and a microcomputer will be able to exchange information without users needing to be aware of the how to communicate directly with the host. This information exchange will be completely transparent to the computer user. LU 6.2 will also be used to write multiple front-end applications including spreadsheets and databases running on client workstations that communicate via Structured Query Language (SQL) commands with a shared database running on a server or host.

#### **Dedicated and Pooled LUs**

Some gateways permit dedicated LUs, an approach that guarantees that a given LU will be available to a LAN node when it requests one.

Pooled LUs are available to all LAN nodes on a first-come, first-serve basis. On most gateways these pooled resources are freed-up and reusable once a user terminates a session.

#### **Gateway Management**

The types of gateway management features supported vary widely among gateways. Some provide a gateway monitor, which enables a network manager to examine all session assignments and enable or disable devices. Some gateways let the network manager configure dedicated devices to attach to assigned devices automatically when a workstation is initiated. Dynamic device attachment logic retains prior assignments for each device and then attempts to reserve a free device for its most recent user unless or until no other device is available for another attaching user.

#### **LAN Features Supported**

Because gateways were available before LANs became popular, some vendors have put more effort into making their products "LAN-friendly," while other gateway vendors have provided

minimal LAN functionality. Can the gateway recognize directory paths without the need to switch to the proper directory? Does the gateway communicate over NetBIOS or over Novell's IPX protocol? NetBIOS versions differ and problems could arise on a large NetWare LAN if the gateway does not support IPX. NetWare network managers also prefer gateway software that runs in a directory that can be set to Read-Only. In other words, what has the gateway vendor done to make this product easier for network users to operate?

#### **Remote Speeds Supported**

If your gateway is to be remote, the transmission speed it will support is critical. Some gateways still only support 19.2 kbs modems while other support up to 64 kbs. You might have to consider multiple remote gateways to reduce traffic congestion.

PPP replaces SLIP with a standard method for encapsulating IP datagrams. This new protocol means that systems integrators can design direct serial connections between TCP/IP routers at very high speeds ranging from 9.6 Kbs for dial up lines to T-1 and fractional T-1 service.

#### **Open Shortest Path First (OSPF)**

The U.S. Department of Defense's Internet Activities Board sets internet policy for TCP/IP users. The Board has created a task force known as the Internet Engineering Task Force OSPF Working Group to develop a dynamic routing protocol for TCP/IP that will provide features not offered by Routing Internet Protocol (RIP).

RIP has limitations when used with networks of more than 100 routers because of RIP's reliance on the Bellman-Ford algorithm. This approach requires the frequent broadcast of the entire routing table. On large internets with over 100 routers, routing updates take longer and longer and consume increasing amounts of bandwidth.

RIP has other limitations. Packets cannot travel through more than 15 routers from sender to receiver. This protocol selects a single path to each destination and is not capable of considering such factors as traffic congestion, delay, and bandwidth.

The Open Shortest Path First (OSPF) protocol uses a link-state and shortest-path-first algorithm. Each router

broadcasts a packet that describes its own local links. Routers collect information from these broadcast packets to build their own network routing tables. Because these packets describing local links are very short, they cause far less traffic congestion than RIP's approach of broadcasting very large routing tables describing the entire network.

Another advantage of OSPF is that network managers can configure their routers to provide least-cost routing according to whatever criteria these managers define as a "cost." Unlike RIP, OSPF does not limit the number of routers that can be used nor does it limit routing to a single path; loads can be distributed over several different paths to optimize available bandwidth.

OSPF provides far more flexibility than RIP when it comes to type-of-service. This new protocol offers eight classes of service with separate paths available for each path. Network managers can program their routers so that certain types of packets (large file transfers that are not time-sensitive) are sent via satellite with delays that could stretch to several hours. Time-sensitive packets, on the other hand, can be given a class of service that will route them via more expensive phone lines.

The good news about OSPF is that the task force has developed procedures so that large networks can run both RIP and OSPF as a dual protocol stack as a temporary

**Figure 5.**  
**Source Routing Information Within a Token Ring Network Frame**

Broadcast	Length	Direction	Largest frame	Rl1...Rln
-----------	--------	-----------	---------------	-----------

Broadcast	Indicates whether this is a broadcast frame.
Length	Indicates the length of the routing information.
Direction	Indicates to a bridge whether this frame is traveling from source to destination or back again.
Largest frame	Indicates the largest size of the MAC information field.
Rl1...Rln	Route information for each "hop" which includes a 12-bit LAN and a 4-bit bridge number.

solution while they work on converting to the latter protocol. OSPF can route information to RIP transparently so that users are not even aware of this conversion process.

## Routing Features to Consider

It is a jungle out there, with dozens of routers that each claim to be superior. Whether a particular feature is really beneficial or not is a decision that you will have to make as network manager or as the systems integrator designing the entire project. This section examines some of the more significant features routers offer so that you can make an informed decision.

### Number/Type of Local and Wide Area Network Interfaces

The number of interfaces available on a router vary widely according to vendor and model. Proteon's p4200 series router, for example, offers 7 LAN ports and 14 WAN ports. Because routers are protocol specific, this particular router is designed to handle Ethernet versions 1 and 2 to IEEE 802.3 networks and IEEE 802.5 networks to Proteon's own proprietary ProNet-4 and ProNet-50 networks. Multiple WAN ports are so important because WAN links are much slower than local links and the bandwidth is much smaller.

The type of WAN interface on a router is just as important as the number of ports available. A network manager or systems integrator must select the appropriate router model for a specific network design. While many models offer RS-232-C, R-449, and CCITT V.35 interfaces, a few also offer a fiber-optic FDDI interface. In the case of AT&T's StarTroup X.25 router, an X.25 interface to public data networks is even offered.

### Network Management Protocols Supported

On large networks it is essential to be able to gather detailed routing statistical reports as well as "fine-tune" routers for optimal performance. Routers vary widely as far as the network management software they support. Some support the IEEE 802.1 network management standard,

others support Simple Network Management Protocol (SNMP). Some routers support only the vendor's own proprietary network management software. Systems integrators must consider long and hard whether or not they want to be locked into a specific proprietary network management scheme that probably will remain static and not grow the same way as industry supported standards.

### Router Performance

The number of packets per second (PPS) a router can handle is very revealing. There is a lot of overhead involved in routing decisions. The PPS figure takes into account the time required for routers to access their tables and decide on the optimum path. Unfortunately, every router vendor has a different way of measuring PPS. Are you dealing with packets traveling in both directions? Are you dealing only with packets that do not have to be segmented to travel to a network with a different packet size? How many packets are lost? What happens to speed when the router is programmed to permit only certain types of packets to pass?

### Protocols Supported

Because multiprotocol routers are available today, it is essential that the systems integrator consider future as well as present routing needs. TCP/IP has different broadcast formats; will the router's protocol implementation be able to support the different TCP/IP versions on an enterprise network? Similarly, while TCP/IP and its accompanying SNMP management protocol might be acceptable for the present, does the router also support the OSI suite of protocols and its CMOT management protocol? OSI protocols that could prove essential in the future for network routing include ConnectionLess Network Protocol (CLNP), End System to Intermediate System (ES to IS) routing, and Network Service Access Point (NSAP).

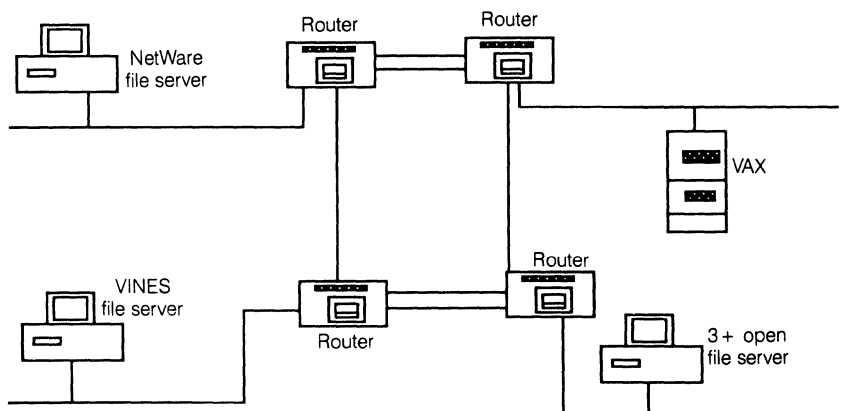
What types of networks will join the enterprise network in the future? If UNIX systems are on the drawing board, then Routing Information Protocol (RIP) is essential because it is the interior routing protocol used on Berkeley-derived UNIX systems. Will there be communication in the future with the Defense Data Network (DDN)? The DDN supports Exterior Gateway Protocol (EGP), which is also known as Request for Comment (RFC) 888 and 904. Security requirements on the Department of Defense network require an IP router that can provide datagrams that support connection to the Blacker interface for secure public data networks (X.25). Figure 8 lists some of the major protocols routers support along with the standard that defines them where appropriate.

### Security

Some routers offer security options that enable you to filter out packets bound for proprietary or secured systems on the basis of their IP addresses. These routers can also be programmed to filter on the basis of message type; this means that electronic mail can be permitted while file transfers can be banned.

**Figure 6.**  
**A Closer Look at a Source Routing Information Field**

Control (2 bytes)	RING 2	BRIDGE 1	RING 7	BRIDGE 3	RING 9	BRIDGE 0
-------------------	--------	----------	--------	----------	--------	----------



**Figure 7.**  
*Routers Link Subnets Together  
for More Efficient  
Management*

## Bridges vs. Routers

One major issue confronting network managers and systems integrators is how to distinguish the need for a bridge from the need for a router. This section compares and contrasts the two types of network interoperability devices.

Bridges are ideal when two networks with different higher level protocols, but the same MAC layers need to be linked together. Bridges are relatively inexpensive and much faster than most routers. They also are much easier to install and to maintain. Once installed, bridges can automatically learn the network location of stations by listening to the source addresses of network traffic.

Bridges are not ideal, though, with large, complex networks for a variety of reasons. Because bridges pass all traffic including broadcast storms, a few NIC problems could bring down a very large internet. Also, because many bridges require a single path between networks, they lack the system fault tolerance that routers' multiple paths provide.

Because more and more networks now are running multiple protocols, a major advantage of a router is its ability to pass packets with specific protocols from one network to another. Routers using dynamic routing schemes are able to adjust to changing network conditions and provide network management functions not offered by bridges.

Another major advantage of a router over a bridge is its ability to perform packet segmentation and reassembly to accommodate intermediate networks with different packet sizes. An example of this situation might be the need to connect two Ethernet networks running NetWare via an Arcnet network running NetWare. Ethernet and Arcnet packet sizes vary considerably in size.

While they are considerably more complex and more expensive than bridges or routers, some situations might require a hybrid of the two devices called brouters. This next section examines this new network tool.

## Brouters

While definitions vary widely by vendor, in this report I define a *brouter* as a hybrid bridge and router that is able to perform both functions. It first attempts to make a routing decision, but reverts to bridge status if unable to do so. Halley Systems' ConnectLAN 202 Local Token Ring Brouter will serve as an example how a brouter operates.

In Figure 9, an IBM 9370 host uses the IBM source routing approach to communicate with an IBM Token Ring. The Token Ring is connected to another 802.5 LAN,

this time with a Novell NetWare file server using the IPX protocol; the brouter uses 802.5 MAC layer procedures to pass its data transparently between these two Token Ring networks.

The NetWare Token Ring Network does not use source routing. The brouter has the ability to recognize source routed frames and forward them per the route defined in the frame. If the packet does not have its source routing information, the brouter is able to provide the best route to the destination address. In a mixed environment, this particular brouter can coexist with other vendors' source and transparent bridges so long as it is the first and last bridge in the chain as pictured in Figure 9. Finally, the packet is then routed to an IBM Token Ring Network that does use source routing.

## How a Brouter Uses Its Routing Tables

Figure 10 illustrates a brouter in action; I use RAD Network Devices' Extended Ethernet LAN to illustrate how brouters use their routing tables. This brouter is attached to an Ethernet or IEEE 802.3 LAN like any other node. Serial links connect the brouters.

RAD's brouters use a database called LAN-table to store addresses for nodes attached to their own LAN. If an address is not detected after a certain period of time, it is deleted from this table. A second database called NET-table contains all the node addresses for the extended network in terms of particular bridges. A third database called ROUTING-table contains directions for the optimal and second-best paths for routing packets to each bridge in the network. These brouters broadcast messages periodically that update everyone on the network as to which nodes have been added, deleted, or modified.

Assume that Node A wishes to send a packet to Node B. The following steps would take place:

1. Bridge #5 uses its LAN-table and NET-table to conclude that the packet has to be forwarded to a LAN connected to Bridge #4.
2. Bridge #5 uses its ROUTING-table to find out that the best path to Bridge #4 is via L7. The packet is transmitted to L7.
3. Bridge #9 uses its ROUTING-table to determine the best way to route the packet. It sends it to L4.
4. Bridge #4 receives the packet, de-encapsulates it, and then sends it to LAN3.

Figure 8.  
Major Protocols Supported by Routers

Protocol	Source
IP	RFC 791, 1009
RIP - IP	RFC 1058
TCP	RFC 793
SNMP	RFC 1065, 1066, 1098
CMOT	RFC 1095
IPX	Novell
XNS	XSIS028112
RTMP	Apple
NBP	Apple
EP	Apple
ZIP	Apple

- If L4 has failed, Bridge #9 is aware of this fact and uses its ROUTING-table to send the packet back to L7, which is the second best path to Bridge #4.
- Bridge #5 receives the same packet it transmitted and will forward the packet using the second best direction—i.e., to L5.
- Bridge #2 to L2; Bridge #3 to L3; Bridge #4 to LAN 3 completes the sequence (courtesy of PAD Network Devices).

### Summarization of Routers and Brouters

Routers function at the Network layer of the OSI model and are protocol-specific. One function they perform is the creation of "fire walls" that prevent broadcast storms on one network from sweeping across the entire internet. Routers can be programmed to provide different classes of service and to use different routers for different types of packets.

Industry committees have developed new standards such as Point-to-Point Protocol (PPP) and Open Shortest

Path First (OSPF) that will make routing much more efficient. While routers differ widely when it comes to features, routing speed, protocols available, programmability, and security are important criteria to use when evaluating routing products.

Routers are hybrid bridges-routers. They can route certain specific protocols and then provide bridging for all other protocols. This versatility makes them very desirable on large internets.

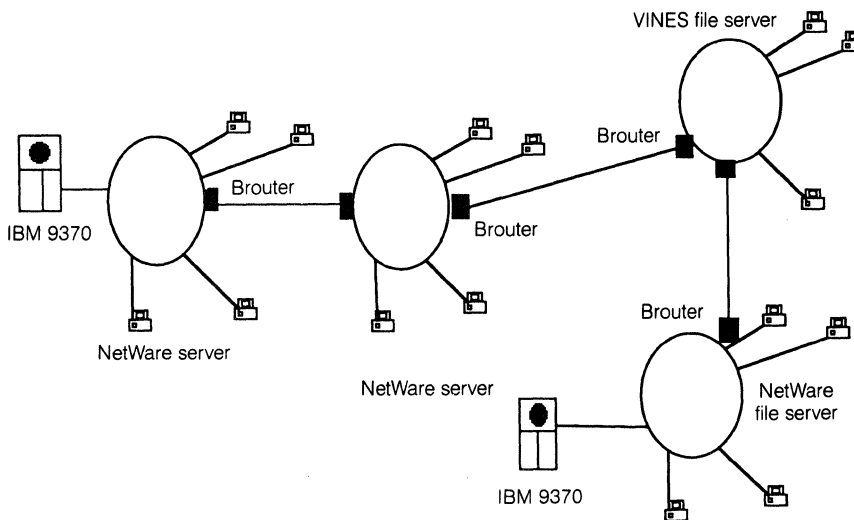
### The Level of a Gateway's Functionality Can Vary

A *gateway* is a device that connects networks which have different network architectures. Gateways use all seven layers of the OSI model and perform the protocol conversion function at the Application layer. Typical corporate gateways connect the PC world of Token Ring, Ethernet, and AppleTalk LANs with the IBM's mainframe SNA environment, with X.25 packet switched networks, or with DEC's DECnet networks.

LAN gateways differ considerably in their functionality. At the very lowest level, a gateway provides terminal emulation so that all LAN workstations can emulate or imitate a dumb terminal. Even in this case, the level of emulation can vary considerably depending upon the gateway. Some gateways handle IBM's block data transmission approach very well and map PC keyboards so that they have the look and feel of an IBM 3270 terminal while retaining the advantage of an intelligent PC workstation by permitting easy switching from terminal emulation to PC operations with a hot key. Some gateways permit a LAN workstation to window several different host sessions and move easily from session to session.

A second level of LAN gateway functionality includes file sharing between LAN and host. Novell has licensed versions of NetWare to several minicomputer manufacturers. Shortly, you should see more and more minicomputers used as NetWare file servers. NetWare for VMS is already available, so it is possible today for a NetWare LAN user to log onto a VAX from a PC workstation, and view and access NetWare files residing on the VAX. To the LAN user, the VAX appears to be just another PC file server. Because the VAX is running NetWare as one of the multitasking processes running under VMS, a VAX user could log onto the computer and exchange information with the VAX

Figure 9.  
Brouters in Operation



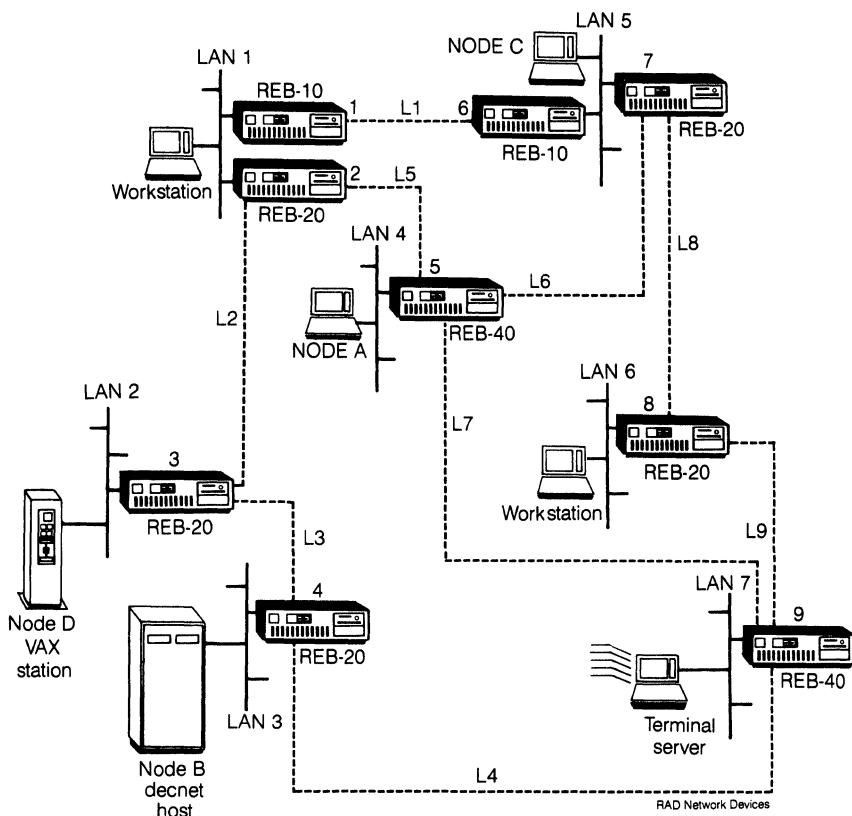


Figure 10.  
A Brouter Uses Its Tables to  
Route Packets

concerning these NetWare files using standard DEC commands. In other words, the NetWare and DECnet protocols are able to communicate with each other in a manner that is transparent to the end user.

At still a higher level of functionality, a gateway would have the ability to provide peer-to-peer communications between microcomputer programs running on the LAN and mainframe programs running on the host. This type of Client-Server relationship will become more and more important in the near future as programs are written that distribute databases among several different machines including LANs, minicomputers, and mainframes with these machines' users communicating with the programs using the same type of user interface.

Later, this report examines IBM's Systems Application Architecture (SAA), its set of specifications for providing a uniform user interface and common communication links among its whole family of computers. At this point, it is worth noting that NetWare 386 already supports IBM's SAA services. That means that when programs adhering to SAA are written, Novell's NetWare 386 operating system will be able to facilitate the type of communications required for programs to exchange meaningful information.

### How Gateways Link Hosts and LANs

Before LAN gateways, companies connected PCs directly to IBM front-end processors via coaxial cabling using expensive PC 3270 emulation cards and software. Many companies invested tens or even hundreds of thousands of dollars in products such as DCA's IRMA board. With a LAN gateway, however, the micro-mainframe connection is much more cost-effective. The gateway board emulates a cluster controller so that each network workstation is seen

by the mainframe as a terminal linked to the cluster controller. The gateway's multiple mainframe sessions are split among the network's workstations so that the channel rarely sits idle. Only the gateway needs to have a circuit card and the software necessary for protocol conversion and terminal emulation.

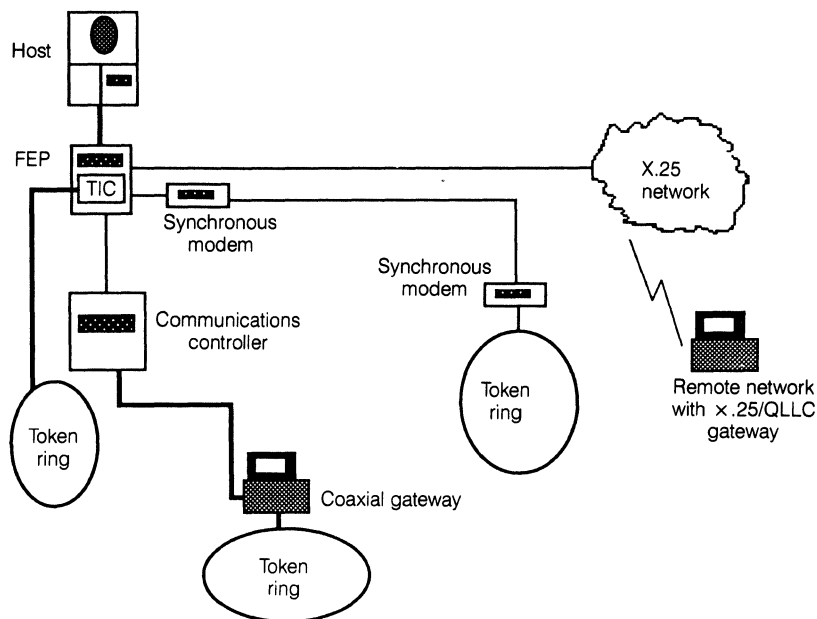
Figure 11 illustrates four different types of LAN connections to an IBM mainframe: Local connection between a host and a LAN gateway workstation linked via coaxial cable to a communications controller; a local connection between a LAN and front-end processor via a Token Ring Interface Coupler (TIC) gateway; a LAN remotely linked to a host's front-end processor via modems; and a LAN remotely linked to a mainframe's front-end processor via an X.25 network.

### Local DFT Coaxial Gateway

With earlier IBM systems, when the host computer requested transmission of a terminal's contents, the controller, not the terminal, handled this request. This approach was known as the *Control Unit Terminal* (CUT) mode of operation. With IBM's 3174 family of controllers, the company introduced the concept of offloading this terminal processing task to programmable terminals. This approach became known as the *Distributed Function Terminal* (DFT) mode of processing. A DFT coax gateway allows a LAN workstation to emulate a Distributed Function Terminal (DFT).

The gateway PC emulates an SNA cluster controller communicating with its workstations emulating DFT terminals. Workstations on the LAN run their own software that allow them to emulate an IBM terminal. This workstation's terminal emulation software accesses the gateway via the LAN's transport protocol. The number of concurrent sessions available with these gateways varies widely

**Figure 11.**  
*Gateways Between LANs and  
Mainframes*



depending upon the manufacturer. While some low-end gateways permit as few as 5 concurrent sessions, Rabbit Software's gateway can handle up to 40 such sessions. It is possible to install multiple coax gateways when traffic is too heavy for one to handle efficiently.

#### The Token Ring Interface Coupler (TIC) Gateway

The highest performance LAN gateway is the link between a Token Ring network and a host's FEP via a Token Ring Interface Coupler (TIC) gateway. The TIC permits a 4 Mbs or 16 Mbs connection depending upon the hardware installed. The key component for this connection is the Token Ring Adapter, not the NIC but an IBM FEP hardware upgrade. An IBM 3745 Token Ring Adapter comes with two 802.5 connections or TICs. High-end FEPs offer additional Token Ring Network connections. The 3745 model 130 LAK gateway offers 4 while the 3745 models 210 and 410 each offer 8 connections. A non-Token Ring LAN such as Ethernet can be linked via TIC gateway at speeds of 2.35 Mbs.

The gateway PC is viewed by the mainframe as a cluster controller; the mainframe polls this gateway PC while it in turn polls all other workstations on the Token Ring network.

#### Remote LAN Gateway

As enterprise networks and wide area networks evolve, remote LAN gateways are becoming very common. A PC on the remote site's LAN functions as a gateway and runs gateway software. This gateway PC functions as a cluster controller and communicates with a front-end processor using IBM's Synchronous Data Link Control (SDLC) protocol via synchronous modem located at both sites. Figure 12 illustrates a typical remote LAN gateway.

The limitation of remote gateways traditionally has been speed. The synchronous modem can dial up the front-end processor at speeds up to 64 Kbs. Companies with heavy micro-mainframe traffic might require multiple remote gateways to solve this congestion problem.

#### X.25 Gateways

Remote LANs can also communicate with IBM mainframes via X.25 gateways. A gateway PC with an adapter

card functions as a cluster controller and runs special gateway software that contains the QLLC protocol, an IBM defined protocol that runs over the X.25 protocol suite. The other LAN workstations emulate IBM 3270 terminals. The IBM host simply assumes it is communicating with a remote cluster controller.

#### Gateways Linking Together LANs

In enterprise networking, it is possible that several LANs might be linked to a company's mainframe, but they might not be linked to each other. When the need develops to link them together, a mainframe can act as a router to connect these LANs. Phase Systems has developed software for NetWare while Micro Tempus has similar software for LAN Manager that enables these LAN operating systems to run as an application under IBM's Virtual Telecommunications Access Method (VTAM) software on mainframes running VM or MVS.

#### 3Com Maxess SNA Gateway

3Com's Maxess SNA Gateway is a LAN SNA gateway that illustrates many of the gateway features you have been examining. The gateway workstation contains the Maxess SNA gateway coprocessor board as well as gateway software. Network workstations have far lower memory requirements because they need only run the 3270 Presentation Services and APPC transaction programs. Because applications are able to run concurrently over a single data link, one user can transfer a 3270 file while other users access programs or run LU 6.2 applications that communicate in a program-to-program mode. The program supports up to 32 concurrent sessions with each workstation able to run multiple sessions. Users have a "hot key" to switch back and forth between 3270, APPC, and DOS applications.

The Maxess SNA Gateway emulates a 3274 cluster controller while its nodes use 3278/9 terminal emulation and have the ability to define their keyboards and screens. It supports IBM IND\$FILE file transfer support up to speeds of 64 Kbs.

The program is intelligent enough to be able to provide both dynamic and static LU allocation. It generates alerts

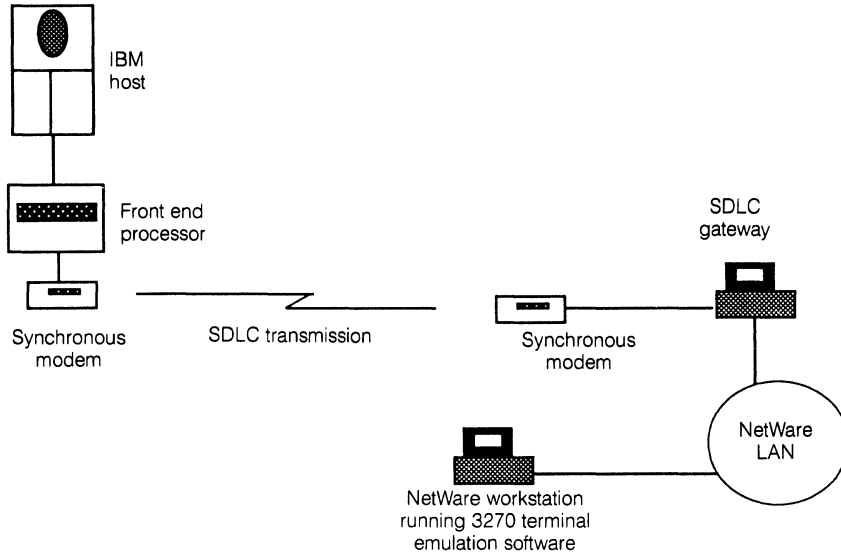


Figure 12.  
A Remote LAN

when a session is not established due to a data link failure or SNA protocol violation. A Response Time Monitor accumulates information concerning the time required for a host and network to respond to terminal users. This gateway can be monitored by IBM's NetView network management program.

### Apple Gateways

Many Fortune 1000 companies have significant numbers of Macintosh computers which they have begun to network, usually with AppleTalk or Ethernet. It is not difficult for a Macintosh on an Ethernet network to be bridged to a PC-based Ethernet LAN. What about linking the Macs to the company's IBM mainframe or perhaps to a department's VAX system? This section examines some of the Apple gateways available to the network manager or systems integrator.

#### Apple Gateways to IBM Mainframes Via PC LANs

In an enterprise network, Macintosh workstations might be linked together on an AppleTalk network using these workstations' built-in LocalTalk interfaces. Because AppleTalk's suite of protocols now supports Token Ring networks with the TokenTalk protocol, Figure 13 illustrates how it is possible to link an AppleTalk network to a Token Ring network with a TokenTalk NIC serving as a bridge. The Macintosh workstations then can use the Token Ring network's gateway to the IBM mainframe.

#### Direct AppleTalk Gateways to IBM Mainframes

AppleTalk LANs can provide gateways to IBM mainframes in a variety of different ways. These gateways all emulate a 3174 Cluster controller and distribute and manage 3270 terminal emulation sessions, but they do so in quite different ways.

DCA's MacIrmaLAN Gateway Server uses a dedicated PC or PS/2-based gateway to run either a LocalTalk PC card or an IBM Token Ring card plus AppleShare PC, IrmaLAN, and MacIrmaTalk software. Each Mac workstation runs MacIrma Workstation client software and MacIrmaTalk Startup software. Companies with a wide range of LANs including AppleTalk requiring gateways to a host might want to consider Tri-Data's Netway gateways. A Windows-based 3270 gateway supports Macintosh workstations running AppleTalk and PCs running NetWare concurrently. Tri-Data's NetWay products already support the installation of both Ethernet and Token Ring cards concurrently.

#### AppleTalk Gateways to IBM AS/400 Minicomputers

Andrew KMW Systems is the first vendor to offer an AppleTalk gateway to IBM's popular AS/400 minicomputer. This twin axial gateway supports up to a maximum of six emulation cards in a Mac II gateway. As Figure 14 indicates, Macintosh AppleTalk workstations are linked to a Macintosh AS/400 gateway, which is directly connected to

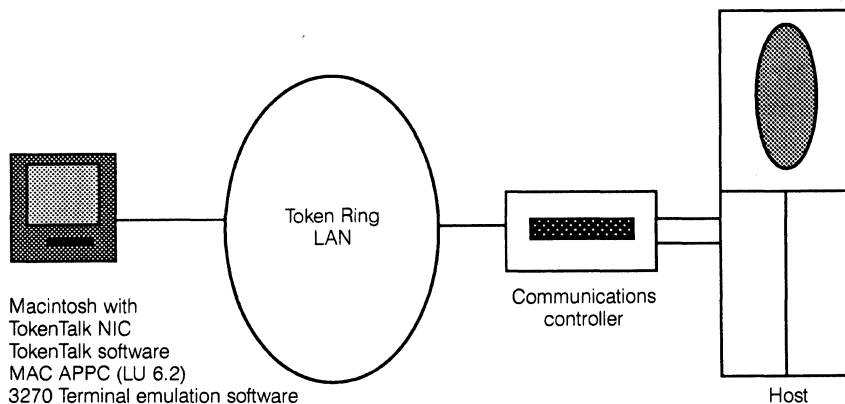
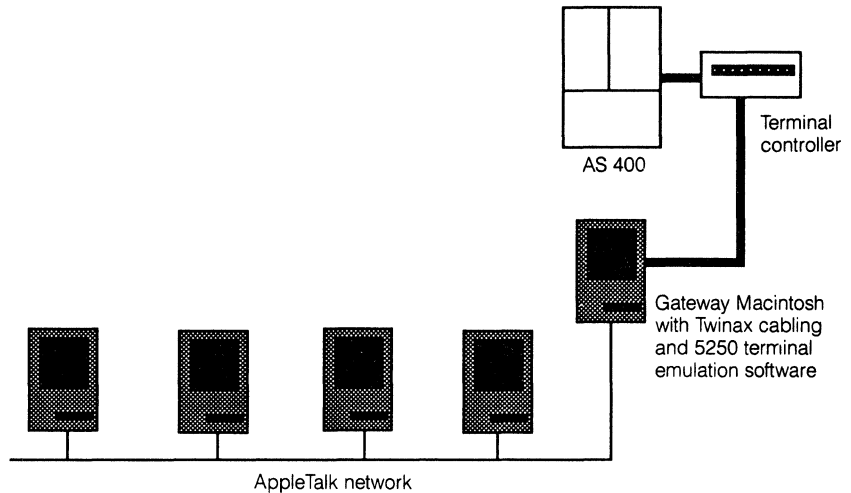


Figure 13.  
AppleTalk's SNA Gateway Via a Token Ring Network

Figure 14.  
Connecting an Apple Talk  
LAN to an AS/400



a terminal controller. The controller is either directly connected or connected via 9600 bps synchronous modems to the AS/400.

### MacAPPC

Apple has developed MacAPPC, a set of programming tools that should result in more efficient Mac-to-host communications in the future. Programs written using APPCs could exchange information and pool processing power regardless of whether they were running on microcomputers, minicomputers, or hosts. MacAPPC implements SNA's LU6.2 and PU2. 1, NAUs required for peer-to-peer communications under SNA. The problem with MacAPPC, though, is that it probably will be a while before you see a significant number of host programs using APPC that can interface with Mac programs that use APPC. Because of the expense of host-based programs, companies will be slow to dump these programs if they are performing adequately.

Apple also has developed MacWorkStation, a tool that lets host computer users utilize the Macintosh interface, filing, and printing features. Programmers access the Macintosh toolbox, including control over windows, pull-down menus, dialog boxes, and other features of the Macintosh user interface without having to write Macintosh applications. A company using MacWorkStation could develop applications so that its mainframe users would have a Macintosh interface.

### Linking Macs to the VAX World

Digital Equipment Corporation's LanWorks for the Macintosh integrates Macintosh computers on an AppleTalk network with DEC's VAXes and DECnet/OSI network. VAXshare, the DEC file server program, looks like AppleShare on a Macintosh. DEC has agreed to change the name of the LanWorks product because it is already owned by another corporation. At the time this report is being written, the new name has not yet been determined, so I will continue to refer to this connectivity tool by its original name. This user interface makes it easy for Macintosh users to manipulate the data coming from the VAX.

LanWorks links the Macintosh micro world with the VAX world via DEC's Mailbus enterprise messaging system and All-in-1 Mail. The electronic mail package supports the CCITT's X.400 family of protocols, which means that this mail service could provide an indirect means of sending mail from an AppleTalk network to IBM's PROFS and SNADS.

Under LanWorks, the VMS server software includes VAXshare File Services and VAXshare Print Services. The File services software enables VAX computers to serve as file servers for Macintosh clients and is Apple File Protocol based. The print services package enables Macintosh and VAX users to share any DEC or Apple networked printer.

DEC has announced several related products that can be used in conjunction with LanWorks. Macintosh workstations appear to the VAX computer as Phase IV non-routing end nodes. This means that the VAX can route packets from Macintosh workstations on Ethernet networks by encapsulating them within DECnet packets, just as it would encapsulate information from DEC workstations. DECnet management software can be used to monitor the AppleTalk communications coming from the Macintosh workstations.

### Summarization of IBM LAN Gateways

LAN gateways to IBM hosts come in a number of different types, including local DFT coaxial connection, direct connection via Token Interface Coupler (TIC), remote link via modem, and X.25 connection. Mainframes can also be used as routers to link two LANs that have mainframe gateways.

In order to understand mainframe gateways, you must know something about IBM's Systems Network Architecture (SNA), its layered network architecture. Historically this has been a master/slave system with all communications initiated by the host. IBM's LU 6.2 and APPC make it possible for peer-to-peer communications in the future between LAN programs and mainframe programs.

IBM's Systems Application Architecture (SAA) is a set of specifications for consistent user interfaces and program interfaces for its entire computer family. The release of SAA-compliant services on NetWare 386 and a set of SAA tools for the Macintosh will facilitate this communication between LAK program and mainframe program.

AppleTalk LANs can communicate with IBM hosts via their own gateways or by connecting first to a Token Ring Network via TokenTalk and then using this LAN's gateway. AppleTalk LANs also can communicate with DEC VAX networks via several products developed by Apple and DEC. These products incorporate AppleTalk's suite of protocols in DECnet so that an AppleTalk user of a DEC computer's services will view a screen resembling that of AppleShare. ■



# An Overview of Communications Servers

## In this report:

Today's Communication Servers .....	2
How Communication Servers Operate .....	3
Digital Private Branch Exchanges .....	6

## Datapro Summary

Although most networking equipment directly interfaces to Token Ring and Ethernet networks, there are some devices unable to link directly to networks. Communications servers can help because they are relatively inexpensive and are used to link asynchronous systems to a LAN. The following is an overview of communications servers (which are also regarded as asynchronous multiplexors) with a historical perspective on LANs and the technical characteristics of these devices.

## The Evolution of LANs

Problems often arise in a network environment when a device does not support a direct network attachment. Asynchronous terminals such as VT100s, 200s, and 300s, asynchronous modems and serial printers, all do not support an attachment to any kind of network. During the early 1980s when LANs were entering the commercial world, most computer devices did not directly support networks.

Throughout the 1960s and 1970s, as businesses expanded, the computer became a necessity in most businesses. For the data communications department of any business (with the exception of IBM customers), most employed one of the oldest and most prevalent standards known as the RS-232-C specification for connection of serial devices. This specification is a standard from the Electronic Industries Association (EIA). Terminals were directly connected to hosts using this standard. RS-232-C devices, such as terminals, modems, and printers, constituted a major part of the investment that businesses had made in data

processing equipment. Most businesses, large and small, have spent billions of dollars acquiring equipment with this type of interface. Refer to Figure 1.

During this time, each connection to the host required a cable to be run to each terminal. This represented two problems:

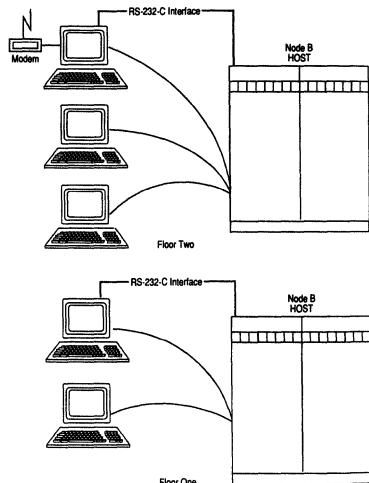
1. *Cost:* This tends to be a major consideration in installing and maintaining a computer system. A large percentage of the cost of installing a large network can be expenditure for labor for cable companies pulling cable.
2. *Multiple-host access:* What if the user wanted to access multiple hosts or decided to move the business office from one floor to another? Before communication servers and LANs, new cable was generally pulled to the new location. For accessing multiple hosts, a business would commonly use A-B switches, or pull multiple cables to the same terminal (a very high cost) or might use a digital private branch exchange (DPBX, to be explained at the end of this report; see also Figure 2).

Cable cost was not the only reason for businesses to incorporate LANs into their operation. Two other factors are also involved:

1. Modems used to be connected directly to the user's terminal or PC. Or even

This Datapro report is a reprint of "Appendix G: Communication Servers," pp. 241-254, from *Local Area Networking* by Matthew G. Naugle. Copyright © 1991 by McGraw-Hill, Inc. Reprinted with permission.

Figure 1.  
Terminal-Host Connection Without Communications Servers



This presented many problems for users. A user would submit a print job to the host, then would have to leave the office, walk down to the computer room, and pick up the printout. While the low cost of dot-matrix printers have eliminated some of these problems, a majority of the printers are still connected directly to the host computer.

**Communication Servers**

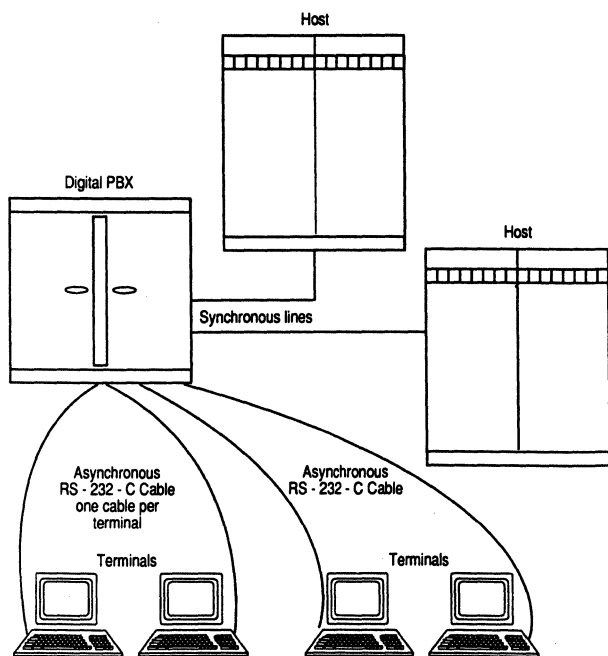
Commercial businesses were enthused by LANs, but all the equipment mentioned above had to be able to connect to the LAN. Hence, there was a special device developed for those pieces of equipment that allowed a connection to a network. It is known as a *communication server* (also known as terminal server or comm server). Communication servers probably represent one of the oldest techniques used to connect any device to a network. Their original purpose was to connect asynchronous hosts and terminals across a LAN.

A communication server can be regarded as an asynchronous multiplexor. On one end of the communication server are RS-232-C ports that connect to asynchronous devices such as terminals, modems, printers, hosts, and even a personal computer's comm ports. On the other end is the attachment interface to the network (Ethernet or Token Ring). Refer to Figure 3.

Some still consider this device as a gateway. Communication servers are not gateways, for gateways take input from one protocol and convert the input into another form for output. Conversely, a communication server takes the input data, encapsulates (does not convert) it into a network packet, and then sends that packet to the network. When the packet arrives at the destination device, the destination device strips the network information off the packet and submits the data to the receiving device in its original form.

In the early 1980s, communication servers were also known as "milking machines." Looking at Figure 4, we see that the connection from the host to the comm server does resemble a milking machine. While a connection to a host such as this had many advantages early on, today it represents many disadvantages, particularly hardware interrupts on the host device. RS-232-C devices interrupt the host processor via a hardware interrupt. Many hardware interrupts on a mini or mainframe host will slowly consume many of the host's processing cycles, thereby reducing the efficiency of the host.

Figure 2.  
DPBX Connections



worse, the company would not buy all their personnel a modem, and special terminals were set up so that users had to walk over to the terminal to use the modem. This represented a major expense for companies in that only one user had access to the modem at a time and a company with many employees had to buy many modems.

2. Printers were usually located within 50 ft. of the host. (This is because the specification for connection of asynchronous devices states that a connection run between the two devices shall be no more than 50 ft. end to end; parallel printers were located within 25 ft.)

**Today's Communication Servers**

Communication servers today are still attaching to hosts, terminals, printers, and other serial devices, but their use is taking a different turn. With most terminals being replaced by personal computers, and mini and mainframe computers having built-in network controllers, communication servers are providing communications connectivity in a different manner. They are being used in three basic areas. One predominant use for them is for modem and printer pools.

Modem pools are groups of modems clustered together for use by anyone who has access to the comm server. By connecting modems directly to a communication server,

Figure 3. Anatomy of a Communication Server

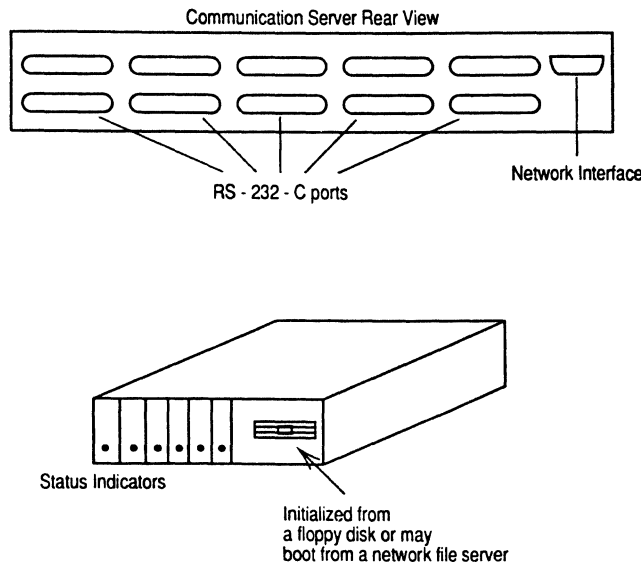
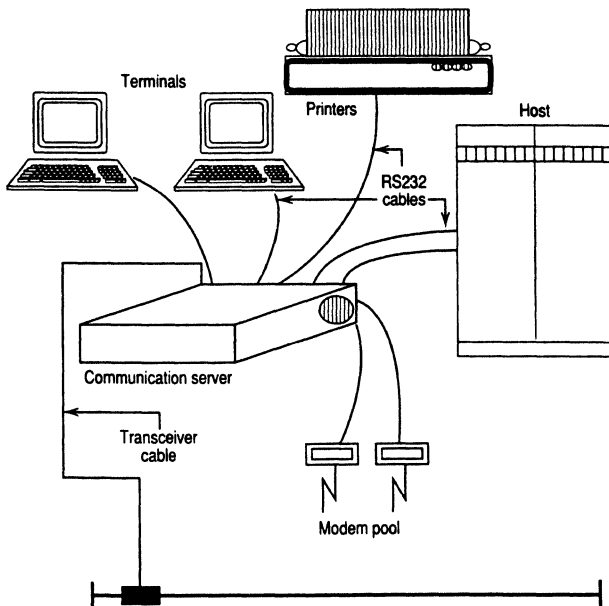


Figure 4. Different Devices Attached to One Communication Server



any user with access to the network may have access to one of the modems (provided that no user has already established a connection to that modem port on the communication server). Instead of a company buying one modem for each user on the network, they can buy a few modems, connect them to a communication server, and then have the users access the modems through the network. This is extremely efficient, for few modems were used 100% of the time. With a connection to the comm server, modems are now accessed by multiple users, thereby increasing the use of the modem. Even if the modem port on the communication server is busy, most networking companies have developed their servers to queue requests to a particular communication server port (see Figure 5.)

For printers, communication servers can provide remote printing capabilities. For example, (referring to Figure 6) a communication server may be connected to the host computer. Another communication server may be placed somewhere on the network. This communication server may have a printer attached to it. A virtual circuit (session) will be established between the two servers. The print job that is submitted on the host is printed to the communication server and that server will send the data over the network to the communication server with the printer attached.

Some hosts with an internal network controller (direct attachment to the LAN) have the capability to send the data over the network without the communication server attached to it.

### How Communication Servers Operate

Let's explore exactly how communication servers work. Communication servers are devices that allow any piece of data communications equipment supporting the RS-232-C standard (terminals, modems, printers, hosts, and personal computers) to attach to a network.

Let's say you are sitting at a terminal device (also known as your local device) and you wish to connect to some resource out on the network, say, a host (also known as the remote device). It does not matter what the operating system of the remote host is running. Communication servers have their own internal operating system that enables the data that you transfer from your local device to the remote device to look the same. Referring to Figures 5 and 6, you first issue some type of connect command. This command enables the communication server to connect to a remote device.

Once the connection is made between the local and remote devices, the user will never know that they are running through a communication server. The communication server from that point on is used only as a vehicle to transport the user's data to the host or remote device and to accept data from the host or the remote device. It will appear as if the users are directly connected to the host computer modem, etc.

Depending on the protocol suite, communication servers can establish connections to other communication servers or communication servers can establish connections directly to a host, provided the host has some type of internal networking capabilities.

Communication servers cannot only establish connections to devices on the local LAN but also traverse through bridges, routers, and other network-extending devices.

Therefore, if your site in New Orleans wishes to connect to a host in Washington, DC, the same "connect" command issued for connection to a local device will establish

a connection to the remote device. Some communication servers support scripting so that menu systems can be built so that users can pick a menu item and the connection will be established for them. 3Com communication servers offer an extensive customizable script menu system. They also support multiple protocol suites such as TCP, XNS, OSI, and LAT. Remember communication servers are independent devices, and it does not matter what the operating environment (vendor independent) they are placed in.

Communication servers have two types of connection points. One is the connection to the network. The other is the ports that are connected to the serial device. The number of ports available is dependent on the networking vendor. The number of ports ranges from 2 to 128, depending on the user's individual needs and what their networking vendor can provide. Refer to Figure 3.

Configuring a communication server port so that it communicates correctly with a terminal, modem, or printer is the same as configuring the port as if it were to be connected directly to the host. All we are doing with a communication server is functionally bringing the host's ports to the user instead of running a terminal cable to the host. The communication server acts as an interface between two devices and a LAN.

During the configuration process, some type of script will ask a series of questions enabling you to configure a port for the particular device that you wish to connect to it. Whatever the asynchronous device is configured for, the port on the terminal server must exactly match it.

The network port cannot be configured. It will attach to the network cable plant that you have installed. Each asynchronous port is usually individually configured. Some configuration parameters you should be familiar with are baud rate, parity, data bits, and flow control.

In the asynchronous world, you will hear terms such as baud rate (for the purposes of this report, baud and bits per second (bps) will mean the same thing), male-female connectors, data communications-terminating equipment

Figure 5.

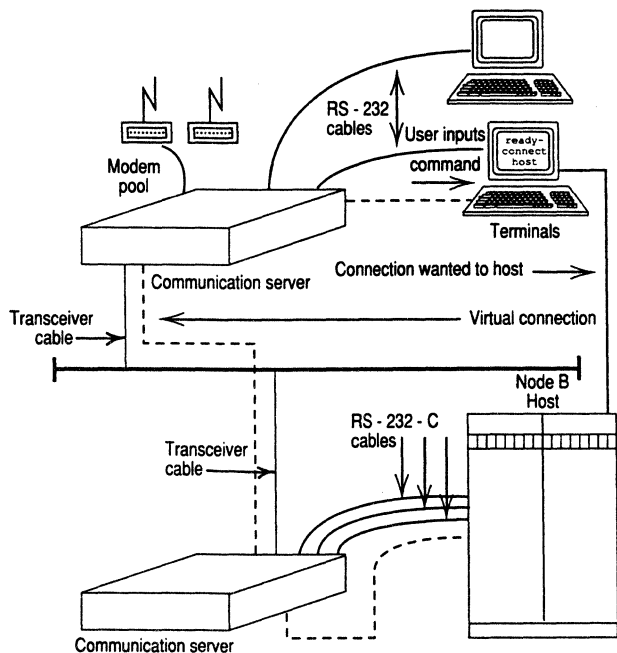
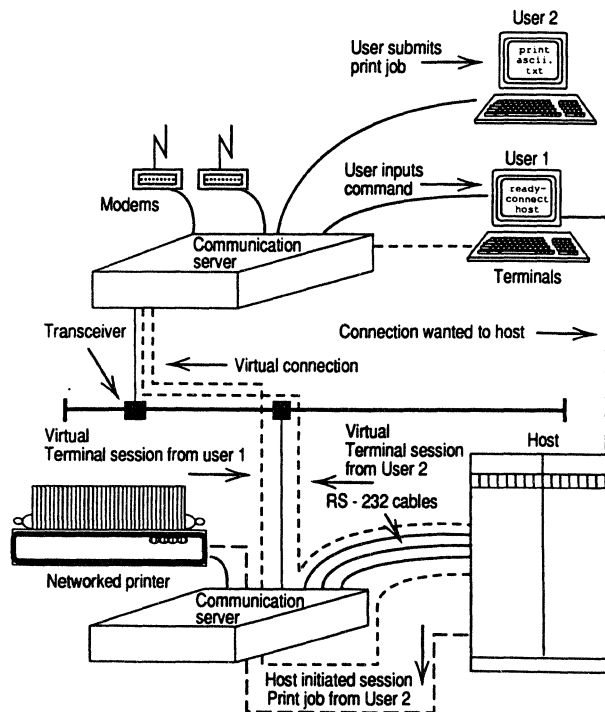
**Terminal-Host Session Establishment**

Figure 6.

**Multiple Communication Server Sessions**

(DCE), data terminal equipment (DTE), hardware-software flow control, and communication server aggregate throughput. All of these terms are used to define each port on the communication server.

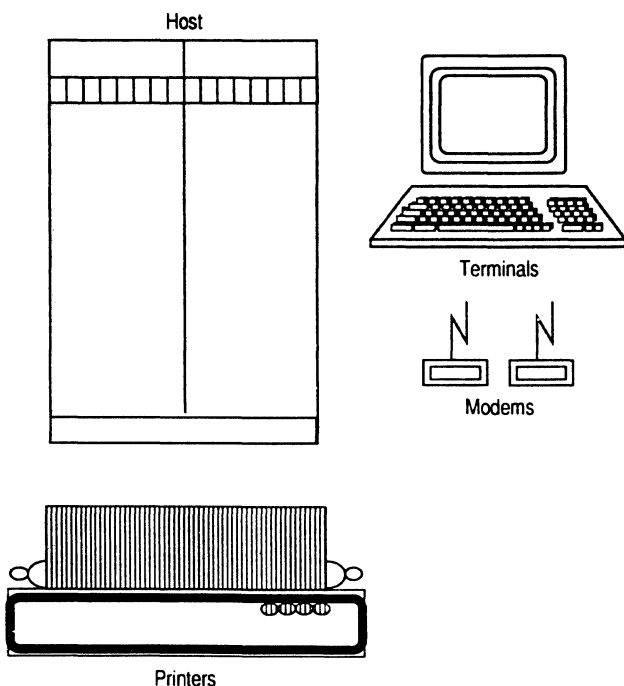
The baud rate is the speed of the port. Most ports on a communication server support 75 to 38.4 kbps. This is on a per port basis. One port may be configured to handle 9600 baud for terminal connection, while another port can be configured to handle 2400 baud for a modem attachment.

One point to emphasize here is that the baud rate does not have to be the same on each end of the connection. For example, if your terminal is running at 9600 baud and you want to connect to a comm server that has a modem attached to it, the modem does not have to be at the same baud rate. The modem could be running at 2400 baud. The communication server will take care of the speed mismatch. The parameters at each port must match the device that is to be directly attached to it. Most comm servers now support something called *autobaud*. If the comm server supports autobaud, it will figure out the speed at which you are running. The advantages to this are many. For example, every time you move a terminal, the user will not have to reconfigure the port. Some comm servers can be configured only by someone who has the privilege to change the parameters.

The data bits indicate how many bits will be used to represent a character. This field ranges from 5 to 8 bits per character. Five bits per character were used in very old teletype systems. Most characters are now represented by 7 or 8 bits.

Characters are represented as a series of 7 bits. For the extended character set, 8 bits is used. A parity bit is added to the data to ensure the integrity of the data. The parity is used for checking errors on the asynchronous bits. Parity types are odd, even, mark, space, and none (no parity

Figure 7.  
Devices That Communication Servers Attach to a LAN



added). The most common type is 8 data bits with no parity. Some comm servers have a capability known as auto-baud (mentioned above) and *autoparity*. This allows a comm server port to determine the baud and parity setting without the administrator specifying it.

Another common configuration is the null-modem cable. Asynchronous devices are known as either data terminal equipment (DTE) or data circuit-terminating equipment (DCE). This terminology specifies the interface of the equipment. For example, an asynchronous terminal is defined as DTE. By specifying this, we know that the equipment transmits on a certain wire and receives data on another wire. It will also determine what special signals this equipment will set up. DTE devices usually initiate sessions, and DCE devices usually accept connections.

A modem is known as a DCE-type device. This specifies that it, too, will transmit on a certain wire and receive on another. According to the way it is set up, a DCE device will transmit on the same wire that a DTE device will receive on. The DCE device will receive on the same wire that a DTE device will transmit on. DCE devices will also have certain signals to present to the DTE device about its status.

This DCE-DTE connection is specified to allow devices to talk to each other. If a DCE device were connected to a modem (which is a DCE-type device), neither device would be able to communicate with one another (both will try to transmit and receive on the same wires). When would this type of connection exist? The ports on the comm server are DCE-type devices. Instead of changing the connector interface for the device to be attached to the comm server, a special cable is used to connect these devices together. It is called a null-modem cable. This cable

will reverse its signals by simply moving the wires around to their respective positions. This is the most common problem found when hooking a modem up to a comm server. This cable fools each device into thinking that it is connected to a DTE device. A terminal is a DTE device and is connected to a comm server using a straight through cable. Hosts can either be a DCE or DTE.

Each communication server port has the capability of handling multiple sessions from the same physical port. This enables the user to establish one remote connection, put that connection on hold, establish a connection to another remote device, and so forth. 3Com terminal servers allow up to eight sessions per physical port. An analogy would once again be the phone system. The common office phone has the capability to make one call, put that call on hold, and make another phone call. The comm server acts the same way.

A communication server port may be in one of three states: command mode, data transfer mode, or listen mode. *Command mode* indicates that a device is attached and is not connected to any other device. If this device is a terminal, the command processor of the terminal server will prompt the user for input. These commands may be to connect to a remote device, change some of the parameters on the port, or show some status configuration. The command mode state means that the port is alive and is waiting for input from the user.

*Data transfer* or *connected mode* indicates that a connection has been established between two devices and data may be transferred from one device to the other. *Listen mode* indicates that the port is available, as is waiting connection from a remote device, for example a modem port waiting for a connection from a terminal.

One other point to note is the type of terminal emulation that is required. Network designers should be aware that most comm servers emulate simple ASCII terminals such as the VT100 series if they emulate at all. Some comm servers do have the capability to attach to specific host interfaces. The network designer should investigate thoroughly the host connection and ensure that the comm server has the capability to emulate that type of terminal.

### Multi-Purpose Devices

We know that communication servers can support terminals, modems, hosts, printers, and virtually any RS-232-C device (refer to Figures 7 and 4), but the unique quality of these servers is that they can support all these devices on a single communication server. In other words, you do not have to buy one communication server for printers, another for modems, another for the hosts, and so forth. All the devices mentioned above can be connected on one communication server as shown in Figure 4. On larger networks, most network designers will separate their communication servers as to the application; one for the host, one for the terminals, and one for the modems. This is for convenience management only and is not a characteristic of the communication server.

If you want to extend the connection between the local device and the communication server, there is a standard to support this: the RS-422/449 connection. This standard was intended to supersede the RS-232-C standard but to date has failed to do so. This standard supports connections between devices of up to 1000 ft. Some comm servers support this standard. 3Com CS/1's support it.

And that is what communication servers provide. They are inexpensive and provide a great service for those devices that cannot connect directly to a network.

Communication servers are used to connect asynchronous devices to LAN. This does not mean that a LAN may only control communication servers or that they should be separated from those devices that can attach directly to the LAN. With the correct network operating system (TCP/IP), for example, communication servers and the directly attached network stations may communicate with each other. In any case both reside on the same LAN harmoniously.

### Digital Private Branch Exchanges

Since communication servers are commonly compared to digital private branch exchanges (DPBXs), a comparison between the two is in order. It is commonly thought that digital private branch exchanges are primarily used for telephone systems, but that isn't necessarily true. Digital private branch exchanges offered many of the same functional characteristics of terminal servers in the data communications environment in the 1970s, 1980s, and even today.

Private branch exchanges (PBXs) were first developed to connect calls between parties on the same premises and to switch calls to facilities outside the premises through the public telephone network. Prior to the PBX, every telephone call that was placed resulted in the call being placed through the outside telephone system by using a local switching office of the public telephone system. An increased demand on the telephone system soon produced a burden on the public telephone system, and PBXs were developed.

Prior to 1968, the only attachment to the public telephone system had to be accomplished through AT&T. The Carterfone decision of 1968 from the FCC changed all that. The PBX system grew rapidly as more vendors entered the market. The PBX was redesigned to accept not only the analog devices of telephone systems but also, eventually, to accept digital data, and the digital PBX was invented.

Prior to the evolution of LANs in the early 1980s, DPBXs offered many attractive benefits for terminals desiring access or connection to a local host processor and the ability to switch between host processor and other devices such as modems (see Figure 2).

As Figure 2 shows, a connection to a DPBX is accomplished by running a cable between the device and the DPBX. Another connection is made to one or more hosts. (Sounds similar to the communication server technology, doesn't it?) The DPBX multiplexed connections from the terminals to the host(s).

DPBXs are centralized switches that allow multiple-host connections and terminal connections. A connection was made between a terminal and host by electrically switching the circuit to reach the host port. This is known as circuit switching. Networks are packet switching. This allowed the user access to one host and multiple hosts. Connections from the DPBX to the hosts were usually made via a synchronous connection at speeds of 64 kbps. Because of this link, digital DPBXs are always located near the host processor to which they are connected. Unlike a communication server attached to LAN, there is no distribution of PBXs throughout a site. DPBX capabilities are limited though when compared to a terminal server.

These disadvantages include a slow-speed link between the DPBX and the host processor. Terminal cables still had to be run for long distances, for the DPBX was still located near the host processor.

DPBXs represent a centralized switch. If the DPBX went down, many users on the DPBX would be idle until the DPBX came back on line. However, the whole DPBX seldom failed. Usually the individual cards in the DPBX would fail, causing only those users connected to that card to become idle.

Even with the limitations of the DPBXs, this was impressive during the 1970s and early 1980s. With the emergence of new technology in the 1980s (Ethernet), having a speed of 10 Mbps and enhancing the same features of a DPBX, while distributing the servers over the network, resulted in the demise of the digital PBX, at least for the purposes of interconnecting computers. ■

# An Overview of Network Management

## In this report:

Facets of Network Management .....	2
OSI View of Network Management .....	5
Integrated Network Management System .....	7
Unified Network Management Architecture (UNMA) .....	9

## Datapro Summary

The complexity of today's networking environments makes them difficult to manage. The situation appears to be intensifying as networks and associated technologies proliferate through out organizations. According to one research group, 75% of 300 Fortune 1000 companies surveyed plan to increase their network management budget by at least 48 percent in the next year. This illustrates the need for network management, and just how willing users are to pay for solutions.

## Executive Summary

Network management is necessary to provide an adequate quality of network service in terms of bandwidth, transmission delay, error rate, user transparency, and reliability, as well as network adaptability to ever-changing technology and user needs without service disruption.

There are three facets of network management: organizational, technical, and functional. The main issue of the organizational facet is how to organize a network management process. The main issue of the technical facet is how to cope with difficulties of managing extremely complex, heterogeneous, multivendor internetworks. From the functional standpoint, network management is a set of operational, administrative, and support tools that provide:

- Fault management
- Configuration and name management
- Performance management
- Accounting management
- Security management

This Datapro report is a reprint of Chapter 9, "Network Management," pp. 199-220, from *Strategic Information Systems* by Henry Eric Firdman. Copyright © 1991 by McGraw-Hill, Inc. Reprinted with permission.

The International Standards Organization (ISO) looks at network management as an extension of the basic Open Systems Interconnection (OSI) Reference Model. Unfortunately, OSI network management is more a promise than a reality. In the meantime several vendors have been addressing one of the most demanding problems of the corporate information infrastructure: integration of many heterogeneous structures into an integrated network management system (INMS). Several INMS products have been available for some time, with IBM's Netview and AT&T's Unified Network Management Architecture (UNMA) being the most popular, at least in the Fortune 1000 community.

Here are some of the major concepts that are or will be commonly incorporated in INMSs:

- Addressing mostly the part of the network management system (NMS) integration that deals with connectivity
- Providing a consistent and transparent view of the INMS's total network management capabilities as if they were provided by a single system
- Accommodating tiered integration architectures that permit hierarchical network management authorities

- Supporting management services, such as configuration and fault management
- Allowing users to build their own management applications that drive, or create presentations from, their individual NMSs
- Using a modular approach to accommodate unforeseen changes in environment, user needs, NMS relationships, and so forth.

Even in an ideal OSI environment, there would be very few people who could understand in real time what is going on in a large internetwork, or interpret on-line the stream of management information provided by the INMS or individual NMSs, or just support normal internetwork operations around the clock. The solution to these problems is applying artificial intelligence (AI) technology.

In the next several years, AI will not only make integrated network management a reality, but it will also help integrate network and information management, providing the glue necessary for putting all pieces of the corporate information infrastructure together.

After the internetwork is designed and implemented, it should be managed. Network management provides the following:

- Adequate quality of network service in terms of bandwidth, transmission delay, error rate, user transparency, and reliability
- Network adaptability to ever-changing technology and user needs without service disruption

For many years network management has been provided by carriers and major computer vendors. However, with the advent of PCs, workstations, and LANs, as well as the proliferation of a multivendor environment, such a simplistic approach is becoming impractical for all but very simple internetworks. The fact is that users must take care of their internetworks or hire system integrators to manage network functions.

For most users network management is an extremely complex problem, and they seem to be ready to pay for its solution. According to recent research by Business Research Group, more than 75% of 300 Fortune 1000 companies intend to increase their budgets for network management by 48% from 1990 to 1992. In other research by Index Group, Inc., network management has been identified as an area of major concern for both IS and telecommunication managers.

The ideal solution to network management problems is seen by users as so-called *end-to-end network management*. This method consists of a single internetwork operator console from which any component of the internetwork, including both subnet and host components, may be operated, troubleshoot, controlled, and reconfigured remotely. End-to-end network management also implies significant involvement of on-site knowledge-based systems (KBSs) that provide assistance to network management personnel and keep the internetwork operator console from being overloaded with unnecessary details of internetwork operation.

The idea of end-to-end network management is far from implementation. Estimates for making it a reality range from four to ten years. In the meantime, internetworks have to be managed by using available management tools.

## Facets of Network Management

Unfortunately, network management problems cannot be solved by purchasing more hardware or developing more software. The situation is complicated by the fact that network management is not just a technical problem. Rather, it comprises three separate but highly interrelated facets: organizational, technical, and functional. Let us look at these facets of network management in more detail.

### Organizational Facet

The main issue of the organizational facet is how to organize a network management process. This issue involves addressing and solving problems such as the following:

- Where to locate points of internetwork state control
- How to collect information about internetwork state and performance and make it accessible to management
- How to determine the minimum information that should be available to management to understand what is going on in the internetwork and how to deliver this information
- How to allocate network management personnel

From the organizational standpoint, network management should be considered in the context of the corporate information infrastructure that becomes increasingly decentralized, providing greater end-user freedom and information sharing and exchange. Managing an internetwork may mean managing hundreds of LANs, bridges, routers, gateways, backbones, and WANs that may be geographically remote and heterogeneous, reside in the subnet or the host, and be owned by different organizations within and outside the enterprise.

When corporate divisions and departments make networks an intrinsic part of their fiefdoms, control over the overall internetwork performance becomes more and more difficult. The situation is complicated by technical problems of isolating a problem in the internetwork to the customer or carrier equipment, or to a subnet or host. As a result, the trend is shifting from local to centralized network management.

It comes as almost a paradox that a more decentralized information infrastructure demands more centralized network management. Several important factors contribute to this paradox, making centralized network management a necessity:

- The entire enterprise relies more and more on its information infrastructure: hence, internetwork.
- Skilled network management personnel are hard to find and keep at each local site.
- No significant network management budget surge can be expected in the near future.
- Decentralized network management can hardly provide anything close to optimal internetwork utilization.
- The number of stations, such as PCs, workstations, and servers, that have to be networked rapidly increases.

The push for centralized network management is extremely strong, especially among IS managers. Centralized network management should provide a corporate-wide view of what is going on across the internetwork. The following operations are most often cited by IS managers as necessary to be carried out in centralized fashion:



- Access to a complex and constantly changing LAN/host network
- Utilization control and dynamic resource allocation
- Single-console monitoring and troubleshooting
- Traffic pattern analysis
- Performance measurement of heterogeneous networks for accounting and cost-justification purposes

As a result many IS managers want to remove operational autonomy from end-user divisions and departments and, sometimes, even from telecommunication departments. Users usually resist losing control of their network, and the conflict between IS and end-user organizations seems inevitable. In some cases top management resolves the conflict in a simple way: by outsourcing network management.

Indeed, the reasons for which IS managers push centralized network management are not entirely altruistic. For them, centralization is an excellent chance to regain control over the computational resources (or what we call the corporate information infrastructure) they lost with the advent of PCs. Regaining this control is essential for the good of the enterprise. Building and managing the corporate information infrastructure is the business of IS organizations for the 1990s, and network management is obviously included.

Centralized network management, however, is not a panacea. Table 1 summarizes its advantages and disadvantages. When deciding on a network management organization, the IS manager must analyze these advantages and disadvantages; assign different weights to them, depending on business requirements; and select the most appropriate degree of network management centralization.

**Table 1. Centralized Network Management Advantages and Disadvantages**

Advantages	Disadvantages
Concentrates scarce maintenance skills at one site	Most faults occur at local sites (workstations, system and application software, etc.).
Standards and interfaces are easier to enforce and maintain	For nonvoice communications, maintenance is actually end-user support rather than central engineering.
Security is easier to provide and enforce.	Frequent changes in traffic patterns and workstation locations are hard to handle.
Control over internetwork performance and utilization tighter	Responsiveness to end-user needs is worse.

At any rate, network management calls for reconsidering the corporate organizational structure. Interrelationships among the IS department, the central telecommunication department, and on-site network management personnel have to be rectified; and the reporting structure has to be in accordance with the requirements for network management.

By all indications the new organizational structure will hardly be more robust than the old one. Some of the proposals to establish a matrix management structure, with

some employees operating networks on-site but reporting to the "central authority," seem to confirm this grim prediction.<sup>1</sup>

One interesting approach to resolving the trade-offs between centralized and local network management is *network partitioning*, with the corresponding changes in the corporate organizational structure. According to this idea the whole internetwork can be divided into an *access network* and a *transport network*.

The access network is building- or site-located and typically a LAN. It may be linked to other access networks by bridges. Each access network should have a manager who is responsible for user services and network operations, growth, and maintenance.

The transport network consists of geographically remote networks connected through a backbone. It is typically a WAN linked to other WANs (including public packet-switching networks) through routers or gateways. The transport network should be run by a central group responsible for network operation and maintenance, all-corporate security, and the issuance of guidelines for access network managers.

### Technical Facet

The evolution of computer and telecommunication technology has led to the development of extremely complex, heterogeneous, multivendor internetworks. As their complexity grows, so do the technical difficulties of managing these internetworks.

Managing an internetwork means controlling the performance and quality of service of hundreds of LANs, bridges, routers, gateways, Tx ( $x = 1, 2, 3$ ) multiplexers, switches, backbones, and WANs that may be geographically remote and heterogeneous. Some of these internetwork components may be parts of the subnet, while others reside in hosts.

There are two sources of technical network management problems:

- Each internetwork component is a complex system with its own troubleshooting, configuration, and maintenance problems.
- The internetwork consisting of these components requires overall management that must concentrate on inter- rather than intracomponent problems.

Specific technical solutions to network management problems are in great degree dictated by organizational decisions. For example, the following LAN management functions are required regardless of what organizational decisions have been made:

- Monitoring and control of server activities
- Analysis of disk usage
- User access to resources
- Performance analysis
- Configuration and fault management

However, the method of providing these functions depends heavily on the organization of the network management process. In the case of completely centralized network management, all of the LAN management functions listed should be carried out from a remote internetwork operator console. If, however, only the first three LAN management functions need to be carried out locally, the last two would be remote.

As another example, even in centralized network management, just collecting all alarms and alerts from all LANs may not be enough. Instead, alarm analysis and interpretation should be done locally, and only the summary must be sent to the internetwork operator console.

Conversely, technological feasibility determines which of the desirable organizational network management options will really work. For example, even though end-to-end network management is well accepted and desirable, it is not feasible today; thus, more realistic trade-offs between desirable and feasible organizational solutions should be considered.

### Functional Facet

The final facet of network management addresses the issue of what should be involved in a typical NMS. From the functional standpoint, network management is a set of operational, administrative, and support functions and corresponding tools that:

- Keep the internetwork operational
- Adjust its architecture to current user requirements and traffic characteristics
- Fine-tune its performance
- Account for its utilization
- Protect it from unauthorized interference

In other words, to provide the internetwork service quality and adaptability, an NMS should support the following functions.<sup>2,3</sup>

- Fault management
- Configuration and name management
- Performance management
- Accounting management
- Security management

Let us look at these functions in more detail.

### Fault Management

Fault management provides four basic services that promote continuous, reliable internetwork operation:

- Fault detection
- Fault diagnosis
- Fault correction
- Fault administration

The *fault detection* service detects fault messages in the network management data stream or receives such messages from other network management functions. Fault messages are converted in an internal format and passed to the fault administration service for logging and to the fault diagnosis or correction service for further actions.

The *fault diagnosis* service attempts to find the fault cause and initiate fault correction at the fault correction service. This service tries to do its job in a fully automated fashion by using active diagnostic tests and/or rules. However, it will provide on-line assistance to the human operator if it is unable to diagnose a fault on its own.

The *fault correction* service uses predefined fault correction rules to restore internetworking. This service also tries to do its job automatically—for example, by switching to the duplicate facility. However, if it is unable to do so, the

fault correction service will provide on-line assistance to the human operator within the limits of its capabilities.

The *fault administration* service is used to build and maintain the fault diagnosis and correction rules and the fault history database. It also provides trend analysis and help for human operators.

### Configuration and Name Management

Configuration and name management is the fundamental part of network management responsible for all kinds of modifications of internetwork components, such as equipment, processes, and services. Whenever it receives a request for internetwork modification, it checks the current component state, confirms the modification validity, performs the modification, and finally validates it. Configurations may be modified to reduce congestion, avoid faulty equipment, or respond to changing user needs.

Configuration and name management functions include:

- Defining internetwork components
- Assigning names to internetwork components and managing them
- Initializing and terminating internetwork components as well as managing their states
- Defining control states and sequences for the entire internetwork
- Managing on-line state modifications for the entire internetwork and its components
- Providing on-line monitoring and reporting of modified states for the entire internetwork and its components
- Maintaining the current state and inventory of all internetwork components

### Performance Management

Performance management is the third fundamental part of network management responsible for analyzing and requesting modifications of internetwork components to provide performance improvements in terms of throughput, delay, and resource utilization. Performance management functions include:

- Collecting performance statistics from all internetwork components, including those of the NMS
- Creating and maintaining the database of historical performance statistics
- Developing performance evaluation criteria and thresholds
- Analyzing current performance statistics aimed at performance fault detection and generation of performance alarms and fault events
- Analyzing long-term trends based on the correlation of current performance statistics and historical patterns
- Initiating internetwork components operation mode and configuration modifications in response to performance fault events
- Monitoring on-line performance of the entire internetwork and its components

### Accounting Management

Accounting management provides a fair distribution of operating expenses among internetwork end users based on service usage. It also provides customization of routing

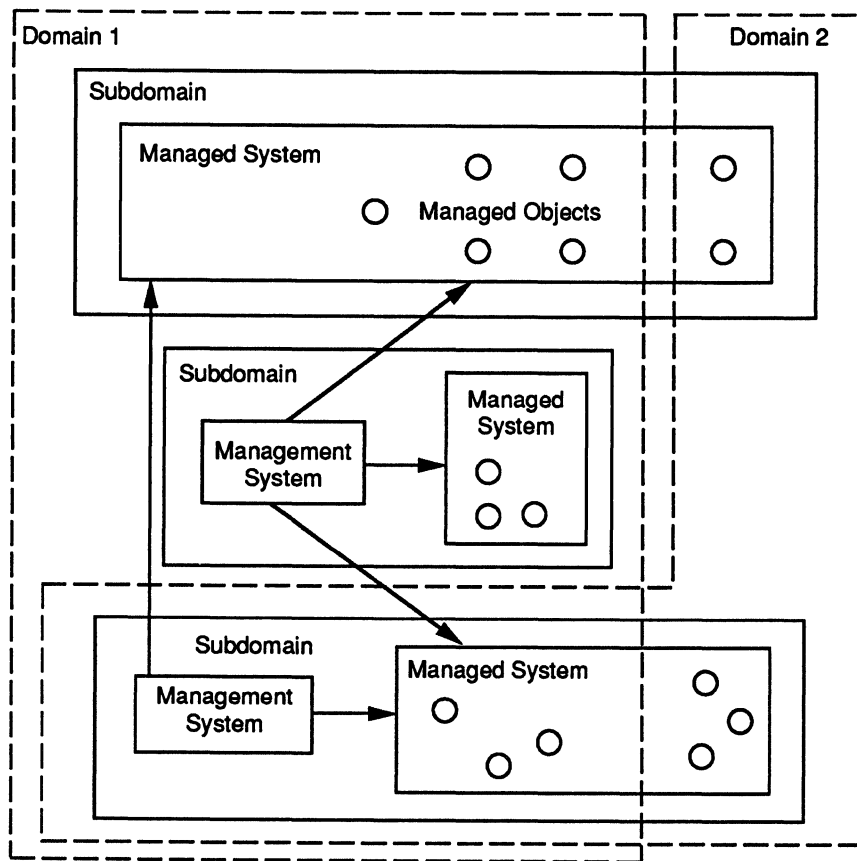


Figure 1. Management Domains

and classes of service by the end-user request. Accounting management includes the following functions:

- Associating tariff schedules with use of internetwork resources based on system utilization statistics
- Calculating costs of internetwork services for users
- Organizing credible billing procedures for services used, based on billing audits
- Calculating combined costs for use of multiple resources.

**Security Management**

Security management implements internetwork security policies. Its basic functions include:

- Distributing security-related information, such as encryption keys and access privileges
- Reporting security-related events, such as network intrusion, violation of access privileges, and access to and update of protected information or services
- Managing security-related mechanisms and services

**OSI View of Network Management**

The ISO looks at network management as an extension of the basic OSI Reference Model. The OSI Management Environment that consists of tools and services required to manage internetworks is discussed in terms of three underlying models:<sup>3,4</sup>

- Organizational model describing how OSI Management may be distributed administratively among management environments

- Informational model providing guidelines for a formal definition of internetwork components and their relationships
- Functional model describing network management functions and their interrelationships

**Organizational Model**

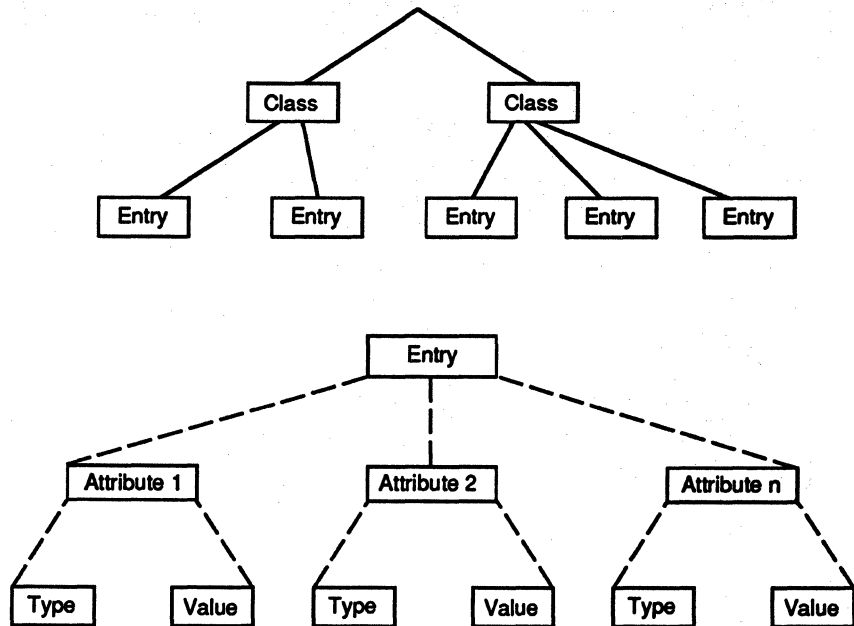
The OSI organizational model is based on the concept of an *abstract object* and an object-oriented representation and computation model.<sup>5</sup> Abstract objects can communicate through *abstract ports*. Abstract ports may be symmetric or asymmetric with respect to services they supply or consume.

One of the fundamental abstract objects is called a *management domain*. The large internetwork can be divided into a number of management domains for administrative autonomy, accounting, or security reasons. For example, three major network management domains are the customer premises (host), the LANs, and the interexchange network (WAN).

Management domains may interrelate in a variety of ways, such as embedding or overlapping (Figure 1). A management domain may consist of one or more *management systems*, zero or more *managed systems*, and zero or more *management subdomains*.

A managed system can be decomposed further into one or more *managed objects*. A managed object is a resource monitored and controlled by one or more management systems. In fact, a managed object is what we called previously an internetwork component, that is, a piece of equipment, a process, or a service. The concept of a managed

Figure 2.  
Management Information  
Tree



object is recursive in that a managed object may be embedded in another managed object. For example, an LAN interface card is embedded in an end-user workstation. Both are managed objects.

A management system is a process that performs monitoring and controlling functions over managed objects and/or management subdomains.

A management subdomain is administered by an administrative authority that could be a public or private organization. The administrative authority is responsible for the creation, modification, and maintenance of managed objects, relationships among managed and management systems, relationships among managed systems and managed objects, and security mechanisms for access to managed objects.

### Informational Model

The core of the informational model is a management information base (MIB). The MIB is required to store and manage diverse information about the OSI system, such as its configuration, current and historic performance data, accounting information, and fault logs.

The MIB represents its information as a collection of *entries*, each describing a managed object. Entries are organized into a management information tree (MIT) shown in Figure 2. The tree hierarchy represents decomposition of managed objects.

Each managed object is specified by three characteristics:

- A list of <attribute, value> pairs. Each attribute must have at least one operation defined for it: for example, read or modify. A value may have its own structure
- A list of operations that can be performed on a managed object. Operation validity and failure may be included in the object specification
- A list of messages that a managed object can issue or be reported on to one or more management systems. Messages can be triggered by error occurrence, exceeding parameter thresholds, elapsed timers, etc.

### Functional Model

The functional model defines the five underlying specific management functional areas (SMFAs) of network management similar to network management functions discussed in "Functional Facets" earlier in this report. In this section we will present the list of SMFAs with emphasis on the OSI procedures available for each of them.

- Fault management. Includes procedures for:
  - Reporting fault occurrences
  - Logging event reports
  - Scheduling and executing diagnostic tests
  - Tracing faults
  - Initiating fault correction
- Configuration and name management. Includes procedures for:
  - Collecting and disseminating data on current state of OSI system and managed objects
  - Modifying OSI system, subnet, and layer attributes
  - Changing the OSI system configuration
- Performance management. Includes procedures for:
  - Collecting and disseminating data on current performance of OSI system and managed objects
  - Maintaining and analyzing performance logs
- Accounting management. Includes procedures for:
  - Informing users about costs
  - Setting accounting limits
  - Combining costs for multiple resource usage
- Security management. Includes procedures for:
  - Authorization and authentication
  - Access control
  - Encryption and key management
  - Maintenance and manipulation of security logs

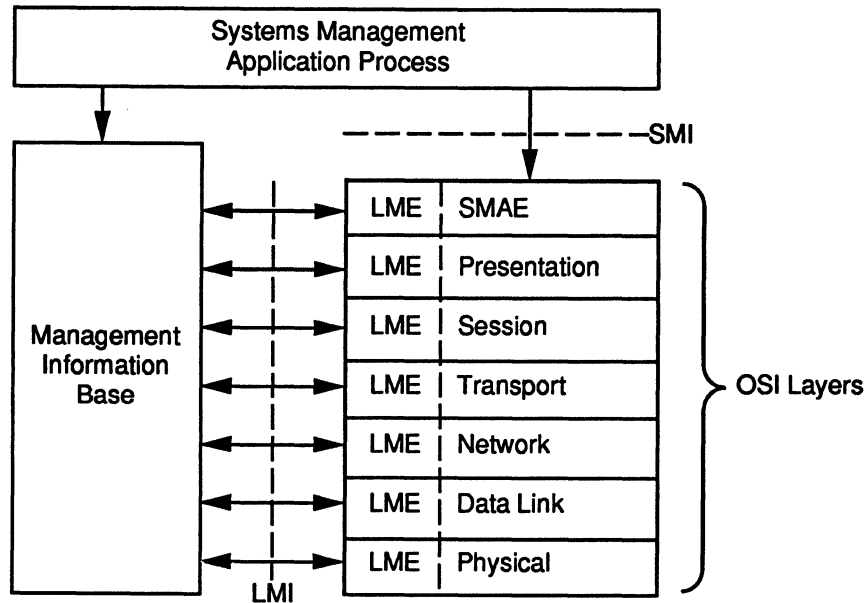


Figure 3.  
OSI Management  
Architecture

LME—Layer Management Entity  
LMI—Layer Management Interface  
SMAE—Systems Management Application Entity  
SMI—Systems Management Interface

### OSI Management Structure

Figure 3 shows the architecture of an OSI system that can participate in OSI management. The architecture provides OSI management from the systems, layer, and protocol perspectives.

*Systems management* provides continuous system operation adapted to changing user requirements and environmental conditions. Systems management is responsible for managing the communication capabilities and communicating entities.

The systems management application process (SMAP) is the local process responsible for implementing systems management functions. It has access to an overall view of OSI system parameters and capabilities in the MIB and can manage all aspects of the OSI system.

The systems management application entity (SMAE) is the application layer entity. SMAE is responsible for communications between systems management entities using for that purpose application layer protocols.

Systems management functions are usually layer-independent or involve multiple layers. Examples include:

- Changing system or network configuration
- Transmitting accounting information
- Requesting comprehensive diagnostic tests
- Coordinating modification of parameters of several layers

*(N)-layer management functions*,  $N = 1, 2, \dots, 7$ , provide the integrity of layer protocols and modification of layer parameters. These functions usually affect the overall operation of the layer and are not used for a single communication instance.

A *layer management process* may be a separate process or a process provided as part of the SMAP. *(N)-layer management entities* communicate through the systems management or *(N)-layer management protocol*. Examples of layer management functions include:

- Reading or modifying layer parameters.
- Testing layers.
- Activating layer services.

*(N)-protocol operations* are used to manage a particular instance of communications so that the system will return to its previous state after this instance is finished. Examples of protocol management functions include:

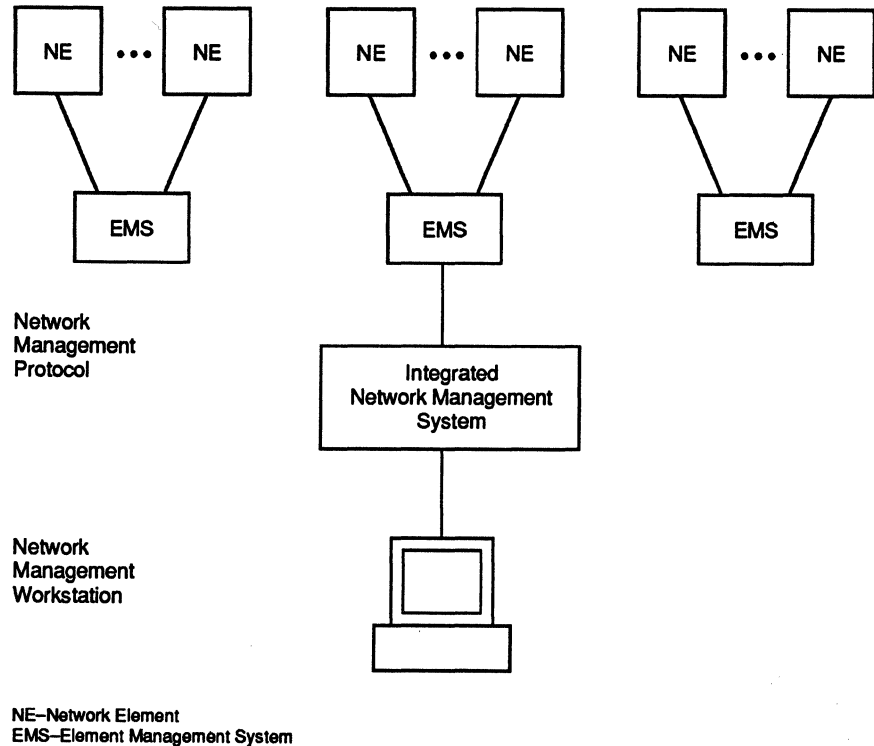
- Modifying connection establishment parameters applicable to a particular communication instance
- Reporting errors and performance data obtained in the instance of communication

### Integrated Network Management System

Like the OSI Reference Model, OSI management is more a promise than a reality. Today, OSI-compliant NMS are almost nonexistent, although more vendors claim they will incorporate OSI standards in their products and prospects seem bright.

In the meantime, several vendors have been addressing one of the most demanding problems of the corporate information infrastructure: *interdomain management*, or integration of many heterogeneous NMS into what is called an *integrated network management system* (INMS). Several INMS products have been available for some time, with IBM's Netview and AT&T's UNMA being the most popular, at least in the Fortune 1000 community.

Figure 4.  
Unified Network Management Architecture



The following concepts are or will be incorporated in INMSs:<sup>3</sup>

- Addressing mostly the part of NMS integration that deals with connectivity
- Providing a consistent and transparent view of INMS's total network management capabilities as if they were provided by a single system (these capabilities are really a combination of many heterogeneous NMS capabilities)
- Concentrating on integration of monitoring and reporting of managed objects and events across individual NMSs
- Accommodating tiered integration architectures that permit hierarchical network management authorities
- Supporting management services similar to SMFAs (see "Functional Model" earlier in this report), namely configuration and fault management
- Allowing users to build their own management applications that drive, or create presentations from, their individual NMSs
- Using a modular approach to accommodate unforeseen changes in environment, user needs, NMS relationships, and so forth

AT&T and IBM, as well as other INMS vendors such as DEC and Hewlett-Packard, develop their INMSs gradually, offering specifications for interfacing other vendors' products with their own systems. For example, IBM's Netview/PC links non-SNA devices to Netview, and

AT&T provides one-way limited connections to Accumaster Integrator. Another evidence of IBM's "good faith" toward OSI standards is that its OSI Communications Subsystem, reportedly the only OSI solution provided by IBM for accessing OSI devices through Netview until 1992, is priced at \$300,000!<sup>6</sup>

The vendors' main excuse for the slow introduction of OSI-compliant INMSs is that OSI standards for network management are not yet ready or are too immature to be fully incorporated in commercial products. Indeed, there is a significant grain of truth in this statement.

While waiting for the advent of end-to-end network management and INMSs to provide it, internetwork users have few intermediate options:

- Using TCP/IP's Simple Network Management Protocol (SNMP), especially in a UNIX environment and for internetworking LANs and WANs together
- Adding ad hoc features to existing NMSs, such as connectivity tools (for example, connecting Netview to any non-SNA box) or expert systems for fault and configuration management
- Using services of network management integrators, such as EDS; Network Management; Ernst & Young's Network Strategies; or major hardware vendors such as IBM, AT&T, DEC, and Hewlett-Packard
- Outsourcing, that is, letting other companies take over the entire network operation (the recent Merrill Lynch with IBM and MCI, or Eastman Kodak with DEC alliances are two examples of this option)

## Unified Network Management Architecture (UNMA)

In perhaps the most comprehensive attempt to comply with OSI management standards and to provide easy migration to an OSI INMS, AT&T has defined a UNMA.<sup>4</sup> Its primary focus is to integrate a variety of independent, and sometimes even mutually hostile, NMSs. Accumaster Integrator is the first partial implementation of the UNMA.

The UNMA is a three-tiered architecture (Figure 4). Network elements (NE), such as the customer premises (host), local exchange network (LANs), and interexchange network (WAN), make up the lowest level. Each NE may have its local management capabilities and is managed by an element management system (EMS).

The middle UNMA level comprises the EMSs. These systems may support different architectures and may be provided by different vendors. The UNMA currently accommodates existing NE-EMS protocols. The future evolution implies standardization of NE-EMS interfaces and protocols.

The highest UNMA level is an INMS that gathers together management functions from all EMSs. The flow between EMSs and the INMS is defined by a Network Management Protocol (NMP) based on the OSI Reference Model and Management.

The UNMA accepts both AT&T's and other vendors' EMSs. Integration of other vendors' EMSs into the UNMA is achieved by the vendors' compliance to the set of NMP specifications and messages.

Another important UNMA feature is that all network management information can be integrated and presented on either the customer premises or the service provider network. When the integration is made on the customer premises, the customer's portion of management information is passed from the provider to the premises. Conversely, when the integration is made at the service provider's site, customer management information is passed from the premises to the provider. In both cases the integrated information is viewed as a network management workstation (NMW).

The NMW provides the unified user interface, including help screens, input command screens, and graphic output. Using the NMW the network management personnel, whether on the customer or service provider premises, can access information about the network, analyze it, modify the network, and perform other network management operations.

The major goal of the NMW is *user transparency*. In other words, the user will view all management capabilities as provided by a single system. This view will not be affected by whether the user works with the INMS or a foreign EMS or with a customer's premises or a service provider's NMS.

As shown in Figure 7, IBM's Netview and AT&T's UNMA are the two most popular INMSs on the market. They are also the subject of many competitive comparisons. (Actually, the two products are more complementary than competing.)

In the UNMA terminology, IBM Netview is an SNA EMS with some integrating features best developed for other IBM products. In its coverage of the IBM environment, Netview is fairly complete; and its only real competitor in terms of functionality is Net/Master. Its interface to non-SNA management systems is provided through another product, Netview/PC, which is incomplete and hard to use.

On the contrary, Accumaster Integrator is first of all an integrator that depends on individual EMSs to manage their NEs. Accumaster Integrator can be used in both IBM and non-IBM environments. In the first case, Netview may be considered as just another EMS hooked up to Accumaster Integrator. In the latter case, all information obtained by Netview is through the corresponding interface.

Summarizing, in the entirely-IBM environment, Netview and Accumaster Integrator do not compete at all and cannot even be compared. It is certainly a Netview domain, with Net/Master, rather than Accumaster, being the major competitor.

In the entirely-non-IBM environment, Netview and Accumaster Integrator do not compete and cannot be compared either, because Netview is of no use in this environment. The decision whether to use Accumaster Integrator depends on many other factors, the major two being the number of different multivendor NMSs used and their mutual compatibility. The case for Accumaster Integrator is the strongest if the enterprise is using multiple independent and incompatible NMSs from various vendors.

Finally, in the mixed IBM/non-IBM environment, Netview and Accumaster Integrator complement one another. If the environment includes SNA, Netview (or Net/Master) is a must. The decision whether to use Accumaster Integrator depends on what portion of the entire environment is non-IBM. The case for Accumaster Integrator is again the strongest if the enterprise is using, in addition to IBM hardware, multiple non-IBM NEs and the correspondent independent and incompatible EMSs from various vendors. In this case, Netview or Net/Master will be used as one of the UNMA EMSs.

Such a position of Netview does not please some IBM executives. Speaking of IBM support for UNMA, Hellen Hancock, IBM vice president in charge of networking business, said: "Netview doesn't work well as a subsystem."<sup>7</sup>

## Conclusion: AI and Network Management

The ultimate goal of network management is to provide an adequate quality of network service and network adaptability to changing technology and user needs without service disruption. OSI-based standardization of network management that follows OSI-based standardization of networks and internetworking should alleviate problems of continuous reliable internetwork operation.

However, networks have become so complex that, even in an ideal OSI environment, there would be very few people who could understand in real time what is going on in a large internetwork, or interpret on-line the stream of management information provided by the INMS or individual NMSs, or just support normal internetwork operations around the clock. The solution to network management problems comes in the form of KBSs<sup>2</sup> that can contribute to virtually every network management function discussed in this report (see "Functional Facet" and "Functional Model").

The most widely used area of AI applications in network management is fault management. Different forms of AI-based fault diagnosis systems are available today. For example, AT&T Accumaster Integrator uses a KBS that identifies a potential problem in an end-to-end network channel, based on the set of consolidated alarms coming from various EMSs. Other applications currently

available include network layout, congestion control, on-line rerouting, and interpretation of performance information.

In the next several years, AI will not only make integrated network management a reality, but it will also help integrate network and information management, providing the glue necessary for putting all pieces of the corporate information infrastructure together.

---

## References

<sup>1</sup>Kerr, S. "Politics of Network Management," *Datamation* (September 15, 1988).

<sup>2</sup>Ericson, E., L. Traeger Ericson, and D. Minoli, eds. "Expert Systems Applications in Integrated Network Management." Norwood, Mass.: Artech House, 1989.

<sup>3</sup>Brusil, P.J. and L. LaBarre. "Managing Networks." In *ISDN, DECnet, and SNA Communications*, edited by Thomas C. Bartee, 255-95. Indianapolis, Ind.: Howard W. Sams & Co., 1989.

<sup>4</sup>Klerer, S.M. "The OSI Management Architecture: An Overview." In "Expert Systems Applications in Integrated Network Management," edited by E. Ericson, L. Traeger Ericson, and D. Minoli. Norwood, Mass.: Artech House, 1989.

<sup>5</sup>Cox, B. *Object-Oriented Programming*. Reading, Mass.: Addison-Wesley Publishing Co., 1986.

<sup>6</sup>*Computerworld* (June 12, 1989).

<sup>7</sup>*Computerworld* (February 20, 1989). ■



# An Overview of Microcomputers

## In this report:

CPU Architectures.....	2
Operating Systems .....	6
Selection Guidelines.....	7

## Datapro Summary

A microcomputer is a reprogrammable electronic device used for processing information. The information is encoded as a digital signal and manipulated over circuits etched on wafers of silicon. Microcomputer sophistication has progressed from the simplicity of hardwired devices with a few circuits on a single silicon wafer to the complexity of very large scale integration (VLSI) that places over 100,000 circuits on a chip. A multitude of product offerings standardized around the Intel 386-series of microprocessors has created an intensely competitive market with commodity-like characteristics. Price/performance is now the primary issue. The previous issues of compatibility, reliability, and expansion options are less of a concern.

## Technology Basics

Microcomputer technology combines a microprocessor with interrelated subsystems and VLSI (Very Large Scale Integration) support circuitry. A basic microcomputing system comprises the following main components and subsystems: a single microprocessor or Central Processing Unit (CPU) for controlling the system; a system bus for data transfer; a memory subsystem, which temporarily stores programs and data; a mass storage subsystem for permanent storage of programs and files; a digital or analog video/display subsystem for viewing system output; and a keyboard and/or a mouse for data entry and responding to the system.

### Microprocessors

The microprocessor, also known as the central processing unit (CPU), is an integrated digital circuit that processes information sequentially via a series of on/off electronic states. The microprocessor is the single most important component in a microcomputer. It executes program instructions, reads data from memory, and accesses the peripheral devices.

All IBM microcomputers and compatibles use the Intel 80X86 series of microprocessors—the 8088/8086, 80286,

80386SX, 80386, and i80486. Other choices include the Motorola 680X0 series—68000, 68020, 68030, and 68040—typically found in Apple Macintosh computers and technical workstations.

Microprocessor performance is a function of several factors:

- clock speed
- word size
- processor instruction set
- address and data bus widths

Clock speed refers to the frequency of CPU timing cycles (measured in megahertz (MHz), or millions of cycles per second), determined by the number of pulses generated by a quartz crystal oscillator. The CPU manufacturer determines the finished chip's highest acceptable clock speed through empirical testing. During each pulse, the CPU performs one small operation. The higher the clock speed, the faster the microprocessor executes instructions.

## CPU Architectures

Processor Type	Word Size (bits)	Data Bus (bits)	Address Bus (bits)
<b>Intel</b>			
8088	16	8	20
8086	16	16	20
80286	16	16	24
80386SX	32	16	24
80386DX	32	32	32
i486	32	32	32
<b>Motorola</b>			
68000	32	16	24
68020	32	32	32
68030	32	32	32
68040	32	32	32

Word size (in bits) refers to the size of the operand registers internal to the CPU. A larger word size means that larger pieces of data can be manipulated internally, increasing processing speed. With successive generations of microprocessors, word widths have expanded from 4 bits (in the earliest microcomputers) to 8 bits and to 32 bits. The width of the data bus (also known as the I/O bus) determines the amount of data that can be transferred into and out of the CPU on a single pass. A wider data bus means faster performance. For example, the 16MHz 80386SX (with its 32-bit word size and 16-bit data bus) is slower than the 16MHz 80386, which is a "true" 32-bit (word size and data are both 32-bits wide) chip—even though the processors are identical internally.

The processor's instruction set determines what processes can be carried out directly. Simple instructions can be carried out in one or a few processor cycles, while more complex instructions generally require more cycles. For most general-purpose applications on microcomputers, however, a rich instruction set is favored.

Because IBM and IBM-compatibles, as well as Apple Macintoshes, are the dominant microcomputers, their associated microcomputer "instruction sets" are the most important, found, respectively, in the Intel and Motorola series of processors. The Intel 80x86 and Motorola 680X0 series are known as Complex Instruction Set Computers (CISC), because they use complex instructions composed of levels of microcode stored on the chip. As chip complexity increases, however, performance bottlenecks are encountered.

Reduced Instruction-Set Computer (RISC) architectures have been proposed as alternatives to CISC. RISC provides fewer microcoded instructions; because complex instructions can be built from many simple instructions, the circuitry is less complex and higher clock speeds are possible.

The total physical amount of memory that can be addressed directly by the CPU is determined by number of bits in the address bus. Every memory location must have a unique binary address. Microprocessor addressing ranges have increased from a few kilobytes (K bytes) to 4

Gigabytes (G bytes). For example, the 8088/8086 microprocessors, used in the original IBM PC and subsequent IBM-compatible clones, have a 20-bit address bus supporting 1 megabyte (M byte) of memory ( $2^{20} = 1,048,576$ ). A 24-bit address bus (80286 and the 80386SX) supports 16M bytes, while a 32-bit address bus (80386 and i486) allows up to 4G bytes. With more memory, a processor can handle more data without slow input/output (I/O) access and can also run more complex programs.

Operating within the one megabyte address space of the Intel 8088/8086 microprocessors is called working in real mode. The operating system (DOS) and applications were given the first 640K bytes, while memory from 640K bytes to 1M byte was reserved for system administration BIOS, adapters, and video buffers.

To increase the 1M-byte limit of Intel 8088/8086 CPUs, vendors have introduced several schemes that take advantage of *bank switching*. With specialized hardware and software, bank switching uses part of the CPU's address space as a window into extra pages of RAM, and memory is thus expanded. Expanded Memory Specification (EMS) Version 4.0 has become a standard. Expanded memory is often confused with extended memory (the amount of memory above the 8086/8088 1M-byte limit) and conventional memory (the traditional 640K-byte limit of MS-DOS).

Intel's 80286, introduced in 1984, added a protected mode to the real mode of the 8088/8086 microprocessors. In protected mode, the 80286 can address 16M bytes of memory and also use four privilege levels to manage system memory and I/O devices. Applications developed for the 8088/8086 run on the newer processors in real mode and benefit from increased processor speed.

The 80386, introduced in 1986 (in a recent Intel marketing maneuver, the "80386" is now referred to as the "386DX"), added still another mode of operation: the virtual-86 mode that simulates an 8086 machine. Memory management units, included in more recent generations of microprocessors, have eased the implementation of multituser, multitasking operating systems and virtual memory.

Through the 80386's paging functions and virtual-8086 mode, multiple 8088/8086 machines can be simulated (allowing concurrent processing of 8088/8086 applications) under an 80386 protected-mode operating system. The 80386 can run a single copy of an 80286 protected-mode operating system, but it cannot run multiple copies. Therefore, 80386 multitasking can be accomplished more easily with 8088/8086 applications and with applications written specifically for the 80386.

Cache memory and floating-point processing have required additional external chips until the most recent generation of microprocessors. The Intel 80486 (also referred to as the i486) derives significant benefits from VLSI—the 80486 contains over 1.2 million transistors compared to the 80386's 275,000 transistors and includes floating-point, cache memory, and paging functions—allowing for higher speed accesses on-chip rather than through slower external chips and circuit boards.

The i486 is essentially an enhanced 386DX: both provide 32-bit data and address buses yielding 4G bytes of physical and 64T bytes of virtual addressing capabilities per task. Real, protected, and virtual 86 modes function similarly, with nearly all 80x86-compatible software able to run on the 80486 without change. The basic instruction sets are identical with the exception of six new instructions

## Architecture Today and Tomorrow

In recent years, one of the more popular topics for panel discussions at computer conferences and trade shows has been the "RISC versus CISC" debate. Besides having a lot of entertainment value (chip designers defending their favorite architectures the way a goose protects her goslings), these debates provide a glimpse into the future of computer design.

Reduced-instruction-set computers and complex-instruction-set computers have differing instruction-set strategies. RISC processors feature a small number of instructions that each execute in one machine cycle. CISC processors use complex instructions that can take several cycles to execute. RISC proponents argue that you get better performance by executing many simple instructions than by executing fewer complex instructions. CISC proponents argue the opposite.

The RISC versus CISC debate won't be decided by panel discussion; it will be won in the marketplace. And the deciding factor may have little to do with numbers of

instructions and registers, and more to do with parallelism.

### The von Neumann Blues

In 1946, in collaboration with Arthur W. Burks and Herman H. Goldstein, John von Neumann wrote a paper that delineated the concepts on which nearly all computers (both RISC and CISC) have been built since. The paper, "Preliminary Discussion of the Logical Design of an Electronic Computing Instrument," advanced the concept of the stored program and introduced the idea of the program counter. Because it described a processor, the so-called von Neumann machine that had to fetch successive instructions from memory, it also defined the bottle-neck between the processor and memory that survives to this day.

Most people would agree that the memory-processor choke point has been a small price to pay for the 40 years of progress based on the von Neumann machine. Computers have grown more powerful every year and will continue to do so for some time. It didn't matter that computers could do just one task at

a time as long as they kept doing it faster and faster.

Ever since von Neumann defined the digital computer, however, designers have been investigating ways around the bottleneck. By their natures, RISC and CISC entail different solutions to the problems of parallelism.

### Inside, Outside

Since their conception, RISC processors have been evolving toward micro-parallelism, incorporating parallel processing features within the processor. Specifically, RISC processors are becoming superscalar; they can execute more than one instruction at a time.

Like other processors, a RISC processor has many components such as the integer unit and the floating-point unit. And, also like many other processors, RISC processors feature pipelining, whereby many instructions can be decoded while one instruction executes. RISC processors, however, are moving toward pipelines for each unit of the processor. Thus, instructions that use the integer unit are pipelined separately from instructions that use the floating-point unit. Instructions that use mutually exclusive parts of the processor can also execute at the same time. The result is a processor that can execute two or more instructions per machine cycle.

CISC processors also employ pipelining, and newer processors such as the

80486 and the 68040 have many integer instructions that execute in one cycle, but the varying execution times of CISC instructions limit the effectiveness of the superscalar approach to parallelism. Instead, CISC processors, with their ever-larger on-chip caches, are better suited to macroparallelism, where multiple, identical processors are bound together on a common bus.

### Obstacle Course

The problems with superscalar processors involve identifying which instructions are independent and which must be executed in a particular sequence. Superscalar RISC machines will require incredibly complex compilers and instruction-decoding logic. Multiprocessor systems based on identical CISC processors require sophisticated systems software for task scheduling, high-speed buses to limit contention, and workable cache-coherency schemes to ensure data integrity.

In the end, the winner of the RISC versus CISC debate will be the architecture that delivers the best solution to the marketplace. In the future, the quality of the solution delivered by RISC and CISC machines may depend less on the number of clock cycles they use per instruction than on the number of instructions they can execute at one time.

Byte Magazine.

added to 80486. Other changes support on-chip functionality and improve on the 80386's memory management algorithms.

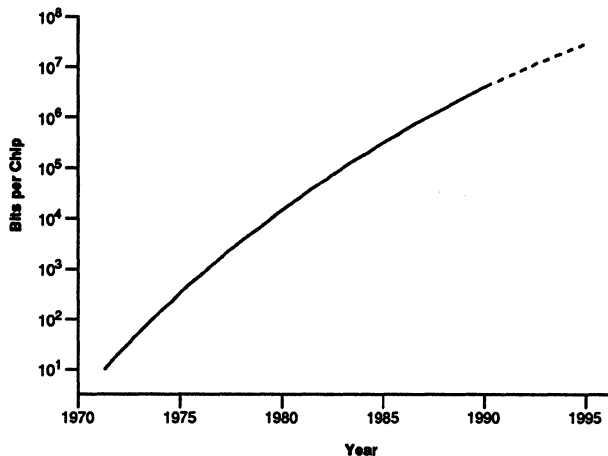
### System Bus

External instructions and data enter and leave the microprocessor via the system bus, a multilane highway of wires that carries data, controlling signals, and electrical power to all the major parts of the system. In traditional PC and AT designs (8-bit and 16-bit busses respectively), the CPU is responsible for maintaining the bus and for moving data around from one peripheral to another. The wider the bus, the greater the amount of data the system can move in a single operation.

Until the advent of the 80386 microprocessor, the AT bus was adequate for the 16-bit demands of the 80286. When Intel introduced the 80386, however, a new bus needed to be developed to handle the 32-bit capability of the 80386 chip. full potential. To provide advanced internal expansion capabilities, and in an attempt to gain market share, IBM introduced Micro Channel Architecture (MCA) in its PS/2 systems in April 1987.

A group of vendors led by Compaq Computer proposed an alternate bus called Extended Industry Standard Architecture (EISA) late in 1988. EISA is an extension of the original AT bus that maintains compatibility with the 8MHz AT standard. Both EISA and MCA provide 32-bit data paths; therefore, both will outperform the current 16-bit AT bus.

Figure 1.  
Production DRAM Density



The storage capacity of RAM chips—a direct function of the number of components per chip—doubles every three to four years. Note that the rate of increase slows with time because of the increasing complexity of fabricating near-micron and submicron technologies.

Source: Byte Magazine.

MCA represents IBM's strategy for the future—an I/O architecture that can fully support advanced microprocessors and can be extended to larger systems. MCA has been implemented on IBM platforms ranging from RISC System/6000 workstations to Enterprise System/9370 mainframes. In comparison, EISA-based systems offer a broader, evolutionary design to technological advance; it provides backward compatibility with existing systems—MCA does not. EISA systems have been available since early 1990 and are slowly being accepted.

### Memory Subsystem

The memory subsystem allows the CPU to receive and store start-up information and accesses the operating system, hardwired programs, and software applications. Microcomputer memories are classified into two major categories: read-only memory (ROM) and random-access memory (RAM).

All ROM is nonvolatile memory—data is not lost when power is turned off. The program code needed for system start-up, including the Basic Input/Output System (BIOS), is stored in ROM. The BIOS is the direct interface between the processor and subsystems such as mass storage, video, and keyboard. In addition, application programs can access the hardware through the ROM BIOS. Most microcomputer ROMs include system setup and diagnostic programs; some also contain the operating system or application programs.

Most microcomputer ROMs are nonprogrammable; their contents cannot be changed. Some kinds of ROM can be erased and reprogrammed. For example, Erasable Programmable ROMs (EPROMs) can be erased by exposure to ultraviolet light and reprogrammed.

RAM, unlike ROM, is volatile—it will hold its contents only while power is on. The CPU can read from and write to RAM, which is usually used to store code and data during an operation. Two kinds of RAM are used in microcomputers: the more common, lower cost Dynamic RAM (DRAM) and expensive Static RAM (SRAM).

DRAM consists of a series of capacitors that stores bits as charges. The presence of a charge represents a “one” bit (on); the absence of a charge represents a “zero” bit (off). Unfortunately, this kind of electronic memory arrangement by itself will completely discharge before the CPU has time to decode it. To be useful in a microcomputing system, therefore, DRAM must be refreshed frequently to maintain the proper charge.

SRAM stores bits in a series of flip-flops or registers that can be switched to either a zero or a one state. SRAM maintains its contents without the need for a refresh cycle, providing a faster access time than DRAM. Once a bit has been stored in SRAM, it retains that state until explicitly changed. Cost and system board space are limiting factors in all memory designs. SRAM's complex circuitry makes it expensive and difficult to implement.

In addition to capacity, memory chips are rated for access time (the amount of time it takes for the chip to deliver data once the request is made) in nanoseconds, or billionths of a second. Ideally, the length of time it takes RAM to deliver data to the CPU should be less than the amount of time it takes the CPU to get ready for it. As microprocessor clock speeds have increased, however, RAM access times have not decreased proportionately. In this instance, the CPU operates with “wait states.” If the RAM circuitry cannot deliver the data fast enough, wait states must be inserted to keep the CPU from trying to act on data it hasn't yet received. Wait states impose notable performance penalties. Consequently, systems designers have implemented various memory arrangements to improve CPU/RAM accesses. Most systems operate between zero and one wait state. Among the methods used are direct memory access (DMA), row/column memory, static-column RAM, interleaved memory, memory paging, cache memory, and write posting shadow RAM.

DMA circuitry allows a subsystem to bypass the CPU and directly read from and write to RAM. Row/column memory (the traditional method) is mapped as a matrix and a particular address is given using a row and a column number. Static-column RAM is a technique using DRAM that operates like DRAM for initial random access and like SRAM for subsequent sequential accesses. If the row address stays the same and only the column address changes, the first row address and the second column address work like static RAM. The original Compaq Deskpro 386 was designed with a static-column memory subsystem.

Memory interleaving is a technique that divides the memory into two or four portions that process information alternatively. The CPU sends information to one section for processing, while another section goes through a refresh cycle. Memory paging is divided into pages—typically 2K bytes. Consecutive memory accesses to the same page are performed more quickly. DRAM access normally requires a row address followed by a column address, thus reducing the time required for memory access. Write posting allows designers to optimize memory designs for the high percentage of writes used in 80486 CPUs; more than 70 percent of all memory bus cycles are writes due to the presence of the 8K byte on-chip cache. Through write posting, the system signals the CPU that the write cycle has completed before memory has actually accepted the data. The CPU can begin executing the next instruction sooner, reducing the number of CPU wait states in the write cycle.

Cache memory, first used in mainframes and minicomputers, frequently accesses data stored in a small amount (typically 32K or 64K bytes) of very fast SRAM (35 ns. or

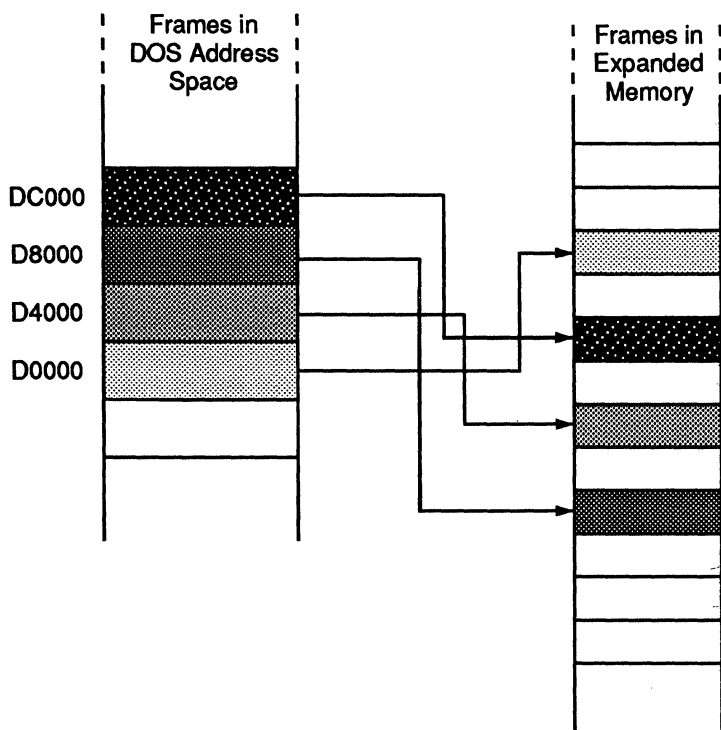


Figure 2.  
EMS Memory Mapping

DOS breaks memory into 64K-byte segments. These segments fall into three primary areas. Conventional memory, also called user memory, occupies the lower 640K bytes. Upper memory, normally reserved for system and expansion ROM, uses the next 384K bytes. Extended memory, which is beyond the address ranges of both DOS and the 8088 processor, starts above 1 megabyte.

Source: Byte Magazine.

less). If data needed by the CPU happens to be in the cache, it is accessed without wait states. The resulting improvement in CPU performance depends on the cache size and on the algorithm used to determine what data remains in the cache. Shadow RAM is not a memory arrangement but a technique that loads system BIOS or video BIOS directly into faster RAM on start-up. The BIOS or video instructions can then be accessed more quickly.

**Mass Storage**

The earliest microcomputers to advance beyond cassette tape for mass storage used eight-inch diskette drives. The 5.25-inch size became popular with the CP/M operating system and a standard with MS-DOS. In 1984, the Macintosh was among the first to use 3.5-inch diskette drives; laptop computers and IBM's PS/2 line have provided further impetus. Zenith has also introduced the two-inch diskette drive, which may approach a practical size limit.

Diskettes vary in their density, or capacity, for data storage. This capacity is partly determined by the microcomputer manufacturer's formatting design (preparing areas of the magnetic surface for data storage). Capacity also depends on whether the medium is single or double sided.

Inexpensive and reliable, 5.25-inch diskettes are circular Mylar disks coated with magnetic particles. The disks are enclosed in a protective PVC jacket that has a write-protect notch and openings for the drive spindle, the read/write head, and the timing hole. The 3.5-inch diskettes are comparable but have a hard plastic case with a sliding metal door to protect the medium.

Hard disk drives are designed for larger storage requirements. A hard disk consists of several circular aluminum platters coated with a film of iron oxide or other magnetized metal and spinning at 3,600 revolutions per minute. The read/write heads hover above each surface on a cushion of air. To read or write data on the hard disk, a stepper motor or voice coil moves the heads above the appropriate track and sector of the disk where the data resides. Hard

disks are usually enclosed in a sealed box to protect the drive heads and the medium from environmental contaminants such as dust.

Hard disk speed, or access time, is measured in milliseconds. The access time is the interval required for a hard disk to retrieve or deliver data from the disk after the CPU signals a request. Another factor affecting speed is the disk interleave, the number of physical sectors between consecutively numbered logical sectors. For example, an interleave of three means that the drive reads every third physical sector as contiguous. The interleave should allow the operating system enough time to process data before the disk's read/write head reaches the next logical sector.

Disk speeds can be improved with a disk cache. Similar to a memory cache, a disk cache is a software implementation that stores copies of recently used disk sectors in a reserved area of RAM. When the system accesses sectors that are already in the cache, no disk access is required and the response time is much faster. If the sector is not found in the cache, the disk is read normally and the caching algorithm stores a copy of the sector in the cache. A disk cache can improve disk access time by as much as 50 percent.

The hard disk controller is the on-board circuitry that controls hard disk operations, determining data transfer rates and the data encoding scheme. Popular data encoding schemes include the common modified frequency modulation (MFM) and run length limited (RLL), the latter providing higher densities and faster transfer rates.

The hard disk interface determines data transfer rates. Seagate's serial ST506/412 interface, used in the original IBM PC XT, has a 5M bit-per-second data transfer rate. For higher performance, both the enhanced small disk interface (ESDI) and the small computer systems interface (SCSI) are popular. Established by a consortium of 22 disk drive manufacturers, ESDI (an enhanced version of the ST506) has a 10M bit-per-second serial data transfer rate. SCSI is actually a system bus with a 32M bit-per-second

transfer rate. SCSI requires a dedicated microprocessor-based controller and allows up to seven daisy-chained devices. The IDE Intelligent is popular because the design allows for inexpensive drives with moderate performance capabilities similar to RLL drives. Developed by Conner Peripherals and Western Digital, the IDE drive (similar to the SCSI drive) controller is integrated onto the drive itself. Unlike SCSI drives, the IDE controller interfaces directly with the computer electronics either through a system board connector or an inexpensive adapter board.

### Video/Display

A video adapter, found either on an expansion board or integrated on the system board, consists of three elementary components: the video controller, the video BIOS, and the video RAM (VRAM). The video controller generates the appropriate CRT scan signals for the graphics or characters that appear on-screen. The video BIOS provides a smooth interface to perform various video and screen tasks such as mode selection, cursor positioning, and character writing. The video RAM (VRAM) is dedicated video memory that contains information defining the screen image.

The video controller produces a series of beam-intensity levels and some control signals. The display uses the control signals and turns them into an on-screen CRT image. CRTs fire an electron beam against a phosphor-coated screen to turn each pixel on and off separately.

With IBM's introduction of the Personal System/2 in April 1987, the Video Graphics Array (VGA) was established as the new display standard. The 640-by-480 resolution is an incremental improvement over the earlier EGA standards 640 by 350-pixel resolution. VGA uses analog signals rather than the digital inputs of all previous color adapters (CGA, EGA, PGA). Analog signals allow the video controller to vary voltage output on-screen, which results in more gradations of colors. For example, the VGA can display 256 simultaneous on-screen colors from a palette of 256,000.

In October 1990, IBM introduced the Extended Graphics Array (XGA) standard providing 1,024-by-768-pixel resolution, 256 colors, and compatibility with VGA and its earlier 8514/A, IBM's first attempt at improved resolution. The 16-bit XGA display adapter is integrated on the system board in the IBM's Model 90 486-based microcomputer and as a 32-bit bus-master Micro Channel card for IBM's Model 95.

The XGA board also has a mode that features a lower resolution (640 x 480) but more colors (64K). XGA is compatible with VGA and 8514/A graphics standards, but not with software that writes directly to the 8514/A registers. IBM claims that the new video board is as fast as or faster than the older 8514/A adapter, which is effectively being replaced.

### Input Devices

The "QWERTY" key arrangement was designed in the 19th Century to inhibit fast typists, who sometimes jammed type bars on the original mechanical typewriters. Alternatives, such as the Dvorak keyboard, have been advocated to improve typing efficiency, but the old design remains entrenched for office equipment from typewriters to microcomputers to mainframe consoles.

In 1986, IBM introduced the 101-key enhanced keyboard. Twelve function keys are included in a single row across the top, and separate cursor-control keys were added. Control and Alternate keys were placed on each

side of the spacebar. In 1987, IBM adopted the enhanced keyboard across the entire PS/2 line.

Apple's original Macintosh keyboard was compact, with little more than the alphanumeric and punctuation keys. This was not much of a limitation, however, since the graphical user interface used a mouse and was common to all Mac applications. With the open architecture of the newer Macintosh models, however, Apple introduced an extended keyboard with 15 function keys to accommodate alternate operating systems and terminal emulation.

Until recently, the keyboard was the dominant input device for IBM and IBM compatible computers. However, the success of the Macintosh and the appeal of new graphical user interfaces (GUI's), especially Microsoft Window's, have made the mouse a popular microcomputer peripheral. Invented by Douglas Englebart in 1963 at the Stanford Research Institute and refined by Xerox at the Palo Alto Research Center (PARC), the basic mouse design remains constant. Other input alternatives include trackballs, graphics tablets, light pens, and touch screens.

### Operating Systems

#### DOS

Since 1981, Microsoft's MS-DOS (and IBM's version, PC-DOS) has remained the dominant single-user, single-tasking operating system for microcomputers, with an existing base of tens of thousands of available applications. The reason for this large base of third-party software is the binary standard that links the Intel 8086, 8088, 80286, 80386, and i486 microprocessors at the object code level. Third-party software vendors provide utilities for MS-DOS that further enhance its capabilities. More advanced systems, however, require a more advanced operating system.

#### OS/2

OS/2, developed jointly by IBM and Microsoft, is a single-user, multitasking operating system. Released in December 1987, OS/2 offers a larger addressable memory than MS-DOS (16M bytes versus 640K bytes) and the ability to run more than one program at the same time. OS/2 takes advantage of the protected mode of the Intel 80286 microprocessor. While OS/2 also runs on the 80386 microprocessor, current versions do not take advantage of that CPU's inherent capabilities.

#### UNIX

Once confined to academic and engineering applications, UNIX was designed as a multitasking, multiuser operating system. First developed at AT&T Bell Laboratories and in use since the early seventies, UNIX has since been optimized to overcome its original weaknesses—an unfriendly interface, lack of multiprocessor support, and poor record/file locking and security capabilities. Written largely in the C programming language, most UNIX implementations are based on one of two versions: the original AT&T version, now labeled System V, and the BSD (Berkeley Software Development) Version 4.2, created at the University of California at Berkeley. Microsoft's Xenix (a version of UNIX) was developed for the 80386 chip. Unlike OS/2, UNIX can take advantage of the 80386's capabilities. Thus, UNIX is an attractive alternative for high-end workstations based on the 32-bit chip. UNIX is gaining broader market acceptance because of its relative portability to 386/486 and RISC-based platforms, making it a logical choice for distributed client/server processing.

### MultiFinder System

The Macintosh operating system, known as MultiFinder/ System, has a consistent user interface across applications. Based on concepts originally developed at Xerox PARC, the graphical interface uses windows, icons, and a mouse. The Macintosh operating system is especially well suited to working with high-quality graphic images, making it an appropriate tool for desktop publishing.

The Macintosh interface's success has spawned imitators. Two such products, Microsoft Windows and the OS/2 Presentation Manager, provide similar Mac-like interfaces for IBM-compatible microcomputers.

---

### Selection Guidelines

Purchasing a microcomputer is no longer a simple decision. Installing client-server architectures, replacing mainframes, and adding intelligent nodes on a network now require strategic planning. Once applications are decided upon, hardware parameters are determined. Implementing an effective LAN will require 80386- or i486-based servers with a true 32-bit operating system. Sophisticated

graphics applications such as CAD/CAM typically require a high-performance 80386 running UNIX. Where cost is a concern, DOS and OS/2 applications running on 80386SX systems are sufficient. In some cases, for simple applications such as word processing, spreadsheet accounting, and medium-sized databases, an upgrade to current equipment is sufficient. The following questions should be considered when selecting microcomputers:

- What hardware/software is needed for present and future needs?
- Will the financial investment be worth the cost of implementation and administration?
- Do the circumstances warrant the immediate purchase of equipment, or will it depend on market conditions?
- Will individual and organizational productivity be increased?

Strategic planning is best thought of in terms of time and technology. The rapid changes in microcomputing technology make these decisions critical to the success of the project. ■





# An Overview of Portable Microcomputers

## In this report:

Mass Storage.....	2
Future Technology.....	7

## Datapro Summary

Portable computers are an exercise in engineering compromise. Portable designers face the challenge of providing desktop features and functions within size and weight constraints and at an acceptable price point. Transportable technology, like its desktop counterparts, has matured; laptop technology, on the other hand, is still improving. Initially laptop systems included Intel 8088 and 8086 microprocessors. Advances in processor technology combined with advances in miniaturization have since enabled portable systems to become roughly equivalent with their desktop counterparts using the 80286, 80386SX, 80386DX, and i486 microprocessors.

## Technology Basics

### Microprocessors

Portable microcomputers have followed in the performance footsteps of their desktop counterparts—initially using Intel 8088/8086 processors, then the Intel 80286 and 80386DX; some transportables use the Intel i486 processor. Because transportable microcomputers are simply scaled-down desktop systems, users now have the option of replacing older desktop systems with transportable units offering increased performance, similar storage capacities, and equivalent display resolution.

Battery-operated laptops use CMOS (Complementary Metal Oxide Semiconductor) microprocessors, which consume less power because the current flows through the chips only when they change states (from off to on). Intel's 80C88, 80C86, and 80C286 are CMOS versions of the 8088, 8086, and 80286 chips, respectively; however, the 80386 and the 80386SX are inherently CMOS designs.

Intel's 16MHz and 20MHz 80386SX processors have been attractive choices for

laptops, especially notebook laptops, because they require less power than their 80386DX relatives and use surface mount technology. The SX's internal 32-bit architecture provides all the capabilities of the 80386, but stores and accesses data in 16-bit blocks like the 80286. The 80386 has a 32-bit data bus and can transfer data between the processor and CPU twice as fast as the SX. The SX is compatible with all 8- and 16-bit software and runs software specifically written for the 32-bit 80386. The SX, like the 80386, can run virtual-86 mode tasks, which work in protected mode and permit multiple simultaneous DOS applications.

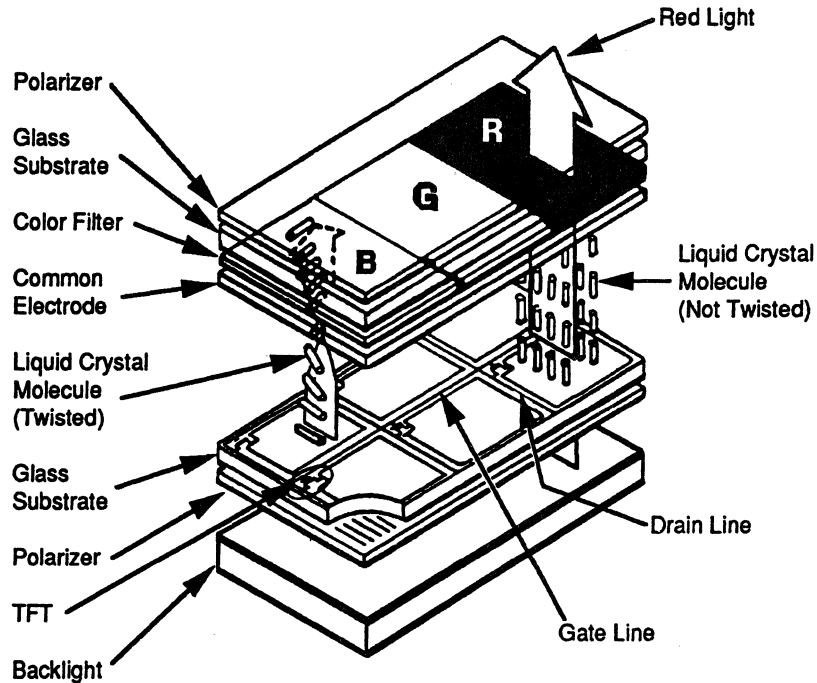
Although the majority of laptops and notebooks use Intel microprocessors, some laptops use 16MHz and 20MHz 80C286 microprocessors (made under license from Intel by Harris Semiconductor and Advanced Micro Devices). Others use the NEC V30 and V20 microprocessors, which are code-compatible with the Intel 8088 and 8086 chips, respectively, but provide enhanced performance.

### Memory

Transportables, like desktops, use the same memory technology (RAM and ROM); however, smaller system boards restrict memory expansion. Consequently, some transportables and all laptops and handhelds usually have less maximum memory

—By George A. Thompson  
Assistant Editor/Analyst

Figure 1.  
Active Matrix LCD  
Technology



than their desktop counterparts. While most transportable microcomputers can accommodate up to 16M bytes of RAM, for example, the average is 5M bytes for laptop systems and 3M bytes for notebooks.

#### Memory storage

Unlike RAM, ROM is nonvolatile (i.e., requires no power to hold its contents) and mask-programmed during the manufacturing process. Laptops, therefore, often use ROM for software applications and sometimes a version of MS-DOS. With these programs in ROM, laptops preserve battery life by not having to boot or load applications from either diskette or disk drives. For example, the ROM in the Zenith MinisPort contains MS-DOS and a file transfer utility; Tandy's 1100 FD includes MS-DOS and the DeskMate interface.

Although ROM is a low cost and reliable storage method, software packaged in ROMs cannot be easily updated. Consequently, ROMs must be discarded and replaced, an expensive solution. Programmable ROMs (PROMs) provide an alternate solution. Similar to ROMs, PROMs are custom programmable by an OEM or end user using a PROM programmer. Two kinds of PROMs are EPROMs (Erasable Programmable ROMs) and EEPROMs (Electrically Erasable Programmable ROMs). EPROMs can be reprogrammed after exposure to an ultraviolet light source but cannot be programmed on the fly unlike EEPROMs, which are erased electrically; however, EEPROMs use complex technology, that results in lower density, higher cost, and questionable reliability.

#### Mass Storage

Transportable microcomputers generally emulate desktop systems using either standard 5.25-inch, 1.2M-byte diskette, or 3.5-inch, 1.44M-byte diskette drives.

All laptops are equipped with 3.5-inch, 1.44M-byte diskette drives. Because conventional disk drives require significant amounts of space and power, mass storage represents the biggest challenge in notebook system design. Among laptops, particularly the "notebook" size, creative

solutions—such as silicon hard disks and 2-inch minidiskettes—decrease weight and improve battery life.

Silicon hard disks, unlike RAM disks, are a portion of RAM powered by an internal battery so that stored programs and data are retained when the system is shut down. Because silicon hard disks retrieve data at the speed of DRAM, access times are much faster than with either RAM diskettes or disks. The NEC UltraLite and the Zenith MinisPort include 1M-byte and 360K-byte silicon hard disks, respectively.

The NEC uses a silicon drive as its primary storage medium; the Toshiba T1200XE and Zenith contain hard drives. The NEC's 1M-byte silicon drive (expandable to 2M bytes) is too small for large spreadsheets or realistic database operations. Credit card-size, battery-backed 256K-byte RAM cards and application ROMs are installed in an external slot. RAM cards, which store data like standard diskettes, are lighter, smaller, and require less power than standard magnetic media, but they are more expensive and more volatile. The application ROMs (0.025-ms. access times) are easily installed and have a negligible effect on battery life. ROM cards are currently available for Lotus 1-2-3, Agenda, Metro Express, XyWrite, WordPerfect, Microsoft Works, and WordStar.

Minidiskettes are 2-inch versions of a typical 3.5-inch, 1.44M-byte diskette. The Zenith MinisPort uses such a minidiskette primarily for backup and for storing working copies of programs.

Hard disk drives, generally standard on transportables, provide another design challenge for laptops. Disk drives are the most power-hungry component of any microcomputer. Disk drive platters must constantly spin at a high speed to guarantee fast and reliable access to data. Although stopping the spin is an alternative, access times become inconveniently long and the system BIOS must be redesigned.

Mass storage devices have shrunk in size without compromising storage capacity. Laptops, once limited to 10M- or 20M-byte hard drives with a slow, 150-ms. average access time (AAT), now routinely offer faster 3.5-inch hard

disks with 20M-byte (27-ms. AAT) and 40M-byte (29-ms. AAT) capacities. A few models even offer 100M-byte hard disk options for 80286 or 80386SX portables.

Portability also holds inherent dangers for hard disk drives, which are sensitive to movement and sudden shocks. Transportable and laptop disk drives are shock mounted and should have a disk "parking" mechanism that places the delicate read/write head on a special landing pad to prevent disk and data damage.

### Keyboards

Keyboard arrangement is perhaps the most imposing challenge confronting portable designers, as the accustomed "QWERTY" arrangement of alphabetic and numeric keys is invariably retained. Transportables, because they are larger than laptops or handhelds, can duplicate full-size desktop keyboards with fewer or no compromises. In laptops and handhelds, attempting to reduce keyboard size results in narrower key spacing, dual function keys, awkward key placement, or all three. For example, the location of the cursor control and function keys sometimes differ from the 101-key desktop keyboard (89 keys plus 12 function keys) layout. In addition, a color-coded numeric keypad may be embedded among the character keys, necessitating the use of a special Function.

The Compaq LTE and LTE/286 keyboards place the arrow keys in an "L" configuration rather than the standard inverted "T." The Compaq and NEC models require Function key activation of PgUp, PgDn, Home, and End keys. Consequently, laptop, notebook, and handheld keyboards often require a new set of keyboard skills and reflexes. With limited use, these compromises are tolerable at best, annoying at worst; but with extended use, (e.g., spreadsheet applications) the difficulty may be unendurable—there may be physical pain or accidental loss of data.

In an effort to relieve the frustration, portable designers have added ports that allow full-size, 101-key desktop keyboards or external numeric keypads to be attached.

The popularity of menu- and icon-based windowing operating environments has created a need for integrated pointing devices on portables. With laptops, unlike transportables and desktop microcomputers, it is difficult to use the serial or bus mouse while in transit or in the field. Innovative pointing device alternatives from laptop and notebook system vendors include trackballs, touchpads, and Isopoint devices.

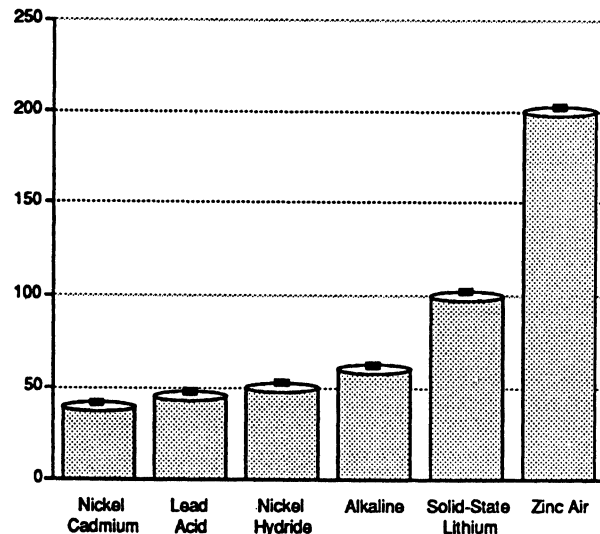
Unlike Intel/MS-DOS laptops, the Macintosh graphical user interface requires a pointing device; consequently, an Apple Macintosh/finder laptop required an alternative to the desktop mouse.

The Macintosh Portable's track ball, essentially a mouse turned upside down, can be positioned on one side of the system or the other providing mouse compatibility to left-handed or right-handed users. A track ball, however, takes up space and does not perform as well as a mouse.

Tawain's DTK Computer Inc. introduced its 16MHz 386SX laptop—the DLP/1—with a touch mouse. Similarly, the Psion MC200 and the MC400 laptops have a 5.5-by 1.9-inch resistive-membrane touch pad above the standard keyboard. The touch mouse allows the user to manipulate the cursor with the pressure of a finger; increasing the pressure increases cursor speed. While the touch mouse is lightweight, it requires considerable surface area.

A new and interesting integrated mouse substitute is the KeyMouse (acquired from Key Tronic). The KeyMouse

Figure 2.  
Energy Density in WATT-Hours per Kilogram



As this graph of power output shows, rechargeable nickel-cadmium batteries (currently favored for portables) and good old alkaline batteries may face stiff competition from the zinc-air, nickel-hydride, and solidstate lithium batteries, which offer combinations of rechargeability, higher power, and longer operating life.

Source: Byte Magazine.

uses the J key on the keyboard as a joystick—the J key is toggled from its normal typing function into joystick mode, and the F key becomes a mouse button. The KeyMouse is implemented on the Quest from Astarte Computer Systems.

The Isopoint, perhaps the most promising mouse-compatible device, allows users to control an on-screen pointer from a small sliding cylinder positioned between two selection buttons in the keyboard. Moving the cursor from side to side involves pushing the cylinder back and forth with a thumb; spinning the cylinder moves the cursor up and down. Invented by Craig Culver of Culver Research (Woodside, CA), the Isopoint was first adopted by Outbound Mac compatible laptop. An improved version can be found on the GRiD 1550sx laptop.

### Displays

Transportable microcomputers, like desktops, have small, self-contained cathode-ray tubes (CRTs) for viewing output. Laptop microcomputers rely on flat-panel display technology as an alternative to the traditional power-hungry, and space consuming CRT. Whereas CRTs fire an electron beam against a phosphor-coated screen to turn each pixel on and off separately, flat panels select a pixel through an electric grid without turning on the adjoining pixels. Flat panels were limited by low-resolution and nearly unreadable characters. The technology has improved considerably (in readability and price), providing a reasonable alternative to CRTs.

There are three kinds of flat-panel displays: liquid crystal displays (LCDs), electroluminescent displays (ELDs), and gas plasma displays (GPDs). LCDs are reflective displays; ELDs and GPDs are light-emitting displays. LCDs, the choice for earliest laptops, are inexpensive to manufacture, are lightweight, and consume less power than either

## Liquid Crystal Displays: A Technology Threatened

While flat-panel displays continue to advance technically, they're facing a major setback on the commercial front. A group of small U.S. manufacturers claims that Japanese companies are dumping flat screens in the U.S., a charge reminiscent of dumping complaints made against foreign makers of DRAM chips in the late 1980s. At that time, the Department of Commerce slapped punitive tariffs on Japanese chip makers that forced the price of memory to skyrocket. Could the same thing happen with flat-panel displays? And if it does, what will it mean for laptop makers—and buyers?

### Seven Against Japan

The group of small U.S. companies behind the dumping complaint call themselves the Advanced Display Manufacturers of America (ADMA). The seven ADMA members are Cherry Display Products, Electro Plasma, Mag-nascreen, Ovonic Imaging Systems, Photonics Technology, Planar Systems, and Plasmaco. In July 1990, the group filed a petition with two U.S. government agencies, the Department of Commerce and the International Trade Commission, claiming that a dozen Japanese companies had sold flat-panel

displays for less than "fair-market value," in violation of the Tariff Act of 1930.

The Japanese companies named in the petition are Toshiba, Hitachi, Fujitsu, Matsushita Electric Industrial, Seiko Epson, Sharp, Hosiden Electronics, Kyocera, NEC, Optrex, Seiko Instruments and Electronics, and Matsushita Electronics. That list comprises a who's who of the LCD screen industry, while the American plaintiffs are small companies that for the most part make non-LCD screens like electroluminescent (EL) screens and gas-plasma displays.

The complaint involves only large flat-panel displays—those with at least 120,000 pixels. It thus includes laptop screens big enough to handle CGA or higher resolutions, but not displays for pocket calculators or those on some home appliances.

### No Price War

Unlike DRAM chips, where U.S. memory makers lost their market in a price war, flat-panel displays have never been a strong point for American companies. The most common displays—and the ones most in demand because of their easy readability and low power requirements—are LCD screens. Liquid crystals were

first discovered by researchers at RCA in 1963, but American companies didn't pursue the technology.

The Japanese companies did, though—in particular Sharp, which introduced a calculator with an LCD screen in 1973.

Since then, the LCD market has largely belonged to Japan. American companies have introduced alternative flat-panel technologies, such as gas plasma and EL displays. Except for sales to the Department of Defense, however, American companies haven't made much headway. Japanese companies have used their lead to develop LCDs into larger and easier-to-read displays: supertwist, activematrix, and most recently, color.

And unlike with DRAM chips, the flat-panel business has seen no price wars. In fact, until recently, the high cost of flat-panel displays made laptop computers significantly more expensive than comparable desktop models. Manufacturers have had trouble keeping up with the demand, especially for higher quality screens. Apple, for example, went to Japan for the Mac Portable's active matrix LCD screen and still had trouble getting the quantities that it wanted.

Other U.S. computer companies also say they've had to go to the Japanese to get LCD technology in the size and quantity they need. And even Zenith, which has bought non-LCD flat-panel displays from U.S. manufacturers such as Planar Systems, says it has had trouble

getting those screens in the quantities it requires.

If Japanese companies have dominated the flatpanel display business for years—and there's more business than Japanese or American companies can handle—why is there suddenly an antidumping complaint? American display makers say prices have dropped so low they can no longer make enough profit to stay in business. But other recent government actions may have as much to do with the petition as the continuing decline in prices.

First, until recently, some American flat-panel makers have contented themselves with selling primarily to the Department of Defense, whose "cost-plus" contracts guaranteed a profit for military suppliers. But because of deficit-trimming budget cuts and the easing of Cold War tensions, fewer military contracts have been available. In addition, some flat-panel makers were hoping they'd get government subsidies to help develop high-definition television (HDTV). Hopes for government support of HDTV development, however, were put on hold when Congress failed to vote for it in 1989.

### Dumpsters?

Are the Japanese companies dumping? That depends on one's definition. The most common definition (and perhaps the least accurate) is "selling a product for less than it costs to produce." Actually selling a product at a loss is unusual, simply because it can't go on for long. If a company loses money on

ELDs or GPDs. Because of their light weight and low power requirements, LCDs have been the prevailing display technology for laptops.

In a rudimentary LCD, each pixel is a single twisted nematic (loosely structured liquid crystal) cell. The cells consist of liquid crystals on glass plates oriented at a 90-degree angle and sandwiched between two polarizing filters, thus forming a *passive* matrix. An electric current causes the pixels to change from a polarization that transmits light to one that blocks light, producing dark pixels on

a gray or white background. The external light source determines character intensity, making an LCD display unsuitable for inadequately lit environments.

To improve readability, LCD manufacturers changed to a "supertwisted" LCD. Supertwisted crystals change the polarization to 270 degrees instead of the usual 90 degrees, providing better contrast. To further improve screen legibility, many laptop displays incorporate a light source in the back to further sharpen characters in dim lighting. Fluorescent backlighting used by all laptops increases the contrast between the image and the display's background. The

every item it sells, it will soon go out of business.

Another definition is "selling products for less than a fair price." What constitutes a fair price, of course, is the big question when it comes to dumping complaints. If a Japanese company sells a display for less in the U.S. than in Japan, is the U.S. price unfair? (The accused companies say their pricing is based entirely on business issues such as market conditions and order size.)

Is the price too low if a company chooses not to soak early customers for the cost of developing the technology and earns back its development costs over, say, a decade instead of a year? (Japanese companies point out that, like many large American companies such as IBM and AT&T, they pay for long-term research over a long period of time. The Department of Commerce, however, insists that "fair pricing" requires that the development costs be paid off in a short time—sometimes as little as a year.)

The Japanese companies complain not only that they're penalized with a double standard, but that they can't even tell when they have violated fairprice guidelines. Under the Tariff Act of 1930 (a protectionist law passed at the start of the Depression in hopes it would shield American jobs from foreign competition), the Department of Commerce refuses to tell foreign companies—or anyone else—the formula used for determining fair prices.

At the International Trade Commission's hearing on the dumping complaint last August, representatives of the Japanese companies also pointed out that LCD screens shouldn't be lumped in with gas-plasma or EL displays because the different technologies have different markets. The ITC rejected that argument. The Japanese makers also argued that active-matrix LCD screens—which aren't made by U.S. manufacturers—should be treated separately from passive-matrix LCDs. The ITC rejected that argument as well.

Several U.S. laptop computer makers testified against tariffs at the hearing, suggesting that they would make LCD displays too expensive to import into the U.S. If that happened, they hinted, they'd have to manufacture their laptop computers outside the U.S., thus dodging the tariffs by importing complete machines instead of the foreign LCD screens. (But those statements may have hurt more than they helped. One Japanese representative commented that the ITC commissioners hearing the case "seemed to feel the computer companies were just being selfish.")

In the end, a panel of four ITC commissioners ruled unanimously that there was evidence that U.S. flat-panel manufacturers were being hurt by the Japanese pricing policies. The case now moves to a Department of Commerce investigation, which may eventually recommend tariffs or other penalties.

#### Tariffs, Higher Prices.

What's likely to happen? If history is any guide, the results of the Department of Commerce investigation are a foregone conclusion. Both Japanese LCD makers and American computer manufacturers now expect that tariffs will be levied, resulting in higher prices for laptops and portable computers made in the U.S. using Japanese flatpanel displays.

As they did when tariffs were imposed on foreign DRAM chips, large American companies that can do so will move product manufacturing overseas to avoid the tariffs, shifting jobs from Texas and California to Taiwan and Korea. But the smaller companies that can't move jobs overseas will not be able to compete.

The tariffs aren't likely to help U.S. flat-panel display makers, since they don't make the technology those large vendors want. But it may have a major impact on prices of laptops from Toshiba, NEC, Sharp, and other Japanese companies. If the Department of Commerce, as seems likely, decides to penalize the Japanese with tariffs on products other than just flat-panel displays—which happened in the case of DRAM chips—all Japanese laptops and portable computers could rise in price. The tariffs might also extend to other products (e.g., LCD-screen TVs).

Another possible effect of tariffs is that U.S. computer makers could face losing access to the state of the art in LCD screen technology for

their own products. If Japanese companies can't sell their own laptops because of high tariffs, they could refuse to sell their best screens to American companies, insisting instead that American vendors contract out laptop production to Japanese manufacturers. That means that U.S. computer companies might have to choose between hiring a Japanese company to build their products or making do with second-rate technology.

In the long run, the cost to produce flat-panel displays—especially LCD screens—will continue to fall. Conventional monochrome LCD screens are getting cheaper to make. Top-of-the-line active-matrix and color screens are still expensive and in short supply, but they will come down in price as manufacturers surmount the learning curve for these products. There probably won't be a glut in LCD screens comparable to the DRAM-chip glut that sparked the memorydumping complaints, since LCD screens can also be used for portable TVs and other consumer products.

Although flat-panel displays are not in the same situation as DRAM chips, they are likely to follow the same path. And in the short run, tariffs will mean expensive laptops and exported jobs—and no real help for the U.S. flat-panel display industry.

This sidebar is a reprint of "Displays—Down the DRAM Drain" by Frank Hayes, from *Byte*, February 1991. Copyright © 1991 by McGraw-Hill Incorporated. Reprinted with permission.

newest laptop displays further improve contrast with fluorescent sidelighting and edgelighting.

LCDs represent color using different monochrome shades, so a screen's ability to display color applications depends on the number of different shades it supports and how it assigns those shades to particular colors. For example, CGA-compatible displays support four shades at medium resolution, 320 by 200 dots per inch (dpi), and two shades at maximum resolution—640 by 200 dpi. Text and graphics may be unreadable if two colors used by an application are assigned the same, or similar, shades. To solve this problem, some portable systems provide palette

mapping utilities for adjusting the default mapping scheme to predefined or user-specified settings. CGA-compatible displays are thinner and cheaper to produce than their EGA and VGA counterparts.

Because battery power is not a fundamental concern in transportables, gas-plasma displays and ELDs predominate in that market segment. GPDs and ELD's provide good contrast and a wide viewing angle. They are also more responsive and more expensive than LCDs.

An ELD consists of a layer of manganese-doped zinc sulfide sandwiched between two electrodes. When voltage is applied, the manganese atoms emit a green-yellow light.

ELD display readability is superior to the LCD's. However, ELDs are difficult to manufacture, consume more power than LCDs, and require higher voltage drive circuits to power the display.

The GPD uses an inert gas (usually a combination of neon and argon) sandwiched between an x-axis panel and a y-axis panel to form a matrix. A pixel is produced by charging wires that intersect at the specific x-y coordinates. When the wires are charged, the gas around the intersection glows bright orange. GPDs require AC power to maintain the glow at a constant level. GPDs draw less energy than an ELD, and images disappear from the screen in significantly less time than with the LCD. This produces a quicker, sharper image that does not depend on any external light source.

### Power Supply

By definition, transportables operate on AC power; laptops require batteries. For laptops, NiCd batteries are the primary power source because, unlike the more common lead-acid batteries which decay at a linear rate, they provide a relatively steady voltage per charge (dropping off near depletion). They are easily recharged and replaceable.

A laptop battery pack is a collection of NiCd cells connected in series. The internal design of a NiCd cell determines its total delivery capacity, its maximum recharge and discharge rates, and the effect of temperature on its operation. When using AC power, laptops replenish the NiCd via a charging circuit (known as trickle charging). When turned off, the charged battery maintains its capacity until the system is used again.

Trickle charging can exacerbate the "memory effect"—a peculiar disadvantage of NiCd batteries. A NiCd battery "remembers" how much power was left when it was recharged. Repeated recharging when a battery is only partially depleted reduces the battery capacity and can render a NiCd battery pack incapable of being fully charged. The memory effect can reduce NiCd capacity by 20 percent. To operate efficiently, NiCd cells must be completely depleted before recharging. The process can sometimes be reversed by putting the battery through repeated charge/discharge cycles.

In the quest to prolong battery operation, laptop designers have devised hardware and software "power management" schemes (e.g., CPU "sleep mode," temporary "screen blanking," and modem and disk drive "power-down") that conserve power by reducing or eliminating current to various subsystems. Some software utilities, which can be keyboard activated, act on the display backlight, processor/memory, modems, and fixed or diskette drives. Some systems provide user-customizable power-saving modes, via the setup program, allowing alternate shut-off time periods. Sophisticated power management techniques prolong the life of the battery packs, which typically must be replaced after a few hundred recharges.

The Compaq SLT 386s/20 provides the Power Conservation Utility (PWRCON), a menu-driven utility that allows the user to set the amount of idle time before the system blanks the screen, puts itself to "sleep," or turns off the hard disk drive. Even with sophisticated power management techniques NiCds average only 3 to 4 hours before they need to be recharged. Such is the case with the SLT 386s/20. Although a fast-charge capability fully recharges the battery in 1½ to 3 hours, the rated battery life is approximately 3 hours.

Toshiba laptops provide a pop-up window displaying the battery status and AutoResume, which is activated at set-up or from the pop-up window, that returns the user to the exact place in the application where he/she was when the unit was shut off. Not having to reload application programs and data files saves battery power. The Toshiba model's optional internal modem can be turned off via the pop-up window or set-up program.

Zenith laptops provide "power rationing" by allowing users to specify which subsystems are required (e.g., disk drive, backlight, keypad, and serial, parallel, and modem ports). A related technique allows clock speed, display time-out, contrast and brightness, and disk spin-down to be managed by the system or by the user.

### Expansion

Internal expansion hinders portability; therefore, both transportable and laptop microcomputers have limited expansion capabilities. With transportables the number and kind of expansion slots is limited only by the enclosure design; they also have expansion buses compatible with their desktop counterparts.

The majority of laptop expansion buses, however, have been vendor specific and cannot work with desktop expansion cards. Options include external expansion chassis attaching to the back or bottom of the system. Even then, the expansion unit is limited to one or two XT- or AT-compatible slots, and the system is usually intended to remain on a desk.

The increasing sophistication of laptops has led to "plug and play" modular expansion chassis that not only add compatible expansion capabilities but also duplicate the features and functions of a desktop system. The laptop system is inserted into a connective shell or base that houses additional diskette and/or disk drives, serial and parallel interfaces, and an AC power supply. It also typically includes external CRT display ports and desktop keyboard ports. The NEC ProSpeed 386 "Docking Station," the Compaq SLT/286 "Desktop Expansion Base," and Toshiba's "Desk Station" represent this kind of laptop expansion technology.

### Operating Systems

Portable microcomputers typically use the MS-DOS operating system, but sophisticated operating systems are not excluded. Recent 80286-, 80386-, and 80386SX-based transportables and laptops are also capable of running OS/2 and UNIX. The GRiDCase 1550sx is bundled with MS-Windows 3. Toshiba offers T/IX, a Toshiba-enhanced version of AT&T's UNIX System V operating system.

### Applications Software

Except for the possible differences in the size of diskette media (3.5 and 5.25 inches), portable computers impose no inherent software restrictions or incompatibilities with desktop software. With adoption of the 3.5-inch diskette as standard desktop media, transferring data between machines is no longer the inconvenience it once was. File transfer programs are available to transmit data via the serial ports of two systems; some programs can also use the parallel ports. Most such programs provide a menu-based user-interface that aids in coordinating the transfer. Some programs allow access to the other system's disk drives; others allow resource sharing of peripherals such as printers.

## Peripherals

While portables take advantage of the same peripherals as desktop systems, the rising popularity of laptops has created a separate laptop peripherals market. Laptop peripherals include every conceivable accessory—external diskette/disk drives, fax cards, network interfaces, modems, tape drives, and printers. System-specific peripherals are not interchangeable, due to incompatible bus structures. External peripherals will work with most laptops and, in some cases, a desktop or transportable.

## Modems

Because laptops are commonly used in business travel, communications and data collection are typical user concerns. Most laptop vendors offer system-specific modems, but third-party sources such as Xircom Inc. offer compact (pocket-size) modems. Modems can link portable computers to the home base or other locations and vastly increase their usefulness. Most of the latest modems operate at 2400 bits per second, and some have Microcom Networking Protocol (MNP), a hardware-based error correction scheme, instead of the slower but less expensive software-based xmodem error correction.

Communicating on the road by modem is not always easy. Hotel phones do not always have modular jacks, and sometimes the hotel's switching causes problems. Foreign telephone systems are especially problematic to the traveler because different countries have different telephone jacks and because the quality of long-distance lines varies. Some travelers route their calls through an electronic mail service to get around poor-quality telephone connections and to save the long-distance charges at hotels. Acoustic couplers that snap onto the receiver instead of plugging into a jack may be needed in hotels and for pay phones (or any phone that cannot be unplugged).

Some vendors also offer compact fax modems. Though the laptop computer's screen may make it impractical to display incoming facsimiles, the cards and software can send relatively high-quality text to any facsimile machine.

## Mass Storage

External diskette drives are available from vendors and third-party manufacturers. External add-on hard disk drives are available in several capacities. Backup capabilities are provided by external tape drives.

Laptop microcomputers now use form factors of 3.5 inches or less, although a few companies (Conner, JVC, Prairie Technology, etc.) now market 2.5-inch Winchester drives. The maximum capacity presently available in 2.5-inch hard drives is 40M bytes. In some instances, 50M- and 100M-byte capacities are available. In the future, 120M-byte capacities will be possible.

## Printers

Portable printers, perhaps the most popular portable peripheral, are experiencing the fastest growth. Last year, 130,000 portable printers were sold, with the expected growth rate at about 20 percent per year. Popular portable printers are available from third-party vendors, such as Axonix, Diconix (a Kodak subsidiary), Hewlett-Packard, and Brother.

The ideal portable printer is light, compact, and rugged; runs for a long time on batteries; and produces good print quality at high speed on ordinary paper. Though no printer has all these capabilities, each printer technology has its pros and cons:

- Dot matrix printers, such as the ones made by Axonix, can print two- or three-part forms, are fast, and use regular paper, but they are noisy.
- Ink jet printers, such as Canon's, Hewlett-Packard's, and Kodak's, are trouble free and quiet, but produce their best print only on special, smooth paper. Canon's BJ-10 Bubble Jet printer provides a maximum resolution of 360 dpi. Because of their superior print quality, ink jet printers are expected to be as successful with portables as they are with desktops.
- Thermal transfer printers are also very quiet, have good print quality, and require little electricity, but the ribbons are expensive.

Most of the portable printers currently available have optional AC or DC power but are power-hungry in comparison to other portable peripherals. Power consumption is a major obstacle, meaning advancements will be slow.

## Other Peripherals

Transportables and laptops can act as standalone microcomputers connected to a network. Network cards are now available, bringing together the data resources of a corporate network and the data collection potential of portables. Toshiba, for example, offers several Ethernet LAN option cards.

More esoteric accessories include acoustic couplers, external numeric keypads, independent power supplies, and car cigarette lighter adapters—enabling a portable to run or recharge in an automobile.

## Future Technology

To date, the history of portable computing has been a quest guided by compromise rather than convenience. Nevertheless, as portable computing enters its third decade, technological advances in three areas—Very Large Scale Integration (VLSI), display quality, and battery performance are moving portables beyond their traditional constraints.

## Integration Technology

As portable designers cram ever increasing performance into an ever smaller package, they are providing solutions that conserve power and space. Advancements in integration technology have increased the number of transistors that can be etched on a single piece of silicon. Increasing transistor densities mean faster switching times, lower power consumption, and enhanced reliability. The 80386SX processor is one such development. Because it requires less power than its 80386DX relatives and reduces space requirements by using surface mount technology, the 80386SX (in 16MHz and 20MHz configurations) has spurred the growth of the portable market.

The tremendous interest in the laptop and notebook portable market has encouraged Intel to develop a version of the 386SX processor designed for laptops and other small computers. In October 1990, Intel introduced the new 386SL Microprocessor SuperSet, essentially a CPU and a corresponding chipset. The SL and its companion 82360 I/O chip operate at 20MHz only, matching the highest speed of the 386SX, which the 386SL closely resembles. The 386SL features enhancements that extend the 386 architecture, without affecting compatibility, to add advanced power management features at the processor level.

Intel has also added a hardware level-only interrupt and a new memory address space to the 386SL. These are reserved for a new interrupt, the System Management Interrupt; with a system management handler, hardware companies can access reserved System Management memory and System Management I/O addresses. These are all fully compatible but invisible to the operating system and application software. The SMI allows suspend and resume operations, peripheral standby, CPU speed control, and uninterruptible power supply capabilities. Despite this extra logic that can remove processor time from the operating system, Intel claims that the 386SL outperforms an equivalent 386SX.

The SL has a main-memory subsystem controller with a 32M-byte address space, a LIM EMS 4.0 memory controller, an AT bus controller, a full cache controller, and support for the 387SX math co-processor. The companion chip, the 82360SL peripheral and power management chip, supports CPU, memory, and peripheral functions, as well as providing programmable features to manage power to prolong battery life. Intel also provides a family of low-power support logic chips.

Currently, a number of chipset manufacturers offer many of the 386SL's power management features in support logic devices designed to be used with the 386SX. However, Intel is now offering more features that chipset companies cannot provide and the ability to link them into the CPU. According to Intel, the approach is inherently safer than that of the chipset manufacturers since it does not have to continually fight with the operating system and applications for control of memory, interrupts, and the CPU.

### Display Quality

LCDs, ELDs, and GPDs share a fundamental problem—limited grayscale presentation. Application software that depends on color forces the display circuitry to represent various colors as contrasting graphics patterns. Initial attempts at color LCDs include passive matrix displays that provide color by placing red, green, and blue filters over each pixel. Color filters, however, reduce the amount of backlighting transmitted through the screen, which can make the display seem washed-out and murky. Passive matrix displays can be found in the following transportables—the NEC ProSpeed CSX, the Toshiba T5200C, and the Sharp Multi-Color 386.

Another alternative LCD technology, however, is emerging as an important development in the evolution of color LCDs—Thin Film Transistors (TFT) displays or active matrix display.

A monochrome TFT, like the one already used in the Macintosh Portable, has one transistor associated with

each pixel. With a color TFT display, each pixel is assigned three transistors (located directly on the plate glass behind the liquid crystal material), one each for the red, green, and blue components. Consequently, unlike passive matrix displays where the individual pixels are controlled by a grid of electrodes, each pixel in a TFT display is activated individually allowing for extremely fast access times with outstanding color reproduction. The size and power consumption of TFT displays precludes them, for the time being, from smaller, battery-powered packages. While small-sized color TFT LCD TVs have been available on the consumer market for almost five years, exorbitant manufacturing costs resulting from low production yields have kept them out of the mainstream market.

Nevertheless, an early example of an active-matrix display is available from Dolch Instruments. Dolch uses a TFT active-matrix display made by Hitachi. The VGA-compatible display is available as an optional replacement for the existing electroluminescent or gas plasma screens on Dolch's 386- and 486-based portables. It carries a heavy price—an additional \$3,995 over the system prices of \$7,995 and \$12,995 for the 386 and 486 models, respectively.

### Battery Power

Because NiCd batteries are a mature technology, engineers are working on several NiCd alternatives: nickel hydride (NiH) and lithium batteries being the two most promising. NiH batteries able to replace NiCds in existing portables will have 50 percent higher capacity (and can be recharged as many times as can NiCds), thereby boosting operating time by as much as 50 percent. To best exploit the higher capacity of NiH, laptop designers will have to make computers that use the same amount or less power.

Another interesting alternative to NiCds are rechargeable lithium batteries. Because lithium is an explosive material, the only lithium batteries manufactured are of the "coin" variety usually found in some cameras, watches, and calculators. However, scientists at the University of California's Lawrence Berkeley Laboratory have developed a new lithium battery that delivers higher power, more recharge cycles, and a longer shelf life than any of the commercial batteries now available. And unlike lithium batteries with a liquid electrolyte, the new batteries won't leak or explode when exposed to heat.

Even with the pending arrival of new battery technologies, most observers say that improved battery performance—especially in the immediate future—will depend on the same criteria valid today: lower power consumption and better power management. ■



---

# Notebook Computers

---

## In this report:

Evolution of the Notebook Computer .....	2
Airline Travel .....	5
Other Travel Concerns .....	6
Mass Storage Tips .....	8
Thoughts on Software .....	8
Notebooks of the Future .....	9

## This report will help you to:

- Be aware of the evolution of the first “portable” computers to today’s light-weight notebook computers.
- Avoid problems when traveling with your computer.
- Know what steps to take if your hotel does not have the proper telephone jack for modem transmission.
- Ensure the security of your data and your computer.

---

## Introduction

The invention of a viable *notebook computer*—one that packs full DOS functions into less than six pounds of hardware—has the potential to revolutionize telecommunications. With the ability to tap into on-line databases, a user has access to information anywhere. The world, suddenly, is smaller. The notebook computer is small enough to be essentially a terminal, connecting any location serviced by a telephone to all forms of information, bringing data to you in the remotest corners of the world, into your hotel room, or an airport lounge.

The Zenith 181, with its backlit screen, used to be a dream come true. It was surrounded by Toshiba, NEC, and Sharp laptops—all of them weighing in at ten pounds or more. The Toshiba T1000 and its less-than-seven-pound heft led manufacturers to think in terms of weight. Now, the wave of the future is this new category: the notebook computer. In the years ahead, no serious telecommuter will be seen without one.

A state-of-the-art notebook computer will certainly be quite different a year from now, as products never stand still. Even at the dawn of the 1990s, though, certain minimum standards and requirements have established themselves, and the consumer should accept no less than the following:

---

This Datapro report is a reprint of Part 1, Chapter 14, “Notebook Computers” pp. 351-366, from *Dvorak’s Guide to Desktop Telecommunications* by John Dvorak and Nick Anis. Copyright © 1990 by McGraw-Hill, Inc. Reprinted with permission.

- A weight of under five pounds;
- Compatibility with existing programs, and the ability to exchange data with larger machines;
- A readable screen, and a comfortable, durable keyboard;
- A battery life longer than the average airplane flight; and
- A price not exceeding that of the average major appliance.

Consumers of the future will probably clamor for a 50MHz, 80486-driven, cellular machine with a 3-D screen that fits in your palm and costs \$499. They will know what a desktop can do, and will want laptops to be capable of the same tasks. Thanks to the trickle-down nature of innovation, laptops will soon have more desktop capabilities, but by that time the desktop will have taken yet another leap. Consequently, there can never be a "perfect" small computer. Nevertheless, the notebook computers of today can handily manage the essential telecommunications tasks, and act as portable data-management terminals that can be attached to any phone.

Notebook computers have gone "over the hump" in terms of power, functions, and weight. This opens up telecommunications to many and diverse new applications. Moreover, even if on-the-go computing has always required compromise, these compromises are no longer debilitating. Your notebook computer may not be as versatile as your home machine, but it performs the tasks at hand well enough so you don't mind.

### **Evolution of the Notebook Computer**

The latest crop of notebook computers evolved from three needs: power, portability, and expandability. Since the beginning, manufacturers have recognized the importance of portability, and have tried—in most cases unsuccessfully—to meet that demand. Sometimes, as with the early GRiD, the machine was too expensive to be popular. An advantage of notebook computing is effortless transportation to remote places; but no one will attempt to scale a mountain with a \$10,000 machine in a backpack.

In the past, the "portable" label was subjective. Osborne Computers, years ago, scored big with a so-called portable machine weighing 20

pounds. The "portable" designation worked only if you were a weight lifter. Compaq's first success was with a similar machine, and one of IBM's first PC failures was a Compaq look-alike called, deceptively enough, the PC Portable. And let's not forget the Otrona, probably the best designed machine among these "luggables." These products missed the mark because of perceived, rather than actual defects. They ran well enough, but most users left them at home on desks, and moved them maybe twice a year. In addition, the standard five-inch screen was uncomfortably small, and purchase of an extra monitor ultimately destroyed the cost effectiveness of these early portables.

### **No More "Luggables"**

These older portable PCs are now found at swap meets, and sold by second-string companies. Desktop PCs, in general, continue to become smaller, if still not easy to move.

"Portable" evolved to "laptop." The laptop label is also a misnomer. Generally, it is applied to machines that theoretically can be used when sitting down, with your knees as a surface. In practice, however, few people use a laptop on their laps; the posture is hardly comfortable or efficient for typing. It's particularly difficult to use one in a plane, seated in coach, when the seat in front of you is leaning back—but anything is possible. For the sake of argument, we can say that the laptop category includes machines that run standard software and have a clamshell-like design. Beyond this, the category has seen more variety than almost any other personal computing genre.

Throughout the development of laptops, the common denominator has been compromise—between power, portability, and expandability. The most significant problem has been weight; even a 15-pound machine seems unbearably heavy by day's end. Even in the most capable units, the battery life was too short, or the screen was too hard to read, or the keyboard was too small. In some of the more powerful laptops, batteries were no longer used at all; manufacturers assumed that anyone who needed that much computing power would always be close to an electric socket. Finally, laptops have never accommodated expandability to any satisfaction.

A laptop's power now approaches that of a desktop PC, but users generally pay a 20 percent premium for portability. Some kind of trade-off is

always needed, and so far there is no one computer that combines the Big Three of *power*, *portability*, and *expandability*.

### Enter the Notebook

Notebook computers, used primarily for gathering data and sending short files, are nothing new. For years, newspaper reporters have toted around remote terminals to transmit late-breaking stories into their newspaper's computer system. No wonder Tandy/Radio Shack's Model 100 was so popular when it was introduced. Journalists accustomed to dragging around an expensive clunker called a "bubble-term" loved the Model 100. It was lighter, easier to use, and allowed limited storage and retrieval of data as well.

Next came the Convergent Workslate, which has recently popped up in surplus outlets for less than \$200. It was a more powerful, but less successful solution. Perhaps the electronic Filofax concept was ahead of its time, or the push-button keyboard was too hard to use. There was a reasonable range of software, loaded through a microcassette, but the LCD screen was too much of an eyestrain for anything besides occasional reference.

While the Model 100 found its niche and the Workslate disappeared from view, no significant notebook-style computers appeared for several years. In 1988, however, British miniaturization king Clive Sinclair came up with the Z-88. Though decidedly nonstandard, the Z-88 was a step in the right direction, and was a bridge between the Model 100 and the full-function notebook units of today. It had no disk drives and could not run DOS software, but files could be easily exchanged with databases on larger machines. The Z-88, which continues to sell, had a sleek look and light weight. Still, its inability to store and back up information on-site prompted users to pay a little more for the admittedly heavier standard DOS laptops.

NEC's UltraLite, announced in late 1988, brought several new issues to the table. Weighing less than five pounds, and less than two inches thick, the UltraLite was not only easy to carry, but could run standard DOS software. It included a modem. The UltraLite also promoted the idea of the silicon hard disk. On this disk, up to 2MB of RAM was configured as nonvolatile storage, which could be saved even when the machine was turned

## When There Is No RJ11 Jack in the Hotel

Portable computing pioneers by necessity have become telephone mechanics, learning how to dissect and rewire various systems on the fly. But this is not how you want to spend your traveling time, especially when there is the danger of crossing wires and disabling the phone altogether. An acoustic coupler, a unit that clamps over a phone's mouthpiece, will allow data transmission, but it adds weight to your luggage and is a hassle to use.

A better idea is to go to Radio Shack and buy a phone line with an RJ11 plug on one end and four alligator clips on the other. In hotels where the phone is hard-wired, you have to open the phone with a screwdriver and hook two of the alligator

clips to the innards. It's not that hard. First unscrew the screws (usually two) on the bottom of the phone. Then take off the top carefully; do not unhook the usually red "message waiting" light. You'll see two to four wires coming out of the wall cord that goes into the phone. Find the green one and the red one. To these, fasten the alligator clips from the matching green and red wires on the cord you purchased. Then pop the RJ11 connector into your laptop modem. Make sure the receiver is on the hook. It can be balanced upside down on the metal bracket that goes up and down and which hangs up the phone. Now, using your regular telecommunications software, you can use the phone with your modem.

off. This hard disk, too small for desktop use, was still large enough to meet most laptop needs.

The UltraLite heralded the debut of two significant laptop features: a silicon hard disk and an embedded file transfer utility. Both will become standard and necessary equipment on the notebook and laptop computers of the future. The silicon hard disk turns RAM into a nonvolatile, battery-backed storage device, and at 2MB it is large enough for fundamental laptop data functions. The transfer utility (UltraLite uses Traveling Software's LapLink) enables files to be sent from a laptop to a desktop through the respective serial

ports. This is a more efficient transfer process than copying to and from a floppy disk.

The UltraLite has no installed floppy drive; rather, it uses ROM cards and RAM cards for file and software transfer. These cards have the potential for efficient data transfer and storage, and could evolve into a software distribution medium.

Despite some disadvantages and nonstandard behavior, the UltraLite is clearly a strong product that increased the ante in the notebook computing game.

### Getting Lighter

Since the UltraLite bid, the Zenith MinisPort saw and raised the bet. (IBM has been notably absent from the game, and innovation has sprung from the NEC/Toshiba/Zenith triumvirate.) Some innovations that seemed like good ideas have been dropped for more familiar approaches; other experiments are being tried. If you step back a bit, the market sorts itself out. It is moving toward smaller, faster, more practical machines with more storage, better keyboards, and brighter screens.

One innovation dropped is the Zenith MinisPort's internal 2-inch floppy diskette drive. This was a nice little drive with a familiar, reliable means of on-the-road data exchange, but Zenith has replaced it now with an internal 20 or 40MB hard disk. There is still an optional external floppy drive for those who need one.

Laptops, as standard DOS machines, have always been touted as alternatives to desktops. In this endeavor, they are already a step or two behind. Notebook computers, on the other hand, have no such burden; no one expects a notebook computer to do everything that a desktop PC can. The appearance of functional notebook computers has changed the concept of portability.

The notebook machines of the future will remain an adjunct to another system. File transfer to a larger computer will take place over a little antenna using low-power radio. (The Japanese have already perfected the technology of wireless RF serial port connections.) We can expect to see the 2½-inch hard disk crop up in superlight machines offering 20MB to 40MB of inexpensive storage, thus turning the notebook computer into a miniature powerhouse.

While the first laptop users traveled in style and faced none of the inconveniences of covered

wagons and dust storms, they were nonetheless pioneers. Taking a computer on a trip was exciting, but you also carried along enough obstacles and inconveniences to make you wish for the good old days of pen and paper.

For instance, I remember a colleague who decided to take his PC Portable to a trade show. The machine required electricity, so using it on the airplane was out of the question—even if the seat-tray would have supported it. The computer barely fit under the seat, to the aggravation of my friend and everyone around him. After landing in Las Vegas, where the flight gates are about six miles from the taxi stand, he weathered his first strenuous physical workout of the day. The next obstacle course was at his hotel; when he arrived, there was no porter in sight. By the time he got to his room, he was too exhausted to compute. In the end, he was able to use the machine for only ten minutes at the show.

Four years later I saw him at the same trade show—this time toting a Zenith SupersPort. He carried it to meetings and around the show floor. However, even at half the weight and twice the power of the PC Portable, the 15-pound SupersPort became quite a burden at day's end.

In the future, then, a notebook computer could be the final trade show solution. It's actually lighter than the press releases you collect on an average day. And carrying it around for a 15-hour stretch won't make your arm longer.

### Coming Down the Pike

Some of the newest machines go far beyond the requirements we suggest for notebook computers of the future. Here are some thumbnail sketches of products that are paving the way to the future.

#### Compaq LTE and LTE/286

Weighing in at six pounds with an 8½- by 11-inch footprint, the new Compaq LTE and LTE/286 notebook PCs are hot. They are a full two-thirds smaller than Compaq's last version, the SLT/286. They have CGA-compatible, backlit supertwist displays, and NiCad batteries that are good for about 3½ hours. Optional battery packs, backups, and quick chargers are available for those who can't just plug into a standard AC outlet before fade-out.

The LTE models are the first notebooks to provide high-speed 20 or 40MB fixed disk drives; a

3½-inch, 1.44MB diskette drive; an 80-key keyboard with standard key spacing; an embedded numeric keypad and 101-key compatibility; and an optional built-in 2400 baud modem.

Both LTE models are the same size and weight. The 286 operates at 12 MHz and can handle all standard applications, including spreadsheets, cost analyses, and account profiles. The slower 9.54 MHz 80286-based LTE is designed more for basic applications like word processing, electronic mail, inventory tracking, and customer databases. Both machines have 3½-inch drives, can read off-the-shelf commercial programs, and can store and transfer data to desktop machines in a standard format.

### **Toshiba T1000SE**

Beating the six-pound goal by ounces, Toshiba released its new T1000SE as the new decade opened. The machine's 1.78 by 12.4 by 10.2 inch size fits nicely into most briefcases. The built-in 11½-ounce nickel cadmium battery is included in the total weight and lasts about two hours before it needs recharging. The AC adapter only adds an extra 10½ ounces.

The T1000SE has a full-size, 82-key keyboard that is pleasant to use. Expansion capabilities abound. The industry-standard 1.44MB, 3½-inch diskette drive makes the machine compatible with existing systems.

### **Poqet**

This little machine gives full service to DOS applications and has a full-size keyboard (or close to it), but would fit in some coat pockets. The Poqet weighs in at just about one pound, making it by far the lightest of the full-service notebook machines. It measures 8¾ by 4¼ inches and can run for up to 100 hours on two AA batteries! It also uses self-powered static RAM and ROM cards, like the ones used by the NEC UltraLite, rather than dynamic RAM. The LCD display is a full 80 columns by 25 lines and has a glareproof surface. The machine has built-in parallel and serial ports.

Poqet is working with programmers to develop ROM-card versions of major applications. These include WordPerfect, Lotus 1-2-3, and Lucid 3D. AlphaWorks plans an integrated package combining a word processor, a database compatible with dBASE III+, a Lotus 1-2-3-compatible spreadsheet with graphics, and a communications

module. The machine comes standard with DOS 3.3, BASIC, PQ-Link (a file transfer ROM utility), and a null modem cable that attaches between your desktop PC and the Poqet's proprietary expansion port.

The Poqet's primary disadvantages are its keyboard and the lack of a built-in modem. If you buy the Poqet, be sure to pick up a plug-in modem. The keyboard was an engineering trade-off. To keep weight down, they chose a nonmechanical keyboard that feels cramped and uncomfortable until you get used to it. Maybe they'll pay the two- or three-pound penalty and offer a future machine with a more familiar mechanical keyboard.

Despite these minor drawbacks, the Poqet is an altogether impressive feat of engineering.

---

### **Airline Travel**

In the past few years travelers have learned a lot about portable computing—mostly through trial and error. As computers have become lighter and more powerful, the standards for laptop travel have also developed.

Airline travel, for which portable/laptop computing was designed, has its own set of foibles. Should you travel with a laptop, in most cases you'll want to check the accessories—external disk drive, adapter, portable printer, and other peripherals—in your baggage. (The exception to this rule might be the adapter. If you are stuck in an airport on a layover, it would be wise to plug into an electric socket and save battery life.) In general, though, carry as little as possible on the plane itself.

Knowing what to check through and what to carry is only the beginning. Airline policies, international travel restrictions, varying power standards, and hotels inhospitable to modems can all become irritants. To defuse troublesome elements, you need to plan ahead.

Your planning can be as fundamental as researching the carry-on restrictions, and determining how strictly the airline adheres to the two-bag limit. If you want to travel light and avoid checking anything—usually a wise move—carrying a garment bag, briefcase, and portable in a separate case will violate the rules. A lot depends on how zealously the flight attendant decides to enforce the letter of the law. Try using your laptop bag as a briefcase.

While jamming a 12-pound laptop into carry-on luggage presents a logistical nightmare, notebook computers solve this problem. Few flight attendants will nail you for a 12- by 10-inch unit slung inconspicuously across your shoulder. In that unlikely instance, it's easy to slip the notebook into a briefcase, or even into the folds of a newspaper.

It is a fallacy that portable computers emit signals that can interfere with an airplane's operation. Most airline employees, fortunately, know this. Still, the pilot controls all activity on the plane, and if for some reason computing is banned on a particular flight, there is nothing you can do.

Airplane laptop users may face a controversy that has nothing to do with the two-bag limit, and that is the possibility that a portable computer can contain a bomb or other terrorist implement. There was an incident in which a portable stereo, checked through as luggage, destroyed a plane in midair. This tragic occurrence has resulted in attempts to prohibit passengers from carrying any electronic equipment on flights.

Realistically, the hazards of traveling with your laptop are pretty unspectacular. Aside from packing correctly, you'll have to deal with getting through security safely, protecting yourself against theft, and finding a comfortable place to work during a layover. As for security, increased or not, the standard X-ray scan will endanger neither the machine nor the disk data. A less commonplace magnetic detector, on the other hand, might affect your data. When in doubt, ask. (Here is yet another reason for you to back up all your files, all the time.)

### **Theft**

Notebook computers have a sleek, expensive look. As a result, they are prime targets for airport thieves and other shifty characters. The obvious rule of thumb is to keep your unit in sight, the same way you would your camera, but that doesn't always work. A thief only needs a second's distraction. The fact that your laptop is not only expensive hardware, but also contains valuable data, makes extra security measures imperative.

The first security precaution you should take is the all-important data backup. The best procedure is to find a backup system and stick to it. A good one is to copy all data files to a floppy disk at the end of each day, and then mail that disk back to the office. Always copy to a floppy, and keep that disk separate in a place that won't attract a

thief—like inside a pair of clean socks. If everything is backed up, a stolen computer won't be quite as heartbreaking a loss.

Protecting the computer is part common sense and part preventive skill. For instance, don't make a conspicuous display of your computer. Another important defense is securing the computer to your luggage, or even to your arm. Sometimes this can be done with the handle, although some of the better-equipped laptops supply a security bracket that can be used to lock the unit to an immovable object. (If your laptop has neither a handle nor a security bracket, you might want to avoid carrying it "downtown" at all.)

### **Encryption**

For those who feel there can never be too much security, look in the back of the major computer magazines for ads about data encryption and security software. For example, Secret Disk, from Lattice Software, will encrypt your files so that they are unusable unless a password is entered.

---

### **Other Travel Concerns**

A notebook computer does not pretend to accomplish all desktop tasks; rather, it is the first personal computer in history with the role of a "second computer." It won't, for instance, handle color graphics or give you access to a million records. An effective notebook will give you just enough power to solve the inevitable on-the-road computing problems. So it makes sense to plan ahead for what your laptop doesn't include.

### **Printers**

One typical travel problem is the need to output hard copy. Portable printers are available, but you'll have to carry around at least another four to ten pounds of inefficient weight. In most cases, you won't need to print much material on the road, but when you do, it helps to call ahead. Many hotels and airlines offer access to printers. For instance, you can stop by United Airlines' Red Carpet Club in O'Hare for a beer and a printout.

These airport clubs aren't free; "initiation" fees are \$100 and up, with similar annual costs. Still, the first time you're rescued from the hard plastic seats of an airport waiting room, you'll consider the investment worthwhile. The alternative, in many cases, is an expensive hotel room.

Airport travel clubs usually can accommodate a modem, enabling you to telecommunicate en route. Once you arrive at your destination, though, you might not be so lucky. Although this situation is changing, don't assume that the phones in your hotel room will allow easy access to a modem. To prevent theft, hotels have bolted everything down, so in many cases the phone is hard-wired to the wall. There has been some progress; some of the larger hotels now include the standard RJ11 jacks in the rooms. As a courtesy to portable computer users, many of the classier places provide an extra open jack.

### Phone Hassles

Solving the hardware struggle may only get you halfway there; a hotel room phone is a breeding ground for software incompatibilities. Telecommunications services like MCI Mail supply toll-free numbers that work throughout the country. As a result, you only have to program your modem to dial one number, regardless of where you travel. That helpful consistency goes out the window as soon as you check into a hotel. Most hotels use the dial number 8 to access a long-distance or toll-free line, but this is hardly universal. Another wildcard is whether a 1 is required before a long-distance call. To accommodate changes you won't know about until you arrive, an on-the-fly mutation of your dial-up file is required.

For this, it helps to choose a communications program that lets you accomplish these dialing tasks easily.

Aside from phone oddities, the notebook computer user must be aware of power aberrations. Some aspects of this problem are only of concern to the international traveler. You'll need to do some research to find out whether your destination country's power standard is consistent with your notebook computer. Happily, most countries share a consistent standard that often extends throughout the entire continent. The French-German connector, for instance, works in approximately 80 percent of Europe and the United Kingdom. Places like Australia, New Zealand, Hong Kong, and Singapore share this standard (perhaps because of their current or former colonial status).

### What to Bring

Buying more than one adapter is unnecessary, because multiadapters that can accommodate many

different standards are available. For a few dollars more, you can pick up a unit that channels electricity from wall plugs throughout the world to your notebook computer. Check with your dealer about the adapter/charger you already own; most automatically switch to and from 220 volts. (Few say so on the box, though, because UL will not approve the idea.) The newest Zenith power supplies, for example, typically work on 220 and 120 volts.

A cheaper tool, the three-to-two prong converter, may be needed to alleviate incompatibilities in the good old U.S.A. Those irritating two-pronged wall sockets are still around, and they tend to turn up in your room when you are most in need of a recharge. An inexpensive converter is one little lifesaving necessity that should be packed and ready to go with your notebook computer.

Another light and essential item is a length of phone cable (the hotel may supply a jack, but don't bet on the wire). Throw in a list of local logon numbers for your favorite E-mail service, in case the toll-free number doesn't work. And don't forget the system manual.

---

### Battery Tips

Extra floppy disks and battery packs are a good idea. These items add flexibility to media portability and power, making the notebook computer more versatile than its predecessors. Access to a removable battery gives a portable computer an almost unlimited operating time; you are restricted only by the number of packs you can bring along. Based on a three- to five-hour operating time per pack, and "average" computer use time, a practical estimate of battery needs is two packs per day. (Should your task involve traveling to the remote wilderness to compose a *Call of the Wild* sequel, that estimate may increase. In any case, the necessary batteries will surely weigh considerably less than the daily food ration.)

While using a notebook computer, you always need to keep one eye on how much power is being consumed, in the same way you need to stay aware of how much gas is in your car's tank. The smartest power management strategy is to always carry a spare charged battery. An intelligent system like the Zenith MinisPort will do some of the work for you. On this model, nonessential functions like serial ports, controllers, and disks shut off automatically when they are not in use. In addition,

you can turn down the video display, turn off the modem, or decrease the processor speed. Cutting the MinisPort back to 4.77 MHz adds another 30 minutes of battery life. Another way to save power is to use fewer programs that constantly access the disk. For instance, disabling any auto-save facility will save power. A disk cache is also a good idea.

A company's battery specifications statistic is similar to a car's miles-per-gallon rating, and should be approached with similar caution. You might not get the sticker rating under standard conditions. Zenith rates the MinisPort's cassette battery life at three hours, using full processor power and full-screen illumination, and estimating a higher-than-average level of disk access. Unlike other laptops, it's not necessary to let the MinisPort's battery run down completely before charging it again. The MinisPort's three-hour battery rating is at least close to the Holy Grail of power duration: the length of a cross-country plane trip. If you take battery-saving measures, and turn the machine off while you eat and read the paper, there will be just enough power to do the job.

---

### Mass Storage Tips

One aspect of notebook computers that is new to the genre is the "silicon hard disk," a parcel of battery-backed RAM that acts much like a standard hard disk. This technology has a good deal of potential; comfortable traveling capacities of 10MB and beyond are within reach. For the time being, however, a notebook will offer little more than 1MB of usable space for applications and data. Although you can never have too much space, this is at least a workable minimum.

Developing a workable notebook-computing software environment is a trial-and-error process, and may take a few tries to perfect. Like any computing setup, it depends on what you want to accomplish. You'll want to make the notebook setup as much like your desktop system as possible, and you can use your home system to refine that setup. First, copy to a separate subdirectory all the programs you want to use on the notebook, and see how much space they occupy. Then delete all files that aren't essential to the program's operation, such as printer drivers, thesaurus modules, and

some overlay files. After making sure that the abbreviated programs function adequately, transfer the whole subdirectory to the notebook with the file transfer utility.

---

### Thoughts on Software

Any self-respecting notebook computer will have one of these programs built into ROM: LapLink, FastLynx, or the Brooklyn Bridge. With these utilities, after the serial ports of a laptop and desktop are connected and the linking software is loaded, both machines' directories are shown on the master system. You highlight the files you want to transfer, and enter the appropriate commands. It is only slightly more complicated than copying data from one disk drive to another.

For a notebook computer to be used in word processing and telecommunications, I assembled a workable, powerful, and compact environment with QEdit and Telix. Try to use the most efficient programs on your laptop. Sometimes it's better to use an editor like QEdit, with all its laptop features such as the big block cursor, than it is to use a regular word processor like XY-Write. You can always transfer the file later to a desktop, and do the final edit with more powerful software.

Another common use for laptops will be data storage and retrieval. For this you'll need some kind of database manager. Unfortunately, though, to manage any standard dBASE-style program, you'll need a hard disk-driven laptop. For the time being, such applications are beyond a notebook's power.

One common application will be to manage a phone list of the people you need to call while on the road. For this, the original SideKick, General Information's Hotline, or Prodex from Prodex Development will all work well. Prodex, which needs less than 100K, has the added advantage of attaching a note field to each name or data entry.

The magic numbers for memory are about 500K if you're using the internal disk to store applications, or 720K if you opt for the floppy. If you want to install a complex environment onto a notebook, such as one that includes a database facility, you will probably need a floppy for each application (word processing, communications, and database).

The ideal software solution for a notebook computer is an integrated program that combines



all needed functions into one neat package. About the best choice is Microsoft Works, an inexpensive four-in-one program that uses about 430K without the dictionary. (Software Publishing's First Choice is similar, and is actually easier to learn. However, First Choice uses over a megabyte, making it a poor choice for the modern notebook.)

WordPerfect Executive is another good option, although it has its own set of drawbacks. Though it features a WordPerfect look-alike word processor, card file (another name for a database), spreadsheet, and scheduler, it has no communications module. But if file and command compatibility mean a lot to you, then the WordPerfect solution is a nifty little package, and the entire program fits on one 720K floppy. You can keep the "main" applications based on a single disk, and use the internal disk for your communications program and data.

---

### Notebooks of the Future

The "notebook" label, as it applies to computers, is no accident. To extend the paper metaphor, this computer is like the spiral-bound tablet that fits in your pocket and contains the day's jottings. When you get home, you transfer all the information to a larger sheet of paper, where it is easier to edit, manipulate, and visualize. There are also a number of

tasks that cannot be easily performed on the computer version of a notebook. A notebook computer, however, allows the data used by the home machine to be gathered with more precision and accuracy.

No doubt innovation will continue to trickle down from the desktop; hard disk-driven notebook machines should be common by the end of 1990. While it's impossible to guess what a notebook will do a decade from now, you can expect technologies like voice and writing recognition, cellular modems, and facsimile transmission to become integrated into portable systems in the next few years.

Today's notebook computer already gives you more freedom—unchaining you from your desk and letting you compute in the sunshine. Notebook computing now allows information to be intelligently sent and received from every corner of the planet, or from the heart of the world's biggest cities, with the same machine. And remember—if you ever need those files sitting at home on your desktop, you can always have your telecommunications program preloaded at home, and put the computer in Host mode. You can then call it from anywhere with your notebook computer, and download the files. ■



# An Overview of UNIX Microprocessors

## In this report:

Motorola.....	2
Intel .....	3
SPARC .....	3
Precision Architecture (PA) and Prism .....	4
MIPS Computer ...	4
IBM POWER .....	5
Intergraph Clipper.....	5

## Datapro Summary

Recent microprocessor developments in the UNIX environment are reviewed. The report discusses the two types of UNIX microprocessors—CISC (complex instruction set computing) and RISC (reduced instruction set computing). Strategies of several leading microprocessor vendors whose products dominate the commercial multiuser and technical workstation markets are examined.

AT&T Bell Laboratories created UNIX using PDP-11 minicomputers from Digital Equipment Corp. For many years, UNIX was confined to a limited number of architectures and applications. However, as improvements were made to its operation and its usefulness became fully understood, UNIX' popularity soared. Now it supports a full range of microprocessors and architectures. Indeed, UNIX has become one of the most popular operating systems for new microprocessor-based hardware platforms.

## CISC and RISC Design

There are two types of UNIX microprocessors. The first UNIX microprocessors were based on CISC (complex instruction set computing) design. In the mid-1980s, the dominant CPU for UNIX workstations was Motorola's CISC 68000 family. While the 68000's use has declined considerably over the last few years,

it remains the most widely used CPU for UNIX workstations. Intel is the dominant supplier of CISC microprocessors for PCs, where UNIX has become increasingly popular.

The other major type of UNIX microprocessor design is RISC (reduced instruction set computing). The first RISC architecture was developed in 1983 by Hewlett-Packard but it did not become a force until the late 1980s. RISC platforms have contributed heavily to the recent success of UNIX, especially in the technical workstation market, where UNIX is most popular. RISC microprocessors are relatively inexpensive but powerful processing elements well suited for tasks performed on workstations. Every major UNIX workstation vendor now bases at least part of its workstation line on RISC architecture. According to most surveys, the number of RISC-based platforms now equals the number of 68000-based platforms (in 1987, the height of its popularity, nearly 70% of the UNIX workstation

market was based on the 68000). Many analysts predict that RISC will surpass the 68000 by the end of 1991.

RISC architecture, in contrast to CISC designs, has limited hardwired instructions, most of which execute in a single machine cycle that can be strung together to perform complex operations. Furthermore, RISC architectures use a high-performance memory subsystem and use optimizing compilers, which absorb some of the overhead previously provided by larger instruction sets and drive parallel processors.

With a single execution level, RISC machines offer a much higher performance than CISC platforms, which require extensive use of microcode instructions to operate. Because UNIX is an extensive and complex programming environment, RISC's single level execution strategy makes it a natural partner. In addition, most RISC designs have built in the demand paging memory management model that is widely supported by most versions of UNIX. Consequently, most of the suppliers of RISC processors and their systems customers are strong supporters of UNIX.

### Motorola

For several years in the mid-1980s the 68000 was used by most of the industry's leading UNIX hardware vendors, including Hewlett-Packard and Sun

Microsystems. However, many vendors have abandoned the 68000 over the last couple of years in favor of a RISC architecture, which provides superior performance for complex tasks. Motorola responded by introducing the RISC 88000 in 1988, but most of its former customers were already using other RISC architectures. Some vendors, such as Sun, invented their own RISC technology while others incorporated a RISC solution developed by other manufacturers. By contrast, very few vendors have adopted the 88000 and, while 68000-based workstations still outnumber all other architectures, its market share continues to erode. Motorola has tried to recapture its share with the 88000 but current trends suggest that this will not happen in the near future.

In addition to RISC competitors, Motorola faces intense pressure in the UNIX market from Intel's 32-bit 80386 and 80486 microprocessors. Motorola's principal advantage, a large base of binary-compatible UNIX software, has been adversely impacted by 386- and 486-based multiuser systems. These are capable of running MS-DOS programs as a subfunction, thus supporting the extremely large number of DOS applications.

**Table 1. An Overview of Popular Microprocessors Capable of Running UNIX**

Architecture	Type	No. of 3rd-Party Systems Suppliers	No. of 3rd-Party Chip Suppliers	Notes
HP (ex Apollo) PRISM	32-/64-bit RISC	0	0	Used in HP Domain DN Series 10000
IBM POWER	32-bit RISC	0	0	Used in IBM RS/6000
Intel 80386	32-bit CISC	over 200	0	
Intel 80486	32-bit CISC	over 60	0	
Intel i860	64-bit RISC	15	0	Often used as a co-processor
Intergraph Clipper	32-bit RISC	0	0	Formerly Fairchild's Clipper
MIPS 2000/3000	32-/64-bit RISC	22	5	
MIPS 5000/6000	32-/64-bit RISC	5	5	Faster versions
Motorola 68000	32-bit CISC	over 100	0	The leading UNIX platform to date
Motorola 88000	32-bit RISC	about 25	1	
Sun SPARC	32-/64-bit RISC	10	5	

### The 68000

The 68000 was instrumental in bringing UNIX to microcomputers and in creating the UNIX workstation market. As the first true 32-bit architecture available for microcomputers, the 68000 offered the performance that porting UNIX to such an architecture demanded. Other features, such as virtual memory management, became incorporated into UNIX as a result of being part of the 68000 architecture. The original workstations from Apollo Computer, Sun Microsystems, and Silicon Graphics were based on the 68000.

The 68000 microprocessors feature a 32-bit register set and linear address space of 4 gigabytes. The current product line supports the following facilities:

- Floating-point co-processors (68881 and 68882).
- Memory management units.
- Direct memory access control.
- Data communications and protocol handling.
- Network control.
- System board interfaces.

### The 88000

Motorola introduced its RISC implementation, the 88000, in 1988 in response to the proliferation of RISC architectures by both rival chip manufacturers and hardware vendors. Although support for the 88000 is currently low, an independent trade group, 88Open, was formed to promote the 88000 in the UNIX market.

The 88000 microprocessor group is based on the 88100 execution unit, which incorporates 320 bit registers, data paths, and addresses. The initial product runs at a 20MHz clock speed, with 25MHz versions now also becoming available, and these enable the 88000 to compete more directly with widely licensed RISC implementations like Sun Microsystems' SPARC and MIPS Computer Systems' R2000 and R3000.

In an effort to retain as many 68000 users as possible, Motorola created a cross architecture binary compatibility standard for the 88000 and 68000. A binary standard expands the software base of applications running on systems that use the same processor. This builds into a large library of existing applications available to any system running UNIX conforming to the standard.

### Intel

Intel is the dominant supplier of CISC microprocessors for PCs. Until the last few years, this has meant that Intel's products were used almost exclusively in the MS-DOS and OS/2 environments. However, the popularity of UNIX on PCs has increased considerably, largely due to the developmental efforts of companies such as The Santa Cruz Operation and Interactive Systems, both leading providers of UNIX versions for Intel-based platforms. The development of graphical user interfaces, similar in operation to the Macintosh interface, has also contributed to the success of UNIX on the PC.

Intel's market dominance is based on the company's 80X86 product line. The most recent versions are the 80386 and 486. The microprocessors are the basis for nearly all IBM and compatible personal computers. Other UNIX vendors that use Intel include Sequent Computer, Unisys, Olivetti, NCR, Altos, and Siemens. Intel has made no significant advances in RISC microprocessor technology.

### SPARC

The most aggressive provider of RISC technology is Sun, the leading producer of UNIX workstations. While the original Sun workstations were based on Motorola's 68000, nearly all of its current platforms are based on a RISC design called Scalable Processor ARChitecture (SPARC). Sun began developing its SPARC technology in 1985. In what some analysts view as a questionable strategy, Sun has licensed its SPARC technology to several hardware manufacturers, including some competitors. Sun's goal is to make SPARC an architectural standard in the UNIX workstation market, thus encouraging independent software companies to produce applications programs for this environment. The SPARC technology (and Sun's licensing program) has been so popular that an independent consortium, SPARC International, was founded to encourage the development of SPARC. The consortium currently has more than 140 members. Hardware vendors shipping SPARC-based products include Solbourne Computer, ICL, Toshiba, and Tatung.

SPARC's major features are:

- Simple instructions, which generally require one arithmetic operation.

- Few and simple instruction formats and three basic instruction formats featuring uniform placement of opcode and register address fields.
- Register intensive architecture.
- A larger "windowed" register file that allows compilers to cache local values across subroutine calls and provides a register-based parameter passing mechanism.
- Delayed control transfer.
- Single cycle execution that allows most instructions to execute in one cycle.
- Concurrent floating point and interface.

### **Precision Architecture (PA) and Prism**

PA, which Hewlett-Packard began developing in 1983, was one of the first RISC architectures developed. PA is the basis of Hewlett-Packard's Series 9000 workstations. However, many of the vendor's other workstations are still based on Motorola's 68000. Hewlett-Packard has licensed its PA technology, but the program is very limited.

Apollo Computer, acquired by Hewlett-Packard in 1989, has developed a RISC architecture called parallel reduced instruction set multiprocessor (PRISM), a cache-based, shared, virtual-memory multiprocessor design. The system's integer processor (IP) features an interlocked, multistaged pipeline that ensures single cycle execution of virtually all instructions. The IP is a 1.5 micron CMOS VLSI design, the floating-point unit (FPU) features a semicustom CMOS register file and a register file and ALU that use emitter couple logic design.

Current PRISM configurations support up to four processors by design, although the architecture supports more. PRISM includes a 64-bit instruction path leading from the instruction cache to the two processing units, thus allowing an optimizing compiler to schedule two instructions for dispatch and execution per cycle. In the PRISM architecture, the floating-point processor receives instructions from a cache, freeing the integer unit for other operations.

PRISM includes standard RISC features such as single cycle load/store instructions, fixed length instructions, and delayed branching. Load/store instructions are implemented on the integer unit, compound instructions on the FPU. Each CPU includes an independent FPU, and the architecture uses large dual caches for data (64 kilobytes) and

instructions (128 kilobytes). Internally, PRISM allows simultaneous IP and FPU instruction dispatch and parallel execution of up to three operations. Communications from CPU to main memory and other processors is through a 150 megabytes per second 64-bit bus. Internal 64-bit data paths allow completion of IEEE 754 double precision floating-point operations in one clock cycle. The series 10000, which uses the PRISM, supports both the IBM PC AT-compatible bus and VME bus.

### **MIPS Computer**

MIPS Computer designs RISC processor elements and licenses its designs to certain systems manufacturers. MIPS also produces UNIX servers and multiuser systems. MIPS, one of the first major companies to concentrate solely on RISC technology, began developing its RISC technology in 1984. The vendor uses a mature RISC implementation and its investment and dedication have made it one of the leading players in the UNIX/RISC market. MIPS' most significant venture occurred in 1989, when it formed a technical alliance with Digital Equipment Corp. As a result of the agreement, all of Digital Equipment's DECstations and DECsystems are based on MIPS' RISC elements. The company also provides its RISC technology to the following vendors: Pyramid Technology, Tandem Computer, AT&T, Bull, Digital Equipment, Prime Computer, Siemens-Nixdorf, and Silicon Graphics.

MIPS currently produces the following versions of the RISC processor:

- The R2000, which has an associated floating-point unit with write buffer chips.
- The R3000 and its associated co-processors.
- The R5000 and R6000, which provide the best performance and are implemented using ECL component technology.

MIPS uses 32-bit custom VLSI chips with 32 general-purpose registers having a maximum address space of 4G bytes. Internally there are five pipelined levels, with 64 kilobytes of directly mapped instruction and data caches. The pipelining structure uses a 64-entry, fully associative translation buffer and four stage 32-bit write buffers. The architecture is based on demand-paged memory management using a 4-kilobyte page size.

The associated floating-point unit is also a custom VLSI unit, which contains 16 floating-point registers and six levels of internal pipelining. It uses a 64-bit double word internal representation, conforming to the IEEE 754 standard.

### IBM POWER

IBM developed its RISC architecture, called POWER, for the RS/6000 workstations, which were introduced in early 1990. There are three slightly different versions of IBM's POWER RISC chip set, which is termed the SGR 2032. However, the processor basis for the three versions is virtually identical. IBM has not announced any plans to license its technology.

### Intergraph Clipper

Intergraph Corp., a manufacturer of graphical workstations and systems, introduced its first 32-bit RISC Clipper microprocessor in 1985. The current Clipper product line is as follows:

- 25MHz C200 Chipset (6 MIPS).
- 33MHz C200 Module/Chipset (8 MIPS).
- 40MHz C300 Module/Chipset (10 MIPS).
- 50MHz C300 Module/Chipset (14 MIPS).
- 50MHz C311 CPU/FPU Chip (20 MIPS).

In early 1991 Intergraph announced the availability of a new microprocessor, the Clipper C4, which includes the C411 CPU and the C421 IEEE-754 compatible floating-point unit (FPU). The 50MHz C4 is upwardly compatible with existing Clipper microprocessors, allowing applications written for these processors to run without modification on the C4.

Clipper microprocessors are used in Intergraph's interactive computer systems and embedded applications. Interactive systems include engineering and graphics workstations, business systems, and board-level products based on the

PC/AT bus, VMEbus, and NuBus. According to the vendor, more than 50,000 Clipper-based systems were shipped by the end of 1990. Intergraph does not license its Clipper microprocessors to other vendors.

## Vendors

### Motorola Microsystems Group

Motorola Inc.  
29000 South Diablo Way  
Tempe, AZ 85282  
(602) 438-3000

### Intel Corp.

3065 Bowers Avenue  
P.O. Box 58126  
Santa Clara, CA 95052-8126  
(408) 765-8080

### Sun Microsystems

25500 Garcia Avenue  
Mountain View, CA 94043  
(415) 960-1300

### Hewlett-Packard

3000 Hanover Street  
Palo Alto, CA 94304  
(800) 752-0900

### Intergraph Corp.

Advanced Processor Division  
2400 Geng Road  
Palo Alto, CA 94303  
(415) 494-8800

### MIPS Computer Systems

950 Deguine Drive  
Sunnyvale, CA 94086  
(408) 720-1700

### International Business Machines Corp. (IBM)

Old Orchard Road  
Armonk, NY 10504  
Contact your local IBM representative. ■





# An Overview of Superservers

## In this report:

Superserver Product Comparisons .....	2
Vendor/Product Line Profiles .....	6
Vendors .....	10

## Datapro Summary

Although LANs have grown from links between a few PCs to vast webs connecting hundreds of machines, none of the devices employed were originally designed to participate in such an interconnection scheme. PCs were conceived as standalone machines; minicomputers and mainframes as the peaks of a hierarchy without peer. As such, these machines have disadvantages making them less than suitable for use as servers in today's networks. Until recently, these devices were the only tools available to perform the tasks of a server: central storage of data files and programs on disk; connection to shared peripherals; and now, as client/server computing becomes a reality, the more demanding job of performing application requests for client machines. Several vendors have addressed the need for machines designed from the ground up as servers. These "superservers" speed processing of server requests using mainframe-like I/O bus architectures and multiple processors dedicated to I/O control.

## What Are Superservers?

Servers on local area networks (LANs) have four primary functions.

*File Services:* They serve to process application and data files which reside on the server but are used by individuals across the network.

*Database Services:* The database server application distributes a database application across a network so that the client provides a graphical user interface for the user while the compute-intensive tasks are being performed by the database network server. This arrangement is also commonly referred to as client/server computing.

*Print Services:* Printing can occur across networks. Control of the printing can occur at the server level.

*Communications Services:* The task of carrying applications and data over the local area network is also provided by the network server.

While network servers provide these tasks, very often these network servers are no more than personal computers, which have become the workhorses for server functions. As users put more applications and data on network servers, the load on them becomes constraining. These personal computers were not originally designed to process the volume of traffic that they are now being asked to handle. As the market pushes further in the direction of an increased load, the hardware manufacturers have responded with processors that are designed to rapidly and efficiently process file, application, and communications services.

—By Bernard J. David  
General Information Services, Inc.

**Table 1. Superserver Product Comparisons**

Product	Apricot FT 386-25	Apricot FT 486	Apricot FT 486-25	Auspex NS3000
System Processor	Intel 80386DX	Intel 80486DX	Intel 80486DX	Motorola (NS3000M); SPARC (NS3000S)
Clock Speed (MHz)	25	25	25	Unavailable
Cache Size (bytes)	32K	128K	128K	16M
Main Memory (bytes)	4M/8M/12M/16M	4M/8M/12M/16M	4M/8M/12M/16M	8M (Motorola); 20M (SPARC)
System Architecture	Apricot	Apricot	Apricot: symmetrical dual processing	Auspex
Expansion Bus Architecture	MCA	MCA	MCA	Unavailable
Diskette Drives	5.25 in., 1.2MB	5.25 in., 1.2MB	5.25 in., 1.2MB	None
Fixed Disk Space (bytes)	350M/650M/1050M	350M/650M/1050M	350M/650M/1050M	1G; up to 10G
Tape Drives (bytes)	150M/525M	150M/525M	150M/525M	150M; 1.3G/2.3G streaming cartridge
Expansion	Opt. 25MHz Intel 387DC	Opt. 25MHz Intel 387DC	Opt. 25MHz Intel 387DC	Up to 2 Ethernet processors
Standard Interfaces	1 serial port/1 parallel port; mouse/keyboard ports; 6 SCSI devices	1 serial port/1 parallel port; mouse/keyboard ports; 6 SCSI devices	1 serial port/1 parallel port; mouse/keyboard ports; 6 SCSI devices	5 concurrent SCSI channels; up to 4 tape drives
Operating System Support	SCO UNIX; Interactive UNIX V/386 3.2; SCO Xenix 2.3; OS/2; NetWare 2.1; NetWare 386; Pick; DOC; C-DOS	SCO UNIX; Interactive UNIX V/386 3.2; SCO Xenix 2.3; OS/2; NetWare 2.1; NetWare 386; Pick; DOC; C-DOS	SCO UNIX; Interactive UNIX V/386 3.2; SCO Xenix 2.3; OS/2; NetWare 2.1; NetWare 386; Pick; DOC; C-DOS	SunOS, Version 4; Sun ONC services; Network File System (NFS); Sun NIS Services; VME; UNIX (SunOS)
Communications	Ethernet LAN card; token-ring LAN card	Ethernet LAN card; token-ring LAN card	Ethernet LAN card; token-ring LAN card	TCP/IP; Ethernet/IEEE 802.3; SNMP
Data Transfer Rates	12 MIPS	20 MIPS	35+ MIPS	Unavailable
Features	Designed as network server or UNIX host; Apricot HyperCache system; security management software; built-in UPS; block-level encryption; timed logon, audit trail, password history	Designed as network server or UNIX host; Apricot HyperCache system; security management software; built-in UPS; block-level encryption; timed logon, audit trail, password history; advanced system controller	Designed as network server or UNIX host; Apricot HyperCache system; security management software; built-in UPS; block-level encryption; timed logon, audit trail, password history	570 NFS IOPS; SPARC or Motorola processor; 2-4 Ethernet ports; 3,000 packets per second IP routing; 16M-96MB ECC primary I/O cache memory; 1G-10GB disk storage; hot-pluggable disk drives; 8-slot 55M bps enhanced VME backplane

The class of machine that has evolved is being called the *superserver*. Its primary distinction from other PC-type servers is its capability to speedily process large amounts of data to and from its hard disks and through its LAN adapters.

### Superserver Functions

The superserver market has been broken down by three main server functions which categorize its market focus. They are:

- file servers
- communications servers
- applications servers

*File servers* generally offer mass storage and shared access to files and applications; network administration services

(backup and archive); system administration (user authorization and file security); and peripheral services (printers, scanners, and faxes).

*Communications servers* support multiple network media (Ethernet, token-ring, FDDI, X.25, and other wide area networks); multiple protocols and operating systems (DECnet, LAN Manager, TCP/IP, AppleShare, NetWare); and electronic mail gateways.

*Applications servers* support horizontal applications (Database-DBMS/SQL); vertical applications (ECAD, MCAD, CASE, and Scientific); and X Windows applications.

Superservers can be either dedicated file servers, communications servers, or applications servers; or they can perform a mix of functions (i.e., a file server can also be a

**Table 1. Superserver Product Comparisons (Continued)**

Product	Auspex NS5000	Compaq Systempro 386	Compaq Systempro 486	Digital DEC 433MP
System Processor	Motorola (NS5000M); SPARC (NS5000S)	Intel 80386	Intel 80486	Intel 80486
Clock Speed (MHz)	Unavailable	33	33	33
Cache Size (bytes)	16M	64K	512K	256K
Main Memory (bytes)	8M (Motorola); 20M (SPARC)	32M (std.); 256M (max.)	32M (std.); 256M (max.)	8M (std.); 64M (max.)
System Architecture	Auspex	Flexible Advanced Systems Architecture	Flexible Advanced Systems Architecture	DEC
Expansion Bus Architecture	Unavailable	EISA	EISA	ISA/EISA upgradable
Diskette Drives	None	3.5 in., 1.44MB; 5.25 in., 1.2M/360KB	3.5 in., 1.44MB; 5.25 in., 1.2M/360KB	3.5 in., 1.44 MB; 5.25 in., 1.2MB
Fixed Disk Space (bytes)	1G; up to 20G	120M/210M/300M/320M/650M	120M/210M/300M/320M/650M	1.2G (max.); 6M-209M-byte hard disks
Tape Drives (bytes)	150M/1.3G/2.3G streaming cartridge	150M/250M/320M/525M cartridge	150M/250M/320M/525M cartridge	320M/525M QIC
Expansion	Up to 4 Ethernet processors; 20M-68MB CPU memory (SPARC)	386 System Processor; 2M/8M/32MB memory module	486 System Processor; 2M/8M/32MB memory module	1-6 CPUs; opt. CD-ROM
Standard Interfaces	10 concurrent SCSI channels; up to 8 tape drives	11 full-size expansion slots (7 EISA); 4 32-bit processor slots	11 full-size expansion slots (7 EISA); 4 32-bit processor slots	7 ISA/EISA slots; 2 serial ports/1 parallel port
Operating System Support	SunOS, Version 4; Sun ONC servers; Network File System (NFS); VME; UNIX (SunOS)	NetWare 386; NetWare 2.1; SCO UNIX System V; SCO Xenix 386; LAN Manager; VINES 4.0; 3+ Open	NetWare 386; NetWare 2.1; SCO UNIX System V; SCO Xenix 386; LAN Manager; VINES 4.0; 3+ Open	SCO UNIX System V; DECnet; MS-DOS; X-Windows system host; SCO Xenix
Communications	TCP/IP; Ethernet/IEEE 802.3; SNMP	Token-Ring Controller; Arcnet; FDDI; Ethernet	Token-Ring Controller; Arcnet; FDDI; Ethernet	Ethernet; DECnet; TCP/IP; NFS
Data Transfer Rates	Unavailable	Unavailable	Unavailable	Unavailable
Features	1,090 NFS IOPS; SPARC or Motorola processor; 2-8 Ethernet ports; 6,000 packets per second IP routing; 16M-96MB ECC primary I/O cache memory; 1G-20GB disk storage; hot-pluggable disk drives; 14-slot 55M bps enhanced VME backplane	Data striping; simultaneous request servicing; parallel data transfers; optimized request management; 32-bit bus master operation; data guarding; drive mirroring; controller duplexing; drive replacement alert system; auto reliability monitoring; automatic data recovery	Data striping; simultaneous request servicing; parallel data transfers; optimized request management; 32-bit bus master operation; data guarding; drive mirroring; controller duplexing; drive replacement alert system; auto reliability monitoring; automatic data recovery	Flexible system architecture; 1-6 CPU expansion; memory expandable to 64MB; disk expansion to 9.2GB using 18 devices; augments VAX and RISC products

communications server). When some speak of superservers, they refer specifically to those servers which have been designed to facilitate network traffic—the communications server. Yet with the integration of tasks in some servers, it makes more sense to create this categorization in the manner discussed previously.

Superservers represent a convergence of the mainframe and minicomputer designs of yesterday and the desktop-based processing seen in microcomputers. Superservers look a lot like minicomputers or mainframes—they serve to process the transactions of a host of users, many of whom are now using diskless workstations (the terminal equivalent of yesterday). Processing occurs at the server level (the minicomputer or mainframe computer), and the results are viewed on the user's terminal. In fact, the only

real differences between the mainframes and minicomputers of yesterday and today's superservers are size and processing capability. Superservers are smaller and can process more information more rapidly than many minicomputers and mainframes. They also can split the way in which processing is done (some on the client and some on the server).

The companies that have created superservers are the mainframe and minicomputer companies that recognize the trend that is occurring; the microcomputer companies that realize this trend is a natural extension of their product lines; and start-up companies that see the market as an opportunity to capitalize on what Forrester Research predicts will be a \$12 billion market by 1994.

**Table 1. Superserver Product Comparisons (Continued)**

Product	IBM PS/2 Model 95 XP 486	NCR S486/MC33	NetFRAME NF100	NetFRAME NF200
System Processor	Intel 80486	Intel 80486	Intel 80386	Intel 80486
Clock Speed (MHz)	33	33	25	25
Cache Size (bytes)	256K	Unavailable	32K	32K
Main Memory (bytes)	8M (std.); 32M (max.)	4M/16M	8M (std.); 32M (max.)	8M (std.); 32M (max.)
System Architecture	IBM	Dual Bus/Interleaved Memory Architecture	NetFRAME	NetFRAME
Expansion Bus Architecture	MCA	MCA	NetFRAME	NetFRAME
Diskette Drives	3.5 in., 1.44MB; 5.25 in., 1.2MB	3.5 in., 1.44MB; 5.25 in., 1.2MB	None	None
Fixed Disk Space (bytes)	400M; up to 2G internal storage	200M/320M/640M	380M (std.); 3G (max.)	380M (std.); 16G (max.)
Tape Drives (bytes)	80M-2.3G tape backup	200M/320M/525M tape backup	1.3G DAT or 2.2G cartridge	1.3G DAT or 2.2G cartridge
Expansion	33MHz system; 2 32-bit MCA expansion slots	Up to 64MB of 80-ns 32-bit memory (EDAC); up to 4GB internal	Intel 376 (386SX) processor; 8M/16MB main memory	Intel 376 (386SX) processor; 8M/16MB main memory
Standard Interfaces	6 32-bit MCA expansion slots; 1 DMA serial port/1 DMA parallel port; 7 internal storage bays	7 32-bit MCA expansion slots; 2 serial ports/1 parallel port	1 SCSI-II (5M bps); Intel 80376 processor; 1 LocalTalk port/1 serial port	1 SCSI-II (5M bps); Intel 80376 processor; 1 LocalTalk port/1 serial port
Operating System Support	OS/2 Extended Edition (OS/2 EE) 1.2/1.3; OS/2 Standard Edition (OS/2 SE) 1.2/1.3; DOS Version 3.3 and 4.0; AIX PS/2 1.2.1; IBM 4680 Operating System Version 2/3; 3+Open; VINES/486; Advanced NetWare 286/386; SCO Xenix/UNIX System V/386	MS-DOS; OS/2; UNIX; NetWare	NetWare 386; LAN Manager; UNIX	NetWare 386; LAN Manager; UNIX
Communications	Ethernet; Asynchronous; X.25; token-ring	Unavailable	RS-232; OSI; TCP/IP; Ethernet; SNA/SDLC; XNS; token-ring; LocalTalk (RS-422)	RS-232; OSI; TCP/IP; Ethernet; SNA/SDLC; XNS; token-ring; LocalTalk (RS-422)
Data Transfer Rates	Unavailable	7 MIPS	3M bps	3M bps
Features	Enhanced XGA display providing 1024 x 768 resolution; PS/2 SCSI 32-bit bus master; 400MB SCSI fixed disk with 11.5-ms seek time; 8 32-bit MCA expansion slots (1 for SCSI and 1 for XGA display adapter); security, audit, and control features	Super VGA video interface (800 x 600); high-performance SCSI II, implemented closely coupled I/O processor; configurable keylock and software password; auto switching international 385-watt power supply; Weitek 4167 Arithmetic Coprocessor	Parity checking on all data paths; error correcting memory; power module redundancy; automatic restart/retry; supports NetFRAME network management and administrative tools (SAM and RCON); remote administration from any 286/386 on network	Parity checking on all data paths; error correcting memory; power module redundancy; automatic restart/retry; supports NetFRAME network management and administrative tools (SAM and RCON); remote administration from any 286/386 on network

The key elements of distinction between companies in the superserver market are the following.

**Processors:** While most superservers use Intel's processors (80386 or 80486, either 25MHz or 33MHz), there are some machines that use Motorola processors. What has been shown repeatedly is that, while on the surface, it looks

as if the processor type has a tremendous impact on processing speed, I/O bus structure and adapters have an even greater impact on speed. While RAM and cache size vary on these processors, RAM can inhibit network and application throughput if there is not enough available.

**Table 1. Superserver Product Comparisons (Continued)**

Product	NetFRAME NF300	NetFRAME NF400	Parallan Server 290 Model 10	Parallan Server 290 Model 20
System Processor	Intel 80386	Intel 80486	Intel 80486	Dual Intel 80486
Clock Speed (MHz)	25	25	33	33
Cache Size (bytes)	32K	32K	128K	128K x 2
Main Memory (bytes)	8M (std.); 64M (max.)	16M (std.); 64M (max.)	8M (std.); 128M (max.)	16M (std.); 128M (max.)
System Architecture	NetFRAME	NetFRAME	Parallan	Parallan
Expansion Bus Architecture	NetFRAME	NetFRAME	MCA	MCA
Diskette Drives	None	None	3.5 in., 1.44MB	3.5 in., 1.44MB
Fixed Disk Space (bytes)	380M (std.); 6G (max.)	380M (std.); 6G (max.)	676M x 2 (std.); 18.9G (max.)	676M x 2 (std.); 18.9G (max.)
Tape Drives (bytes)	1.3G DAT or 2.2G cartridge	1.3G DAT or 2.2G cartridge	1.2G DAT	1.2G DAT
Expansion	Intel 376 (386SX) processor; 8M/16MB main memory	Intel 376 (386SX) processor; 8M/16MB main memory	12 MCA slots (1 EVGA); 8M/32MB incremental memory	12 MCA slots (1 EVGA); 8M/32MB incremental memory
Standard Interfaces	1 SCSI-II (5M bps); Intel 80376 processor; 1 LocalTalk port/1 serial port	1 SCSI-II (5M bps); Intel 80376 processor; 1 LocalTalk port/1 serial port	8 MCA slots (1 EVGA); up to 4 SCSI buses and controllers	8 MCA slots (1 EVGA); up to 4 SCSI buses and controllers
Operating System Support	NetWare 386; LAN Manager; UNIX	NetWare 386; LAN Manager; UNIX	Parallan-enhanced OS/2 1.21; client workstation support for DOS, Windows, OS/2	Parallan-enhanced OS/2 1.21; client workstation support for DOS, Windows, OS/2
Communications	RS-232; OSI; TCP/IP; Ethernet; SNA/SDLC; XNS; token-ring; LocalTalk (RS-422)	RS-232; OSI; TCP/IP; Ethernet; SNA/SDLC; XNS; token-ring; LocalTalk (RS-422)	32-bit bus master; 16-bit Ethernet; token-ring; SNA/SAA	32-bit bus master; 16-bit Ethernet; token-ring; SNA/SAA
Data Transfer Rates	7M bps	8M bps	Up to 70 MIPS	Up to 70 MIPS
Features	Parity checking on all data paths; error correcting memory; power module redundancy; automatic restart/retry; supports NetFRAME network management and administrative tools (SAM and RCON)	Parity checking on all data paths; error correcting memory; power module redundancy; automatic restart/retry; supports NetFRAME network management and administrative tools (SAM and RCON)	Maximum availability and support subsystem; LAN Manager multiprocessing extensions; opt. LAN Manager 2.0	Maximum availability and support subsystem; LAN Manager multiprocessing extensions; opt. LAN Manager 2.0

**I/O Bus Structure/Architecture:** Vendors in this market use the ISA (Industry Standard Architecture), EISA (Extended Industry Standard Architecture), or Micro Channel Architecture or a proprietary bus structure. The bus structure and the types of I/O adapters used have a great impact on network performance.

**I/O Adapters:** While most of the I/O that occurs is through some type of SCSI device, both the type and number of I/O adapters can have a dramatic effect on the performance of the network. Those companies that have developed either a proprietary bus or I/O adapter can often achieve much greater performance than those using industry-standard buses or adapters.

**Operating System:** Commonly, superservers support multiple operating system environments such as MS-DOS, OS/2, UNIX, SCO UNIX, SCO Xenix, or SunOS (UNIX).

Different operating systems can have an effect on the network performance. The LAN running on top of the operating system can also impact on the network. Differences in 3Com, Banyan, IBM, and Novell LANs can impede or aid the network.

**Network Communications:** Direct hooks into and out of the LAN and the I/O processor are achieved through network communications. The most popular vehicles to use are Ethernet and token-ring, which offer network performance that is either 1M or 10M bps (Ethernet) or 4M or 16M bps (token-ring).

**Network Administration:** To facilitate the use of the network, many superserver companies have designed special network management tools to audit, control, and monitor activities either locally or remotely.

**Table 1. Superserver Product Comparisons (Continued)**

Product	Parallan Server 290 Model 50	Parallan Server 290 Model 60	Tricord PowerFrame 30	Tricord PowerFrame 40
System Processor	Dual Intel 80486	Dual Intel 80486	1-2 Intel 80386	1-2 Intel 80386
Clock Speed (MHz)	33	33	25/33	25/33
Cache Size (bytes)	128K x 2	128K x 2	256K	256K
Main Memory (bytes)	24M (std.); 128M (max.)	32M (std.); 128M (max.)	8M (std.); 128M (max.)	8M (std.); 128M (max.)
System Architecture	Parallan	Parallan	Tricord Multiprocessor	Tricord Multiprocessor
Expansion Bus Architecture	MCA	MCA	EISA	EISA
Diskette Drives	3.5 in., 1.44MB	3.5 in., 1.44MB	3.5 in., 1.44MB	3.5 in., 1.44MB
Fixed Disk Space (bytes)	676M x 8 (std.); over 30G (max.)	676M x 16 (std.); over 30G (max.)	385M	385M
Tape Drives (bytes)	1.2G DAT	1.2G DAT	2.3G helical scan SCSI	2.3G helical scan SCSI
Expansion	12 MCA slots (1 EVGA); 8M/32MB incremental memory	12 MCA slots (1 EVGA); 8M/32MB incremental memory	EISA bus upgradable to IIOIP	8MB memory module
Standard Interfaces	8 MCA slots (1 EVGA); up to 4 SCSI buses and controllers	8 MCA slots (1 EVGA); up to 4 SCSI buses and controllers	EISA SCSI controller or IIOIP; 5-slot PowerBus; 7 EISA slots; 1 parallel port/2 serial ports	EISA bus with IIOIP; 5-slot PowerBus; 7 EISA slots; 1 parallel port/2 serial ports
Operating System Support	Parallan-enhanced OS/2 1.21; client workstation support for DOS, Windows, OS/2	Parallan-enhanced OS/2 1.21; client workstation support for DOS, Windows, OS/2	NetWare 386; OS/2; SCO UNIX; VINES; Interactive UNIX	NetWare 386; OS/2; SCO UNIX; VINES; Interactive UNIX
Communications	32-bit bus master; 16-bit Ethernet; token-ring; SNA/SAA	32-bit bus master; 16-bit Ethernet; token-ring; SNA/SAA	Ethernet; token-ring; TCP/IP	Ethernet; token-ring; TCP/IP
Data Transfer Rates	Up to 70 MIPS	Up to 70 MIPS	33M bps	33M bps
Features	Maximum availability and support subsystem; LAN Manager multiprocessing extensions; opt. LAN Manager 2.0	Maximum availability and support subsystem; LAN Manager multiprocessing extensions; opt. LAN Manager 2.0	Virtual memory; coherent cache with zero wait state design; Fast Page Mode main memory for burst-mode transfers; scalable architecture allowing for incremental increases in power, performance, and capacity; data integrity provided by PowerFrame IIOIP; open architecture	Virtual memory; coherent cache with zero wait state design; Fast Page Mode main memory for burst-mode transfers; scalable architecture allowing for incremental increases in power, performance, and capacity; data integrity provided by PowerFrame IIOIP; open architecture

## Vendor/Product Line Profiles

Companies that are active in the market are shown here by their origins.

### Mainframe/Minicomputer Vendors

- AT&T
- Digital Equipment Corp.
- Hewlett-Packard Co.
- IBM
- NCR Corp.

### Microcomputer Vendors

- Compaq Computer Corp.
- Dell Computer Corp.

### Start-Up Companies

- Apricot Computers plc
- Auspex
- NetFRAME Systems Inc.
- Parallan Computer Inc.
- Tricord Systems Inc.

In the following paragraphs, we present profiles of the preceding 12 superserver vendors and their superserver product lines.

### Apricot Computers plc

Apricot has three superservers on the market: the Apricot FT 386-25, FT 486, and FT 486-25. The Apricot FT 386-25 uses an Intel 80386DX processor with a clock speed of 25MHz and cache of 32KB. It performs at 12 MIPS. The system has an optional 25MHz Intel 80387 DX processor.

The Apricot FT 486 employs an Intel 80486 processor. The processor has a clock speed of 25MHz and 128 Hyper-cache which is 128-bit-wide cache. It performs at 20 MIPS. The Apricot FT 486-25 employs symmetrical dual processing architecture and uses an Apricot DSPA 486 with an optional second 25MHz Intel 486 processor. Its performance is 35+ MIPS total system rating.

The Apricot superservers have a standard VGA display with ports for a mouse and keyboard. The Advanced System Controller (ASC) monitors the status of main processors, the standard built-in UPS, mains and battery supplies, and system thermal sensors and controls access to internals.

The SCSI drive system allows for up to six SCSI devices—five full-height bays and one half-height bay. A variety of hard disks and SCSI backup devices can be placed on the system. The serial port controller has up to 32 channels per card with a maximum of 64 channels.

The security system is controlled by a security processor and an access control device (which is accessed with an Apricot infrared security card). The security management software allows for a maximum of 25 user IDs and has such security features as timed logon, audit trail, password history, screen blanking, and block-level encryption (with a proprietary algorithm).

Apricot supports a host of operating systems such as MS-DOS, SCO and Internative UNIX V/386 3.2, SCO Xenix 2.3, OS/2, NetWare 2.1, NetWare 386, Pick, DOS, and C-DOS.

### AT&T

AT&T offers two systems that can be classified as superservers—the StarServer S System and the StarServer E Symmetric Multiprocessor System. StarServer S is based on the Intel i486 processor and the EISA bus architecture; it is backward compatible with the ISA bus structure. It provides 26.5 MIPS performance and supports three operating systems: AT&T UNIX System V (Releases 3.2.3 and 4.0), DOS 4.01, and OS/2 1.2 Standard Edition. It is configured with 8MB of memory as a server and is expandable to 64MB of memory. There are 10 EISA I/O expansion slots.

The StarServer E is expandable from a single CPU (26.5 MIPS) to a four-CPU Symmetric Multiprocessor System (106 MIPS). It has a fully symmetric shared-memory version of the AT&T UNIX System V operating system. StarServer E includes 12 EISA I/O expansion slots and is configured with 8MB of memory, expandable to 512MB.

### Auspex

Auspex, founded in 1987, has been financed mainly through its \$20.6 million infusion of venture capital. The management team includes the founders of Quantum, Adaptec, and Bridge Communications. It views its target market as the network and file access segment of the super-server market.

The company has two different types of UNIX servers: the NS3000 and NS5000 series. Both servers are available with either Motorola or SPARC-based processors. The

NS3000 has a network performance of 570 NFS I/O operations per second (IOPS), while the NS5000 performs at 1,090 NFS IOPS. The user can get two to four Ethernet ports on the NS3000 server or two to eight ports on the NS5000. The IP routing ranges from 3,000 packets per second (NS3000) to 6,000 packets per second on the NS5000.

Cache memory on both systems ranges from 16M to 96MB of primary I/O cache memory. The disk drives are “hot pluggable,” supporting on-line installation and removal of disks. The VME backplane supports either an 8-slot (NS3000) or 14-slot (NS5000) VME backplane. Both systems are completely compatible with UNIX (SunOS), ONC/NFS, SNMP, TCP/IP, Ethernet, and VME.

Some of the features specific to both of these systems include REX, a Remote EXecutive service, a remote copy program (RCP), a remote login program (RLOGIN), electronic mail (SMTP and UUCP), and a TCP/IP terminal emulation (Telnet).

### Compaq Computer Corp.

Compaq is an example of a microcomputer company that entered the superserver market as a natural outgrowth of its own microcomputer business. Compaq's superserver product line is called the Systempro. There are fundamentally two different lines of the Systempro: the 33MHz Intel 80486-based models and the Intel 80386-based models. Compaq offers a total of six models (three in each processor type) with the primary difference between these models being the size of fixed disk (240M/420M and 840MB).

Compaq calls its architecture for the superserver the Flexible Advanced Systems Architecture with Multiprocessing Support (Flex/MP). Flex/MP allows concurrent processing and I/O activity, delivering a high level of 32-bit system performance while maintaining compatibility with industry-standard hardware and software.

The bus structure is based on EISA, a standard Compaq helped establish. In concert with the Token-Ring Controller, it provides support for Ethernet, Arcnet, and FDDI networking through third party vendors. The Systempro uses an innovative fixed disk drive array technology which creates a faster response time to requests from multiple users. It can manage requests up to four times faster than nonarrayed drive systems.

The tower chassis of the unit provides 11 expansion slots and can accommodate 11 mass storage devices. Features such as drive mirroring, controller duplexing, and data guarding provide enhanced data protection. The Systempro models are positioned as client/server servers that perform a combination of file, applications, and communications services.

### Dell Computer Corp.

Dell has combined quality products with mail order convenience to claim a top spot in the IBM PC clone market. Like Compaq, Dell sees superservers as a logical extension of its existing product line. The company offers two superserver models—the 25MHz 425TE and the 33MHz 433TE. Both machines are based on the Intel i486 processor.

Both models include 4MB of memory, expandable to 64MB on the memory board; eight expansion slots accommodating either ISA or EISA adapters; and 11 half-height storage bays for UNIX multiuser systems, workgroup servers, or power workstations.

**Digital Equipment Corp.**

The DEC 433MP is Digital's answer to superservers. While it can function as a network server, Digital also touts it as a multiuser timesharing system, applications/file server, or multiuser workstation for software development.

The system runs SCO UNIX System V/386. It can support more than 100 users. The system supports from one to six CPUs. It supports various networking capabilities including TCP/IP, NFS, DECnet, and PC LAN networking. It can simultaneously run SCO UNIX System V, Xenix, and MS-DOS applications.

The DEC 433MP supports both the ISA and EISA bus structures. It supports from one to six Intel 80486 processor boards which allow for network expansion. The system has four embedded, high-speed serial I/O ports and a high-performance SCSI port.

Memory may be expanded from 8M to 64MB in 4MB increments. Its internal storage can expand to 1.2GB by using six 209MB disks. By using external storage devices, storage can be expanded to 8GB. Storage includes a 320M/525MB QIC tape drive, a 3.5-in. diskette drive, a 5.25-in. diskette drive, and a CD-ROM.

**Hewlett-Packard Co.**

Hewlett-Packard's entry in the superserver market is the HP 9000 Series 800 family of business servers. The machines are based on HP's Precision Architecture RISC (PA-RISC), VLSI technology, and the UNIX operating system (HP-UX).

The HP 9000 Series 800 consists of the following models: 822S, 832S, 842S, and 852S, which are the entry-level and midrange members of the family; and the 850S, 855S, 860S, 865S, 870S/100, and 870S/200, which are the high-end members of the family.

The entry-level and midrange models feature performance ranging from 11 MIPS to 52 MIPS, memory ranging from 8MB to 256MB, and internal disk storage ranging from 335MB to 2.68GB. The processors reside on a single board; the processor module contains VLSI chips including the CPU, control units for the cache, System Interface Unit (SIU), and Floating-Point Co-Processor (FPC).

The high-end models feature performance ranging from 14 MIPS to 100 MIPS, memory ranging from 48MB to 768MB, and disk storage ranging from 42.88GB to 85.76GB. They contain the same single-chip CPU design as the low-end and midrange models.

**IBM**

IBM sees the convergence of both its microcomputer and minicomputer lines in its line of superservers—the IBM Personal System/2 Model 95 XP 486. While it carries a name similar to many of its personal computers, IBM has clearly distinguished this system as a server. It comes standard with an Intel 80486 processor running at 25MHz. The processor chip is upgradable to a 33MHz chip. The standard memory configuration is 8MB of main memory, expandable to 32MB. The standard 400MB SCSI fixed disk has a seek time of 11.5 milliseconds. The PS/2 Model 95 XP 486's data storage is expandable to 2GB. The system is equipped with eight 32-bit Micro Channel expansion slots (one is used for the SCSI adapter, and one is used for its XGA Display Adapter). There are seven internal storage device bays which support either a 3.5-in. half-height drive or a 5.25-in. full-height drive.

IBM has placed an enhanced performance XGA display adapter on the PS/2 Model 95 XP 486 that provides a

1,024 by 768 video display resolution. There is one DMA serial port and one DMA parallel port for use with external communications. The system supports a host of other communications through adapter cards which can be added to the system. It supports communications with Ethernet and token-ring networks as well as terminal emulation on an IBM 3270-type device or an IBM midrange 36/38 processor.

In addition to this network support, the PS/2 Model 95 XP 486 can communicate with many of the Rolm data communications modules to interface with this telephone equipment. Storage on the system is augmented by tape backup (ranging from 80MB to 2.3GB) and a CD-ROM device. A host of printers are supported by the system, including the IBM PagePrinter, LaserPrinter, Quietwriter, and IBM color plotters.

Unlike many IBM systems, the PS/2 Model 95 XP 486 supports a host of operating systems including standard IBM DOS, Versions 3.3 and 4.0; OS/2 Standard and Extended Editions, Versions 1.2 and 1.3; the IBM 4680 Operating System; 3Com 3+Open LAN Manager Advanced System; Banyan VINES/486; Novell's Advanced NetWare 286, SFT NetWare 286, and NetWare 386; and SCO Xenix System V and SCO UNIX System V/386. IBM has also confirmed that this server supports most of the popular third-party products which run under MS-DOS.

**NCR Corp.**

The NCR S486/MC is this minicomputer vendor's superserver entry. It is built on NCR's own understanding of the client/server model and incorporates a 32-bit, 33MHz Intel 80486 processor which is fully expandable to either a Micro Channel-based server or workstation. The system supports a range of operating systems: MS-DOS, OS/2, UNIX, and NetWare.

Six half-height and three full-height drive bays come with the system. All of the drives can be plugged into a "cableless" SCSI bus board. A single SCSI Host Adapter supports up to seven SCSI devices, which include a CD-ROM, optical disk drive, laser printer, and scanner.

When used as a workstation, the system can run at 27 MIPS and facilitate storage up to 4GB internally. The video interface is a super-VGA with 1MB of RAM and seven MCA expansion slots. All of the S486/MC33's peripherals are based on an Intelligent SCSI RISC processor. The system can be expanded to 64MB of 80-nanosecond, 32-bit error detecting and correcting dual-ported memory (EDAC) through the parallel bus with four 16MB memory boards.

The S486/MC33 uses interleaved memory, which enhances memory as additional memory is added to the system. Its dual-ported memory improves the access time when multiprocessing by freeing the processor and enabling memory from one side to be used for the peripherals while the CPU uses memory from the other side.

**NetFRAME Systems Inc.**

NetFRAME Systems is an example of a company that was established solely to serve the superserver marketplace. Its product line includes four superservers—the NF100, NF200, NF300, and NF400. Many of the design elements permit interchangeability between systems, as well as upgradability.

The NetFRAME family has its own, unique, multiple independent bus structure that is shared throughout the product line. It uses the same memory architecture as well. The NF100 is referred to as NetFRAME's "entry-level



with room to grow" model. It uses a 25MHz Intel 80386 processor. The standard configuration has 8MB of error correcting memory, 380MB of disk storage, support for remote console software, and NetFRAME's Server Activated Maintenance (SAM) program, which performs the basic server maintenance functions.

One I/O expansion board provides links to SCSI-II, RS-232, RS-422 (SDLC), and Ethernet or token-ring LANs. Maximum main memory is 32MB, and maximum internal disk storage is 3GB. The NF100 can attach to NetFRAME's external storage systems which enable the system to expand to an additional 16GB of storage. The system itself can accommodate three expansion processor boards.

The NF200, called "high performance in a compact unit," uses a 25MHz Intel 80486 processor. This server is viewed more as a data processing- or application-based server. The NF200 is equipped with 8MB of error-correcting memory, 380MB of disk storage, support for remote console software, and NetFRAME's SAM program. It also has one I/O expansion board. An NF100 can be upgraded to an NF200.

The NF300 is capable of supporting up to five fully loaded Ethernet or token-ring LANs, unlike the NF100 or NF200, which can support only three. It uses a 25MHz Intel 80386 processor and comes with 8MB of error-correcting memory, a 380MB hard disk, and all other system and administration features found on the other NetFRAME systems. The system can hold eight expansion processor boards. Disk storage is expandable to 6GB internally and 42.6GB externally; its maximum memory is 64MB. Redundant DC supply modules are an option that ensures further data integrity and system uptime.

The NF400 uses an Intel 80486 processor; its physical characteristics match the NF300, with 16MB of error-correcting memory. It can support up to eight fully loaded Ethernet or token-ring LANs.

### **Parallan Computer Inc.**

Parallan manufactures a family of superservers intended to serve as platforms for transaction processing, SQL-based database management, and decision support. The servers use a hierarchical bus structure whose central feature is a 64-bit, 200M bps, parity-protected InterProcessor bus that supports the dual Intel 80486 processors. It also supports four banks of shared main memory, a Remote Maintenance processor, and dual-channel SCSI controllers. The server operates such that applications reside on one of the processors while the LAN Manager High Performance File System (32-bit HPFS) and network protocols reside on the other.

The Parallan servers are upwardly compatible and include the Parallan Server 290, Models 10, 20, 50, and 60. The Model 10 has five processors, including one 33MHz 80486 system processor, two RISC SCSI processors, the bit-sliced Intelligent Memory Mover, and the 80C186-based Remote Maintenance Processor (RMP). A 64-bit InterProcessor Bus, 8MB of Error Checking and Correcting (ECC) main memory, 676MB of hard disk storage, one dual-channel SCSI intelligent disk controller, a Micro Channel bus with eight slots, and Parallan's Maximum Availability and Support Subsystem (MASS) are all a part of the Model 10.

The Model 20 has one additional processor beyond the Model 10, a 33MHz Intel 80486 system processor. It differs from the Model 10 in that it contains 16MB of ECC main memory, 1.3GB of hard disk storage, a dual-channel

SCSI controller, and a dual Micro Channel bus. The Model 50 has eight processors—two 33MHz Intel 80486 system processors, four SCSI processors, the Intelligent Memory Mover, and the 80C186-based RMP. It has the 64-bit IP-bus, 24MB of ECC main memory, 5.4GB of hard disk storage, two dual-channel SCSI controllers, dual Micro Channel buses with 12 slots, MASS, and one expansion enclosure. The Model 60 is identical to the Model 50 except that it has 32MB of main memory, 10.8GB of hard disk storage, and three expansion enclosures.

The Maximum Availability and Support Subsystem (MASS) is Parallan's network administration tool. It also enables customer personnel or Parallan support to monitor, control, and tune a Parallan Server 290 remotely. Finally, it gives the system availability for fault-resilient operation.

### **Tricord Systems Inc.**

Tricord Systems makes a family of superservers which has been specifically designed to give the client/server application processing power, flexibility, and implementation of popular PC industry operating systems. The PowerFrame Models 30 and 40 use Intel-based 80486 processors. The PowerFrames include Tricord's scalable Common Multiprocessor Architecture (CMA), which allows users to incrementally increase the power, performance, and capacity of their systems.

The PowerFrame was introduced in mid-1990 and has been designed specifically for networks. The core technology consists of a high-performance memory design, one or two Intel 80486 processors, an EISA bus, a 132M bps Tricord PowerBus, and the Intel 800386-based Intelligent Input/Output Processor (IIOP). Tricord considers the PowerBus and IIOP, both developed by the company, to be the keys to network performance.

Operating systems supported by the PowerFrame are Novell's NetWare, Microsoft's OS/2 and LAN Manager, SCO UNIX, Interactive UNIX, and Banyan VINES. The PowerBus' bandwidth of 132M bps enables both rapid network performance as well as the use of multiple Intel 80486-based processors. It interconnects the 80486 processors, memory, and I/O devices, minimizing contention on the network.

The Intelligent I/O Processor is dedicated to the disk I/O task and enables multiprocessing to occur. It transfers data between memory and disk, which maximizes disk throughput. By performing in this manner, the main CPU can then concentrate on operating system and application tasks. A PowerFrame can support either one or two IIOPs. When it has two, it can support a total storage capacity of 42GB on 28 drives. PowerFrame uses an EISA I/O bus for LAN connections. It can transfer data at a maximum rate of 33M bps. This architecture also allows for use with ISA devices.

The Model 30 is a tower configuration optimized for use as a file, communications, or print server. The Model 40 provides greater expandability.



## Vendors

Listed here, for your convenience, are the addresses and telephone numbers of the vendors whose superserver product lines are profiled in this report.

**Apricot Computers plc**

111 Granton Drive, #401  
Richmond Hill, ON, L4B 1L5 Canada (416) 492-2777

**AT&T**

295 N. Maple Avenue  
Basking Ridge, NJ 07920 (908) 221-8694

**Auspex**

2952 Bunker Hill Lane  
Santa Clara, CA 95054 (800) 735-3177

---

This report was developed exclusively for Datapro by Bernard J. David, president of General Information Services, Inc., Wilmington, DE, a firm specializing in computer connectivity, consulting, evaluation, and product provision. Mr. David is also a lecturer in entrepreneurship at the Wharton School, University of Pennsylvania.

**Compaq Computer Corp.**

P.O. Box 692000  
Houston, TX 77269 (713) 370-0670

**Dell Computer Corp.**

9505 Arboretum Boulevard  
Austin, TX 78759 (512) 338-4400

**Digital Equipment Corp.**

146 Main Street  
Maynard, MA 01754-2571 (508) 493-5111

**Hewlett-Packard Co.**

19091 Pruneridge Avenue  
Cupertino, CA 95014 (800) 752-0900

**IBM**

Old Orchard Road  
Armonk, NY 10504 (914) 764-1900  
Contact your local IBM representative.

**NCR Corp.**

1700 S. Patterson Boulevard  
Dayton, OH 45479 (513) 445-5000

**NetFRAME Systems Inc.**

1545 Barber Lane  
Milpitas, CA 95035 (408) 944-0600

**Parallan Computer Inc.**

201 Ravendale Drive  
Mountain View, CA 94043 (415) 960-0288

**Tricord Systems Inc.**

3750 Annapolis Lane  
Plymouth, MN 55447 (612) 557-9005 ■

# An Overview of Database Servers

## In this report:

Client/Server Architecture .....	2
The Benefits .....	2
The Drawbacks .....	2
Vendors .....	2

## Datapro Summary

The promise of distributed computing on microcomputers depends on the development of the tools. While LAN technology has matured in recent years, distributed applications have still not reached their potential. The heart of distributed database technology is the database server. This report presents a general overview of database servers and client/server computing and presents the current PC-based offerings.

SQL (pronounced "sequel") database servers provide a means to improve communications and decrease network traffic over a distributed database by providing a central holding point (or central database). A database server is similar to a LAN server, offering many of the same features and services.

While the press has heralded each successive "year of the LAN," most potential users have waited to see what all the noise is about. LANs have become a symbol of unused potential rather than a means of optimizing computer systems performance. One of the primary reasons for LAN underuse has been the immaturity of the enabling technologies. The introduction of SQL servers is a key factor in realizing the potential of distributed computing and client/server architecture. In fact, the introduction of SQL servers and the promise of client/server computing may represent the turning point for widespread implementation of LAN technology in the corporate environment.

## Database Engines

A SQL server is a database engine used in a LAN environment to provide both centralized database control and generalized distributed database access. The application's

user interface is stored at the individual workstation, while the database engine is installed on a network server. The engine (or back end) provides services such as security and concurrency control. Separating an application into two functions (front end and back end) allows the workstation to be optimized for ease of use and the server to be optimized for maximum throughput.

When a workstation application requires data for its task, it requests the data from the database server. Rather than give the workstation a copy of the entire database, the server sends only the subset required. Sending this subset reduces both the message traffic across the network and the amount of database processing performed at the workstation level.

Structured Query Language (SQL) handles communications between workstation and server. SQL is a set-based language designed to perform relational queries quickly and easily with as few verbs and objects as possible. It was invented by IBM for use with its DB2 mainframe relational database, but it has been adopted as a de facto standard by vendors of mainframe and minicomputer-based database systems, and it is quickly becoming a standard for microcomputer database management software as well.

The introduction of the OS/2-based Microsoft/Sybase SQL Server signaled a

—By *Karen J. Offermann*  
Associate Editor

new era of LAN-based microcomputer database management software, yet the momentum for this technology is building very slowly. PC-based SQL servers are available only from Microsoft/Sybase, Gupta, Oracle, and IBM. Borland's Paradox Engine is not a ready-to-go server, but it can be used to develop one. We expect to see a dramatic increase in database server support among programmable, multiuser database packages over the next year. While some SQL servers run under MS-DOS, virtual mode multitasking operating systems such as OS/2 and UNIX offer a number of advantages over the DOS environment.

### Client/Server Architecture

Understanding the implications of SQL servers requires a knowledge of the basic principles of client/server architecture. Simply stated, in client-server computing, the server and the workstations can work independently. The server is referred to as the "back end" (or "engine") while the workstation is referred to as the "front end." The SQL query language allows the workstation to request only the data required for its operations. In the past, a data request usually resulted in huge blocks of data moving across the network to the workstation. The workstation would then sort out usable data and discard the rest. This method resulted in slow access time, poor security, and lowered data integrity.

Client/server architecture divides an application into separate processes that operate on separate CPUs connected over a network. Client/server computing is inherently one-way, unlike peer-to-peer communications in which both processors can exchange instructions. The client CPU, generally an intelligent workstation, manages or controls the user interface and communicates commands to drive the activity of a server across a network via remote procedure calls (RPCs). RPCs, the heart of client/server computing, are software capabilities used in distributed applications. Programs operating on a local system can "call" procedures or processes operating on remote systems by ordering messages, translating different data formats, and maintaining the integrity of transferred data.

In the client/server environment, the portion of the application shared among all users resides on the server. The user interface resides on the client machine. Some users define client/server architecture as one where the client depends entirely on the server and cannot accomplish anything without it. Others hold that the server's purpose is to enhance the client's capabilities without any implied dependency. Attitudes about departmental autonomy seem to correlate with attitudes about distributed computing.

Client/server systems are *loosely coupled*—although processes and systems are integrated, an individual CPU runs under its own operating systems. In a tightly coupled system, several processors are integrated under a single operating system. In fact, most distributed systems today are considered loosely coupled. A loosely coupled network containing a server and workstations shares data and other resources. Most integrators aim to provide capabilities in a loosely coupled network that exceed those of the individual workstation, allowing the client to access more memory or to off-load processing.

### The Benefits

SQL servers offer several benefits over single-user systems—most notably speed gains and additional data

storage capacity. In addition, issues of security and data integrity are more easily controlled in a client/server environment.

- Users can access data on other systems without having to leave a currently running application.
- One database engine can provide data to a number of application types, such as database management software, spreadsheets, and word processors.
- When a database server is implemented with a graphical user interface (GUI), the various applications can appear on the desktop as one.
- Multiple data access methods are available in this version of the client/server model including application program interfaces (APIs), which are essentially low-level function calls.
- Languages such as C and Pascal can be used to create custom programs.
- SQL permits optimization of database queries and supports complex manipulation of data, including nested queries, joining a table to itself, and sophisticated sort and calculation capabilities.

### The Drawbacks

As with any new technology, SQL servers have drawbacks. The immaturity of the market should inspire caution among users who are interested in long-term solutions. Not all vendors who venture into the server marketplace are going to remain there. Users should investigate the vendor's financial status and expertise with the technology before committing to an expensive, long-term investment.

While increased speed and productivity are obvious benefits, they come at a price. Database servers are an evolutionary technology and require a significant amount of custom programming and optimization. Database administration is a full-time job that requires greater skills than a single-user system.

- Front-end applications and development tools are only now emerging.
- The implementation of a SQL database server and network is a pioneering effort, requiring custom programming.
- Communications between the client and server puts a demanding load on a LAN; careful capacity planning is essential.

## Vendors

### The Choices

#### Sybase/Microsoft

Sybase, Inc.  
6475 Christie Avenue  
Emeryville, CA 94608  
(415) 596-3500

The Sybase/Microsoft SQL Server is adapted to OS/2 from the Sybase server for minicomputer environments. Sybase offers database servers on a number of platforms, including DOS, VAX/VMS, UNIX, and Mach. These servers have been designed to support on-line transaction processing in multiple environments. To fulfill plans for the OS/2 operating system, Microsoft licensed development and distribution rights for the Sybase server. Instead of providing a total database management system with both a server and a front end, Microsoft offers the server and programming tools—encouraging third-party developers to write front ends.

The SQL Server was designed for online applications with large numbers of users, multigigabyte databases, high transaction rates, and networked multimachine access. The multi-threaded server architecture is optimized to handle multiuser functions such as scheduling, task switching, disk caching, indexing, and locking. The SQL Server handles 30 users per megabyte of memory, minimizing hardware requirements and releasing more memory for disk caching to reduce disk I/Os. Stored procedures reduce network communication since one procedure call replaces many individual SQL statements. Since stored procedures are compiled and optimized in advance, they are processed 5 to 10 times faster than a single SQL command.

The Sybase/Microsoft SQL Server supports triggered procedures, defaults, and rules in table definitions, user-defined data types, and two-phase commit support. The Transact-SQL language adds procedural logic to standard SQL, for enhanced flexibility of SQL procedures. These procedures can be stored for later execution; in fact, entire procedures with computations and variable manipulation can be stored. This storage simplifies programming and enhances performance. However, SQL Server lacks explicit locking and exclusive locks on Select statements.

The SQL Server eliminates the need for data integrity rules in applications. Business rules are defined in the central dictionary and are applied to all applications. The Server supports several mechanisms for enforcing consistency within a single data field: NULL values, defaults, rules, and user-defined data types. For cross-field integrity, the SQL Server supports referential integrity via stored procedures and special stored procedures called triggers that are executed automatically when an insert, delete, or update to a data row is attempted. For distributed updates, the SQL Server provides a two-phase commit service, which guarantees that in case of hardware or software failure, the entire update will be backed out.

The Sybase family of servers is noted for server-enforced comprehensive data integrity, scalable performance, and support for open distributed database management. Back-end triggered procedures enforce referential integrity and relationships between tables the user chooses to define. Control-of-flow enhances its SQL language. Data definition is enhanced with support for user-specified default values, rules, and user-defined data types.

**Complementary Products.** Sybase offers a Sybase SQL Toolset for development and distribution of on-line applications. The SQL Toolset includes the Application Productivity Workbench, the Data Workbench, and the Programmable Libraries.

### Oracle LAN Server

Oracle Corporation, Inc.  
20 Davis Drive  
Belmont, CA 94002  
(415) 598-8000

Oracle Corporation has long espoused the benefits of cross-platform distributed computing, providing its first client/server architecture in 1985. Oracle offers its server technology across multiple platforms, with an emphasis on compatibility and portability. The Oracle solution includes a full set of development tools and a mature SQL implementation. Oracle LAN Server for OS/2 is essentially a large system product ported to OS/2 with only minor changes. The Oracle LAN Server is available for OS/2, UNIX System V/386, Banyan Vines, and for NetWare 386. It is part of a complete database management system, including a front-end interface and a complete set of data management tools.

Oracle's SQL is extremely strong. Support for user-programmable "triggers" help automate the data entry process. Because the program is available on such a wide range of hardware platforms, it is appealing to organizations that incorporate mainframes and minicomputers, as well as PCs, in their systems.

All four versions of the Oracle LAN Server are based on Oracle Version 6 technology, and have been optimized for OLTP support, decision support applications, or combinations of the two. Both horizontal scaling (adding more servers to the LAN for increased performance) and vertical scaling (moving data to another machine running a faster operating system or processor) are supported.

Oracle LAN Server provides a good selection of features, except for stored procedures. Oracle supports use of program arrays as SQL variables. File configuration options include indexing and clustering, fine-tuning performance, backup, and database administration. Compatible servers are available in a number of operating environments, and it supports almost any LAN. Its database link feature supports joined selects from databases in different locations.

Oracle offers a strong SQL implementation, a number of syntax options, special functions, multiple configuration, and backup/recovery options. Programming features include array variable support in SQL statements.

The Oracle LAN Server is available for Banyan Vines, Novell Netware (SPX/IPX and NetBIOS), 3Com 3+ Open (named pipes, NetBIOS), and the IBM LAN Server (NetBIOS). A Netware Loadable Module (NLM) extends the Novell operating system to integrate the Oracle server for Novell, effectively making the server a part of the operating system. Integration with larger systems is enabled through built-in TCP/IP, APPC/LU6.2, and Asynchronous support. In addition, Oracle servers can support non-PC clients and allow transparent data access across the mixed environment.

A number of PC software vendors, including Lotus, Ashton-Tate, and Borland, have endorsed the Oracle Server and are committed to developing versions of their products that will access and update Oracle Server data.

**Complementary Products.** Tools for the Oracle environment include OracleCard for both the Apple Macintosh and Microsoft Windows, SQL\*Plus, SQL\*Forms, SQL\*ReportWriter, SQL\*Menu, Oracle for 1-2-3, SQL\*Calc, Easy\*SQL, SQL\*Graph, Oracle\*Financials, Oracle\*Mail, SQL\*Connect, CASE\*Method, CASE\*Designer, and CASE\*Dictionary.

**Gupta SQLBase**

Gupta Technologies, Inc.  
1040 Marsh Road, Suite 200  
Menlo Park, CA 94025  
(415) 321-9500

Gupta Technologies' SQLBase Server was developed specifically for PCs. SQLBase under DOS became the first SQL database server for PC LANs, while the OS/2 version is relatively new. A major upgrade, Version 5.0, provides referential integrity support to ensure consistent database updates. Gupta also offers the front-end SQLWindows, which features a graphical interface. Gupta is publicly committed to development of graphical front-end tools.

The Gupta product line emphasizes connectivity between databases, with tools that enable SQLBase databases and SQLWindows applications to interact with other programs' databases. Gupta and Microsoft recently announced SQL Windows for SQL Server. Gupta Technologies has concentrated on providing quality front-end tools for data access, intending that a strong presence in the tools market will increase sales of its back-end DBMS. As a result, Gupta offers some of the strongest front-end development tools available.

The SQLBase Server supplies most of the features common among its competitors. Its SQL language implementation includes a variety of special functions, including financial functions not offered by other products. The SQLTalk for Windows interface permits menu access to most administrative functions on DOS workstations. Stored commands can be retrieved and executed in a batch. Like SQL Server, SQLBase lacks explicit lock capability and the ability to acquire an exclusive lock on a Select.

Novell owns 20% of Gupta Technologies.

**Complementary Products.** Gupta's SQL products include SQLBase (a server), SQLWindows (a Windows development tool), and SQLGateway, SQLRouter, and SQL-Host (connectivity software). SQLBase runs under DOS and OS/2 on most PC LANs. Gupta has shifted its emphasis from database engines to development tools, and it is unclear whether the company will continue to emphasize development of its SQL engine.

**IBM OS/2 Extended Edition Database Manager**

Old Orchard Road  
Armonk, NY 10504

Contact your local IBM representative.

IBM OS/2 Extended Edition is an extension of the OS/2 operating system. Unlike products that work with a separate base operating system and LAN operating system, IBM promotes a total IBM solution including the operating system, database, and communication between systems. The database management portion of Extended Edition cannot be used with a different operating system or LAN, and IBM supports the package only with IBM hardware, but an unbundled version of the database server will support PC clones. The Query Manager front end provides an intuitive, menu-driven interface for data retrieval and for all other aspects of database administration.

Ad hoc queries are supported with a built-in query manager, and high-level queries are supported by the SQL implementation. The IBM Database Manager supports a relational database model and is upward compatible with DB2 and SQL/DS systems. Facilities for data entry, editing, queries, and report writing are provided.

Extended Edition includes referential integrity constraints as part of table definition, but otherwise its SQL language is limited, with few special functions and no ability to update a column with a subquery. A workstation program can execute a program on the server by using the remote execute feature. IBM provides options for concurrency control, offers embedded SQL for programming, and provides a powerful interactive SQL interface (available only for OS/2 workstations).

**Borland Paradox Engine 2.0**

Borland International  
1800 Green Hills Road  
P.O. Box 660001  
Scotts Valley, CA 95067-0001  
(408) 439-1622

Borland's Paradox Engine is a unique product—it is at the core of all Borland business products, providing a common data and file handling mechanism, and it is also sold as a standalone product for database server and application development.

The Borland Paradox Engine 2.0, released in March 1991, provides users with ANSI C, Turbo C++, Borland C++, and Turbo Pascal support. Windows 3.0 programs are supported by supplying the Engine as a DLL. Paradox Engine 1.0 applications are completely supported. Applications developed with the Paradox Engine can create, read, and write Paradox database tables, for one or multiple concurrent users. The Paradox engine is composed of a number of elements—a C library file, the Pascal TPU, and a Dynamic Link Library that can be integrated into applications. The resulting application .EXE file provides the speed and size benefits of a conventional C or Pascal program.

The engine's C library includes Borland's Virtual Runtime Object-Oriented Memory Manager (VROOM). VROOM is designed to provide a means for writing programs with improved memory management and overall performance. VROOM programs can operate outside the standard DOS 640K limit, larger applications can take use extended or expanded memory. VROOM permits on-demand dynamic loading and unloading of modules in a running application.

The Turbo Pascal Unit (TPU) is a separate Turbo Pascal Compiler module that can be linked to the Paradox Engine.

An Application Programming Interface (API) provides developers with a library of more than 70 data handling functions that can be linked into applications to enable manipulation of Paradox tables in single-user and multiuser environments.

The Paradox API permits Paradox programs to:

- create, read, and write Paradox tables, records, and fields;
- utilize explicit file and record locking and concurrent execution with Paradox and interactive Paradox Application Language applications;
- import data from serial communications; and
- create standalone (unlimited royalty-free) runtime applications.

The Paradox Engine supplies a C header file (PX-ENGINE.H), and libraries (PXENGTCL.LIB,

PXENGTC2.LIB, and PXENGTCMSL.LIB) for use with Borland C and C++, and Microsoft C 5.1 or later.

Turbo Pascal Units PXENGG55.TPU and PXENG60.TPU, and an overlay file for Turbo Pascal (PXENGINE.OVL) are included for use with Turbo Pascal 5.5 and 6.0.

In order to facilitate development of Windows 3.0 applications, a dynamic link library containing engine functions (PCENGWIN.DLL) and a library (PXENGWIN.DLL) are provided. The DOS C header file (PXENGINE.H) is also used for Windows 3.0 development.

Paradox, the Paradox Engine, and Paradox Engine applications support IBM Token-Ring or IBM PC Local Area Network Program 1.12 or higher, Novell Advanced Network 2.0A or higher, 3Com 3+ and 3+ Open Networks, Banyan Vines 2.10 and higher, AT&T StarGROUP for DOS 3.1 and higher, and 100% compatible networks running under DOS 3.1 or higher.

**Complementary Products.** Paradox SQL Link database front end (1.1) provides a connection to the Sybase SQL server on Digital VAX/VMS and other UNIX platforms. The Paradox SQL Link translates Paradox application language (PAL) queries into SQL queries which are passed to

the server, then returned as Paradox tables. Paradox SQL Link supports IBM's EE 1.2 Database Manager, Microsoft Server 1.0, Oracle Server 6.0, and Digital Rdb/VMS.

ObjectVision is a visual programming tool for Windows 3.0 application development. Designed for nontechnical end users, ObjectVision can read and write Paradox, dBase, and Btrieve data format files.

---

## Conclusions

Microcomputer database management software is moving away from standalone single user systems and into multiuser, distributed systems. The workstation/server environment allows users to maximize the processing power of the personal computer and maintain local control. Although some users may not be ready for implementing a distributed database, this is the time to learn about the technology and begin examining the choices. In the fiercely competitive business world, the advantages of a comprehensive information system that provides ready access to all corporate data may prove the difference between success and failure. ■





# An Overview of Technical Workstations

## Datapro Summary

Technical workstations are designed specifically for complex applications in engineering, software development, artificial intelligence, imaging, document preparation, technical publishing, and financial environments. The workstation market continues to grow at an accelerated rate as users demand greater performance and processing power at competitive prices on their desktops. Vendors vie for the lead, announcing new low-end and high-end systems, lowering prices, and increasing system performance.

### In this report:

Operating System..... 3

Implementing RISC..... 3

PCs versus  
Workstations..... 4

## Technology Overview

### The Technical Workstation

A technical workstation comprises hardware and software designed specifically for complex applications in engineering, software development, artificial intelligence, technical publishing, document preparation, imaging, animation, and financial environments. The workstation combines interactive and automated tools for the design and development of a wide spectrum of products.

These days, a "typical" workstation configuration includes a high-speed 32-bit CPU, the UNIX operating system, from 4M to 10M bytes of memory, 100M to 600M bytes of disk storage, a 19-inch monochrome monitor with resolution of 1,024 by 1,024 pixels, X Windows or equivalent, and Ethernet networking capabilities.

The following paragraphs describe some important workstation characteristics and features.

### Performance Range

Technical workstations deliver anywhere from 1 (older models) to 128 (for a multiprocessor system) million instructions per second (MIPS) of performance. For example, the popular entry-level models such as Sun Microsystems' Sun-3/80 and Hewlett-Packard's HP 9000 Model 340 run at rates

ranging from 3 to 4 MIPS. At the upper end, the Stardent 3000 runs at 128 MIPS, and the HP/Apollo Series 10000 runs at 30 MIPS (these performance figures are for multiprocessor systems). "Average" single processor workstation performance, however, ranges from 10 to 14 MIPS.

Most vendors now also provide performance figures in millions of floating-point operations per second (MFLOPS). For example, the Sun SPARCstation 330 provides 2.6 MFLOPS, and a fully configured Stardent 1040 provides 40 MFLOPS.

### Central Processing Unit

The technical workstation's central processing unit (CPU) features a 32-bit microprocessor, a floating point co-processor, a graphics processor, a display management unit, and a memory management unit (MMU).

Today, workstations employ several different types of microprocessors: Motorola 680X0, Intel 80386, and those based on reduced instruction set computing (RISC) technology. Several workstation product lines—such as Digital's VAXstation line—use proprietary chips, although the company has recently committed to supporting an open systems environment.

### Motorola 680X0

Most workstations use an industry-standard chip for the central processor. Right now, the most widely employed is

Motorola's 32-bit MC68030, ranging from 25 megahertz (MHz) to the recently introduced 50MHz version. For example, Hewlett-Packard's newest workstation offerings, the HP 9000 Models 345 and 375, use the MC68030 running at 50MHz. In addition, HP has stated its intention to upgrade those systems to support the MC68040 when it becomes available.

### Intel 80386

Most high-end PCs used as technical workstations have an Intel 80386 microprocessor. Sun Microsystems, however, is the only major workstation vendor to use an Intel microprocessor in its systems (Sun386i). (Hewlett-Packard's Vectra PC is also based on the '386, however.) Although the Sun386i is based on an Intel microprocessor, Sun has done little to enhance the system and has chosen to focus on its RISC-based workstations instead.

### RISC

Increasingly, workstations such as IBM's RISC System/6000, Hewlett-Packard's HP 9000 Series 800 and Apollo Series 10000, Sun Microsystem's SPARCstations, Digital Equipment's DECstations, Silicon Graphics' IRIS workstations, and Stardent's workstations use microprocessors based on a reduced instruction set computer architecture. The RISC architecture offers faster instruction execution than the traditional, microcode-heavy complex instruction set computer (CISC) architecture, which has come to characterize superminicomputer-class systems. The RISC architecture speeds up computer operation by reducing the size of the instruction set and maximizing the number of instructions that can execute in a single machine cycle. (More information on RISC appears in the Implementing RISC section.)

### Floating-Point Co-Processor

A floating-point co-processor—an accelerator subsystem—provides increased speed and accuracy in numerical computations. It is especially useful because workstations often perform processing for computation-heavy applications including econometrics, chemical analysis, and aerospace engineering.

### Floating-Point Accelerator

Many workstations support a floating-point accelerator, which improves the system's speed on compute-intensive applications.

### Graphics Processor

The graphics processor increases the speed of graphics task execution. It is built specifically to handle complex graphics symbol manipulation in drawing/diagramming, modeling, image processing, and simulation tasks.

### Display Management Unit

A display management unit translates system-level images into screen images. It is responsible for providing high screen resolution for tasks such as text processing, typesetting, graphics, and imaging. It also provides the coloring for the image display.

### Memory Management Unit (MMU)

The memory management unit supports the virtual memory architecture of a technical workstation, which permits programs whose size exceeds physical memory to run on the workstation. In effect, the memory management scheme treats disk storage as part of main memory.

The MMU works in conjunction with the operating system to bring virtual memory pages into, and remove them from, main memory for mapping processes (programs or portions of programs), for doing reads and writes, and to provide the microprocessor and I/O devices with access to processes in system memory.

### Graphics Processors and Accelerators

Graphics processors work closely with the workstation CPU to perform all graphics applications. Graphics accelerators allow users to perform applications such as imaging, high-end visualization, and simulation. They facilitate the simultaneous acceleration of multiple on-screen windows.

### System Memory

In order to run applications effectively, workstations require a minimum of 4M bytes of memory. Over the past few years, we've seen maximum memory expansion capabilities grow tremendously. In a number of cases, main memory now can be expanded up to and beyond 128M bytes. For instance, the Digital DECstation 5000 accommodates up to 120M bytes of memory and the Sun-3/470, IBM RISC System/6000, and others each support up to 128M bytes of system memory.

### System Buses

Most workstations use either the de facto industry-standard Multibus or VMEbus to support peripherals such as disk storage units, magnetic tape drives, printers, display monitors, and communications controllers. By using an industry-standard bus, workstations gain access to a wide range of third-party devices, which are frequently less expensive than those supplied by the workstation vendor.

### Disk Storage

Workstations must provide a substantial amount of disk storage to support the complex applications they execute. The program code and data files used in running demanding CAD/CAM/CAE, imaging, animation, and other applications consume large amounts of disk storage space.

Technical workstations support more disk storage than microcomputers or even, in some cases, traditional minicomputers. Workstations commonly accommodate one to four fixed-disk storage devices with formatted capacities ranging from at least 71M bytes to 800M bytes and higher. Many of the newer high-performance workstations support as much as 3G to 10G bytes of disk storage. In contrast, many PCs support one or two fixed-disk storage devices with formatted capacities ranging from 20M bytes to 40M bytes.

### Display Units

Technical workstations support high-resolution monochrome, grayscale, and color display monitors. Typically, technical workstations use displays with screen resolutions averaging 1,024 (horizontal) by 1,024 (vertical) pixels or 1,280 by 1,024 pixels. In contrast, IBM PC microcomputers support monitors with typical resolutions of 600 by 480 pixels.

### Input/Output Devices

Workstations support a variety of input/output units. The options depend upon the workstation configuration, workstation application, and designer requirements. For example, input units supported include a keyboard, mouse, dials, tablets, and digitizers. A keyboard enters alphanumeric

data, text, and graphics-generation commands. A mouse manipulates data and objects on the display screen, eliminating the need to enter commands through the keyboard. Tablets create freehand drawings on-screen and digitizers create online images of graphics residing on source documents.

Workstations also support plotters, printers, and video; still or motion film units for hard copy output; and either a diskette or magnetic tape for data storage and retrieval.

### Operating System

Most workstations use a UNIX operating system based on AT&T's UNIX System V or the University of California at Berkeley's 4BSD version; some workstation operating systems combine both UNIX versions. UNIX is the workstation operating system of choice because it is a readily available system tool, licensed easily from AT&T, saving vendors the expense of developing a proprietary operating system from scratch. Furthermore, UNIX is widely employed in workstation-based applications because UNIX gives users access to a growing base of off-the-shelf software developed for those environments. UNIX also has intrinsic features well suited to software development—a task for which workstations are frequently employed.

Some vendors offer both a proprietary, multiprogramming, multitasking operating system and a UNIX operating system. The former is offered to satisfy the installed base and the latter to appeal to the growing number of customers (government, for example) that require UNIX. Digital Equipment Corporation offers both its VMS operating system and ULTRIX, its implementation of UNIX, for its VAXstations and DECstations. Future Digital products, however, will undoubtedly concentrate on ULTRIX.

### Windowing System

Most workstations are multitasking, running more than one application program at a time in on-screen windows, and supporting multiple users. Windowing features are a must for technical workstations, and most workstation vendors support the industry-standard X Windows interface definition. Windowing features allow the simultaneous display of multiple unrelated events, permitting the designer to view and study several displayed documents, or views of a single design. With windowing, the engineer typically selects the desired window or object via a mouse, keyboard, or some other interactive device. The window can be enlarged to show its contents in greater detail.

### Graphics System Software

Graphics system software is required to control graphics processing and create graphics applications. It offers system-level support for the creation and manipulation of designs and images; supplies application developers with utilities and subroutines for creating graphics applications and incorporating graphics into existing applications; and offers end users the tools for creating and editing drawings, drafts, designs, and images and for running simulations.

Graphics system software generally conforms to industry standards such as Graphical Kernel System (GKS), which handles both 2-D and 3-D graphics processing; GSPC Core Proposed Standard Graphics Software System (Core), which supports 2-D and 3-D graphics; and Programmer's Hierarchical Interactive Graphics System (PHIGS), which provides for three-dimensional graphics computing.

### Database Management Systems

Workstations use standard relational database management systems (DBMSs) for data manipulation and applications development. Database systems such as Ingres, Unify, and Oracle are widely installed on technical workstations not only because of their data manipulation and programming qualities, but also because they can run under UNIX.

### Local Area Networking

Local area networking is key in the workstation scheme, since networks allow individual workstations to share programs, databases, and peripherals, such as mass storage devices, printers, and telecommunications lines for interacting with a host computer. Most workstation vendors support IEEE 802.3 Ethernet, the industry standard.

In addition to Ethernet, some workstation vendors also support a proprietary network. Proprietary networks maintain compatibility with existing network schemes and meet distributed processing requirements (for example, data transfer speeds and system addressing capabilities) not provided by Ethernet. For example, HP/Apollo Domain systems support both the Apollo Token-Ring LAN and an Ethernet network.

### Implementing RISC

Reduced instruction set computing (RISC) technology is one of the most talked-about architectural schemes for providing improved price/performance. Recently, several major computer system vendors have introduced RISC systems. RISC, however, is not a computing panacea; it is simply one of many architectural implementations designed to produce faster and more price-competitive systems. These systems are only helpful if they are used on applications that take advantage of RISC's enhanced performance.

Interest in RISC has become intense, because the technology can deliver increased processing power at lower costs than those for more conventional architectures. Hewlett-Packard/Apollo, Sun Microsystems, MIPS Computer Systems, Digital Equipment, Silicon Graphics, and IBM are some of the most active vendors in this technology area, introducing workstations based on RISC technology or employing RISC-like features.

As RISC research and development progresses, a common definition and a set of design principals are emerging. The fundamental theme of a RISC architectural design is maximizing the speed of the CPU by implementing "single-cycle execution." The fastest rate at which a computer can execute instructions is one per machine cycle. RISC designs strive to execute each instruction in no more than one machine cycle. The single-cycle execution rate is the basis for the performance gains made possible by RISC.

For manufacturers, RISC provides a means to develop a system significantly faster and at a lower cost than a conventional system. The cost of designing a new computer is closely related to the complexity of the instruction set: the more complex the instruction set, the more hardware logic is needed to implement and support it. Designing and developing the hardware and microcode to support complex instructions are time-consuming tasks. Microcode and processor design, development, and testing require many worker-hours and consume a major portion of any development budget. Because RISC designs are much simpler

than conventional designs, computers with RISC technology require less time to design and are less expensive to build.

By reducing the cost and time for building and delivering computers, RISC computer manufacturers can pass the savings onto the customers in the form of shorter and less expensive development cycle times and (consequently) improved price/performance. RISC manufacturers can deliver products more quickly and at a lower price than those vendors that build their systems using complex instruction sets.

While RISC can save money for developers and users, there are drawbacks in implementing and buying it, as well. Despite RISC's favorable price/performance ratio and rapid development times, established computer vendors with significant installed bases of conventional machines have trouble migrating to RISC designs. The reduction of the instruction set leaves out complex routines required by older systems; those routines must be compensated for before the software can be ported over. In shifting to a RISC architecture, vendors risk stranding their existing customers with large investments in software that might not run on the new machines. Therefore, RISC vendors need to ensure complete compatibility with earlier lines of computer systems, an expensive and difficult task.

Most RISC machines run under a UNIX operating system, although it is generally agreed that there is little about UNIX that makes it operationally better for RISC. Rather, the relationship between UNIX and RISC is primarily economic. Vendors have ported UNIX over to the RISC architecture because it enables them to come to market with a product much more quickly and at a lower cost than if they undertook the massive effort of developing a proprietary operating system from scratch. Thus, although

UNIX must be modified for the RISC architecture, times and costs for software research and development are still kept to a minimum.

### PCs versus Workstations

Personal computer-based technical workstations, commonly called technical PCs, have quickly become an accepted segment of the market. In fact, their increasing speed and functional capabilities, along with their lower prices, threaten sales of low-end workstation systems.

An abundance of third-party hardware and software has turned the IBM Personal Computer, 386-based systems, and compatibles into design tools for what-once were traditional workstation applications. Hardware accelerators and high-resolution graphics controller boards, as well as the right software package, transform personal computers into high-powered workstations.

The Compaq 386, IBM PS/2, and Apple Macintosh are growing in popularity and now also threaten the traditional workstation vendors. In fact, these systems compete head-on with entry-level desktop systems from Sun, Hewlett-Packard/Apollo, and Digital Equipment. As PCs move up in performance, entry-level workstations are dropping drastically in price.

Each workstation "class" provides some interesting benefits. For example, PCs provide a vast array of applications software. The peripherals PCs require are also relatively inexpensive. More important, most PCs are fully IBM compatible.

Workstations, on the other hand, offer users high resolution and fast processing—attributes that are often missed on a technical PC. Most workstations support superior networking capabilities, as well as a wealth of sophisticated software. ■

# The Evolution of LAN Operating Systems

## In this report:

LAN Operating Systems .....	3
NOS Market Issues .....	4
NetWare .....	5
LAN Manager .....	5
VINES .....	6

## This report will help you to:

- Understand the evolution of LAN servers which produced today's client/server computing model.
- Know how the network operating systems (NOSs) manage server operations.
- Compare popular LAN operating systems to determine which best meets your organization's needs.

The network operating system (NOS) is the key component in most local area networks (LANs). This report provides an overview of the evolution of LAN operating systems, plus a look at the three leading NOSs on the market.

## Introduction

End-user computing, from the early 1980s to today, has evolved from unconnected mainframes, minicomputers, and PCs to networks of devices providing peer-to-peer

communications. Likewise, servers on LANs have evolved to what is today known as the client/server computing model.

This evolution has resulted in the popular acceptance of LANs by the business community. The proliferation of LANs has brought to the forefront the importance of the software that provides the basic network features and functions—the network operating system (NOS). LAN operating systems are the key network component and now drive many LAN purchase decisions.

A lively battle is now being waged for dominance in the LAN operating system market. Novell, with its ubiquitous NetWare, is the undisputed market leader. Software giant Microsoft has launched a challenge with its OS/2 LAN Manager product. LAN Manager, initially licensed as an OEM product to several LAN vendors (IBM, 3Com, and

---

This report was prepared exclusively for Datapro by Victoria Marney-Petix. Ms. Marney-Petix designs and delivers on-site seminars on LANs, internets, and network management and consults with vendors on new product development and competitive strategies from her office in Fremont, CA. She teaches in the Telecommunication Program at San Francisco State University Extension and is the author of *Networking and Data Communications* (Prentice-Hall, Inc., 1986).

Ungermann-Bass, among others), has been a disappointment. Microsoft has since changed its strategy—it is marketing LAN Manager directly and plans a massive assault on Novell's leadership position. Meanwhile, Banyan's VINES has won popular acclaim but only a small market share.

---

### Server Evolution and Client/Server Computing

Computing resources in the early 1980s existed as unconnected personal computers (PCs) on end-user's desks or as centrally located mainframes connected only to each other. The local area network (LAN) revolution brought the PCs into peer-to-peer communication with each other and eventually brought even the mainframes into the LAN. These LANs used servers<sup>1</sup> to bring printer sharing and local file access to end users.

The earliest servers consisted of software slapped onto the cheapest hardware platform available; they were designed for printer sharing and generally were not even dedicated to that task. The dedicated but still PC-based server was a step up. Today's high-performance servers are generally resident on dedicated platforms with workstation or minicomputer<sup>2</sup> engines, giving network managers more MIPS, faster processing, multitasking operations, and other performance and reliability advantages.

As business needs changed, network architectures evolved to serve those needs. The distributed server brought faster response time to the network's users and less stress to the internet infrastructure (bridges, routers, etc.) as end users directed more requests to local servers and fewer requests to central mainframes. Of course, distributed servers do increase the network management burden because there are more devices to manage.

This LAN proliferation irrevocably altered the way PC users thought of their desktop machines, but it did not really alter the way that mainframes were used in the corporate internet. The mainframe, even as the 1990s begin, is still seen as the core of the corporate information strategy, the mainstay of application management and information dispersal.

The final step in server evolution—called the client/server computing model—carries the server to its logical ultimate role as the primary focus of information dispersal in the internet. Users in a

client/server computing environment can access information stored anywhere in the internet, providing an entirely new information infrastructure.

### Rise of Client/Server Computing

In the client/server computing model, the server and its clients share the computing role between them. The server performs database access and intensive computing tasks (the back-end processes), and the client performs the display and user interface tasks for those calculations (the front-end processes). In a standard server model, the server receives a database request and transports a copy of the entire database to the requestor. This clogs the network and raises security concerns. In the client/server computing scenario, only results are transmitted, conserving network bandwidth. The client/server computing model allows each device to contribute what it does best and distributes the computing load throughout the network.

Each part—both server back end and user front end—can be developed, migrated, and serviced separately. The parts are developed so as to optimize a particular function, either database retrieval or information display, so that each device in the network can be used most effectively. Both the use of computing resources and the use of corporate funds are leveraged with low-cost, high-MIPS server platforms and a preserved investment in hardware, software, and training.

The client/server computing model helps bring essentially flat networks into a more hierarchical framework for easier and more effective management. As networks grow in size, segmenting managed resources into logical management domains allows managers to automate some tasks, zero in on problem areas faster, and spend time analyzing and planning rather than just fire fighting.

The software entities that manage server operations are called network operating systems (NOSs). In the same way that a device operating system manages the interface between a device's basic functions and its user applications, the NOS manages the interface between the network's underlying transport capabilities and the applications resident on servers. To manage a network, especially a modern network containing distributed servers, you need a powerful and effective network operating system.

Figure 1.  
The NOS in the Network Software Stack

	Novell Stack	TCP Stack
Layers	Applications	Applications
7	Net Man	Net Man
5-6	NetWare	LAN Manager
4	SPX	TCP
3	IPX	IP

### LAN Operating Systems

Network operating system software resides at the equivalent of the OSI Reference Model's Session and Presentation Layers, with some Application Layer functions thrown in. Figure 1 shows where the typical NOS fits in some popular protocol stacks, using Novell's NetWare and Microsoft's OS/2 LAN Manager<sup>3</sup> as examples.

NOSs evolved primarily to solve an existing problem: giving multiple users access to a PC LAN's servers. Because the earlier mainframe- and minicomputer-dominated networks did not stimulate the development of servers, they did not depend heavily on NOSs. The evolution of the NOS was intimately tied to the market success of the PC LAN.

The NOS has now moved beyond the confines of the purely or primarily PC LAN. In the typical hybrid internet, the NOS now regulates every user's access to information.

#### Server Architectures

As servers become the major platform for disseminating information in the corporation, the NOS becomes a critical tool in the network management arsenal. The following list shows the critical management needs in server design.

#### What Must a Server Deliver?

- Reliability
- Maintainability/upgradability

- Performance
- Services blind to user's
  - Device OS
  - Transport protocol
  - Upper layer protocols

Reliability and performance are clear needs in any network that carries mission-critical applications. If the server is the repository of critical applications, the network manager must be able to move users, services, and the hardware platforms themselves to new locations quickly and easily.

The ability to deliver services to any client device, regardless of its device operating system, is a clear necessity in a multivendor, multiplatform network environment. But users want more—including the ability to access servers using multiple transport protocols (Layer 4 in Figure 1) and multiple high-layer software (electronic mail, remote terminal access, directory services, etc.) protocols.

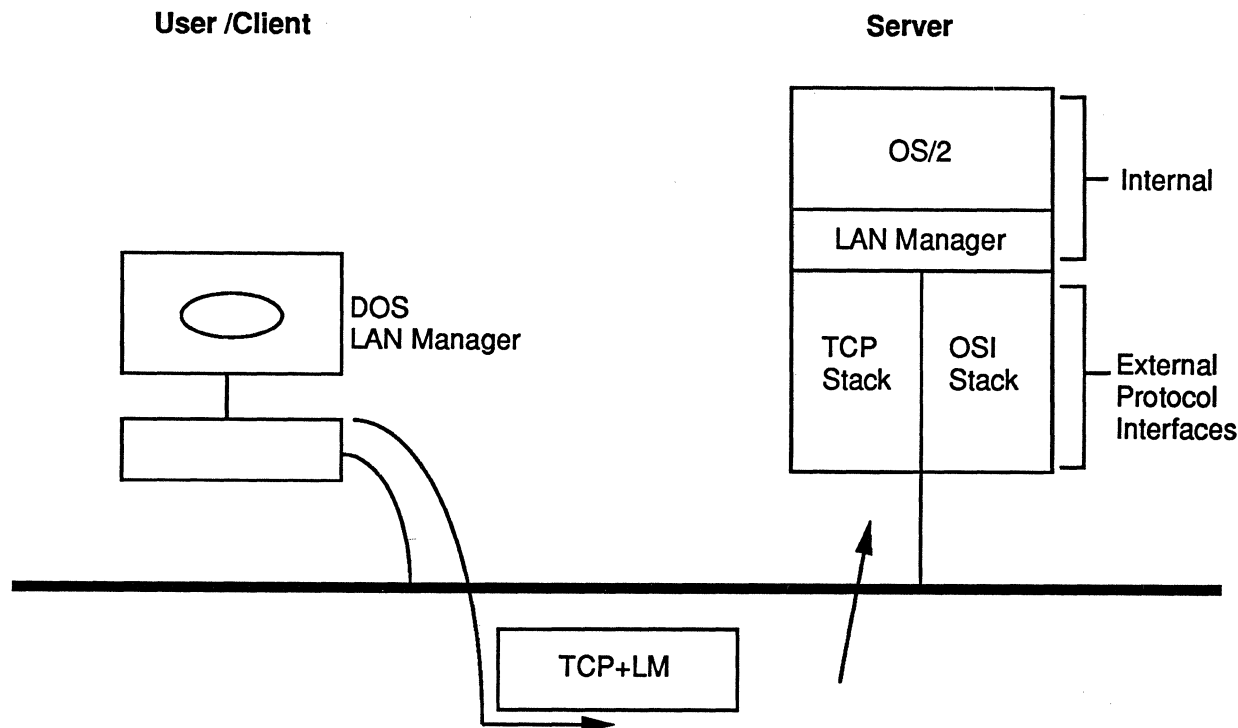
The NOS gives servers a chance to offer users uniform presentation of offerings. Figure 2 shows an end user and a server, both running their respective portions of Microsoft's OS/2 LAN Manager. The internal server engine should not affect the server's capability to offer software services to end users of DOS or any other operating system. In Figure 2, this particular end user has a TCP stack for transporting the packet containing the request; the server has front-end processing for both TCP and OSI packets. Another server might handle SNA and DECnet, for example.

#### Management in the NOS

In PC LANs, and small- to mid-sized LANs in general, a NOS is seen as primarily a server management tool. However, having an effective LAN operating system is an even more valuable asset to an internet manager. A single-server interface—which an operating system provides—eliminates an entire *class* of software incompatibility problems in the internet. In addition, an effective LAN operating system with management capabilities can isolate the remaining software problems to the server itself.

The general principle of good internet management is to embrace layer management—as the OSI model requires—so that problems at a particular layer are resolved or contained at that layer or

Figure 2.  
Client/Server Operations



the layer immediately above. The operating system can contain some problems at the layer it manages and prevent them from proliferating to the application or network layers and therefore out into the internet.

### NOS Market Issues

Now that we have looked at NOS management capabilities in general, we need a marketplace road map.

The most popular LAN operating system today is Novell's NetWare. Both IDC of Framingham, MA, and Dataquest of San Jose, CA, believe that NetWare's installed base accounts for almost three quarters of the total LAN operating system installed base. The most recent version, NetWare 386, runs on servers using the 80386 and 80486 chip and, if the appropriate add-on software is present, it can be accessed by end-user computers using the MacOS (Macintosh), DOS, UNIX, and OS/2 device operating systems.

Microsoft's<sup>4</sup> OS/2 LAN Manager uses a server engine running OS/2 as its device OS but is designed to be accessible to users running other

device OSs. With LAN Manager available from over 40 OEMs, most network managers can buy this NOS from their existing LAN supplier. (In addition, Microsoft now sells its version of OS/2 LAN Manager directly.) Each OEM's version contains slightly different features, allowing varying levels of interoperability with other products. AT&T's LAN Manager/X, for instance, is a UNIX version of OS/2 LAN Manager that allows devices running with DOS, OS/2, or UNIX operating systems to access the server. Microsoft's basic LAN Manager is available for devices running under DOS or OS/2.

LAN Manager, as a relatively recent entrant into the marketplace, has a relatively small installed base. Dataquest's projections see it moving into second place within five years, however, a figure few independent analysts dispute. LAN Manager, because its multiple licensees also have the right to make custom modifications, may eventually find market troubles in the incompatibilities among its many flavors. Since LAN archrivals



3Com and Ungermann-Bass are both LAN Manager licensees, for example, they need to differentiate their versions to establish a competitive advantage for their customers.

IBM's entry, OS/2 LAN Server, is also based on Microsoft's OS/2 LAN Manager, but with major modifications. IBM and Microsoft recently deepened their long-term technical alliance by announcing LAN Server's coming migration to code compatibility and a common set of APIs with Microsoft's LAN Manager. Users can expect to treat LAN Server as a subset of LAN Manager through the 1990s.

Now that we have a road map, let us take a more detailed look at the three top LAN operating systems: NetWare, LAN Manager, and VINES.

**NetWare**

The most recent NetWare version—named NetWare 386 because it is optimized for 80386 microprocessor performance—eliminated many major shortcomings of earlier versions, primarily by making it easier to add new users and new servers. Each server can theoretically serve up to 250 users, although users requesting complex tasks will reduce that maximum number considerably. NetWare's key features are as follows.

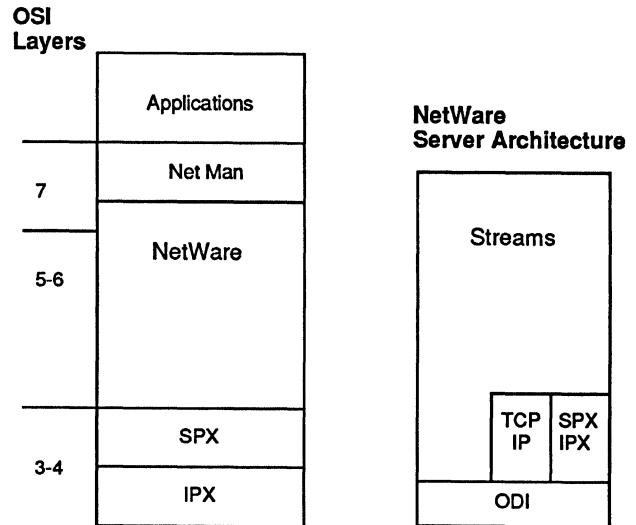
**NetWare Features**

- Centralized administration possible
- Extended File Salvage
- Fault tolerance: hot fixes, disk caching
- High-performance file system
- Supports multiple transport protocols

NetWare servers can use dynamic memory allocation and a high-performance file system to speed performance, while centralized administration improves security and boosts the efficiency of the management staff. Novell has a message handling system (MHS) for electronic mail subsystems and NetWare SQL as a database manager.

NetWare's Extended File Salvage could be a considerable boon in small networks: a server actually recovers only the space allocated to "deleted" user files when it runs out of free space. Thus, users with second thoughts may be able to get their discarded files back if the server's memory is not heavily loaded. Of course, the manager of a *Fortune* 100 internet will probably never have enough

Figure 3.  
NetWare Server Architecture



server memory to waste in saving deleted files for careless end users. This is clearly a powerful feature only for underused servers.

The server engine's architecture is based on UNIX' Streams and the Open Data Link Interface (ODI) codeveloped by Novell and Apple Computer. Figure 3 shows how the parts of the NetWare architecture fit together. Streams can now encompass the TCP/IP stack, AppleTalk Filing Protocol (AFT), Sun's Network Filing System (NFS), and Server Message Block (SMB) and LAN Requestor software. Novell has announced that NetWare will eventually be capable of supporting X.400 electronic mail and other OSI protocols, including the TP4 transport protocol.

**LAN Manager**

If you peruse LAN Manager's various features in the following list, you will see many of the same selling points that NetWare claims. Fault tolerance, a high-performance file system, and centralized administration are all important to network managers, but they are not unique to LAN Manager.

**LAN Manager v. 2.0 Features**

- Centralized administration
- HPFS
  - Fault tolerance
  - Disk mirroring

- Hot fixes
- Domain Server
- Security
  - Access control lists
  - Audit trails for charge-backs
- Named Pipes
- Peer services

LAN Manager's more interesting features include the domain concept, Named Pipes, multitasking and multiprocessing operations, and the ability to audit specific Named Pipes. A network manager sets up a domain of servers and can then act on them as a group; this automates moves, adds, and deletes of users, applications, and security authorizations. Multiprocessing operations put multiple processors to work within one server. In LAN Manager, one processor will be devoted to certain operations—disk I/O, for example—and all disk read/write requests will be automatically routed to that processor, leaving the second processor free for other work.

Named Pipes is a powerful LAN Manager tool. A pipe is a Presentation Layer entity that allows process-to-process communication; it is created for a specific communication and disappears when the communication is complete. A Named Pipe is a permanent logical structure that gives users the opportunity to redirect interprocess communication in regular ways. Once a Named Pipe is listed in a server directory as a shared resource, it can be accessed by any user who needs it. Named Pipes gives developers a simple, generic, high-level interface for their API development. Managers can create scripts to automate their audits of Named Pipes for departmental charge-backs.

Peer services are becoming more important in segmented, workstation-oriented networks, which are typically the larger networks with the most complex topology. If a workstation can act as a nondedicated server to peers for a single low-level function—fetching mail, perhaps—while still acting as a standalone device for its end user and as a client to other dedicated network servers, the network's efficiency increases. Multiprocessing, especially the symmetrical form, helps a server's manageability by removing the processor as a potential I/O performance bottleneck. In symmetrical multiprocessing, the processors share the load

equally; the asymmetrical form dedicates each processor for certain tasks so they cannot load-share dynamically.

### VINES

VINES' primary niche is the large, complex network, so until recently Banyan was actually leading rather than responding to user needs. Only in the past three years has buyer understanding been equal to the tasks that Banyan designed VINES to accomplish.

### VINES Features

- Fault tolerance
  - Disk mirroring, hot fixes
- High-performance file system
- Security features
- Extensive peer services
- Symmetrical multiprocessing
- Global name service

VINES performs especially well compared to its rivals in networks with large numbers of users and servers and heavy traffic. It is particularly easy for network managers to add, delete, and move users and server offerings. In small networks, VINES can be slower than NetWare, and it definitely requires more memory in user devices.

One great VINES selling point—and a key to the “large networks” design goal—is StreetTalk, a global naming service. A global naming service, such as the CCITT/ISO X.500 Directory Service standard, provides a “yellow pages”-type service for all network users. Users log on to the network with a globally known logon name and access list, so they can access servers anywhere in the network. In networks without a global naming service, moving users, servers, or services puts a significant work load on network management staffs.

Neither NetWare nor LAN Manager compare favorably with VINES in this instance. NetWare 386 was slated to have a global naming service in late 1990; users frequently cited this lack as a major source of dissatisfaction. In NetWare, LAN Manager, and LAN Server, users log on to an individual server, not the network as a whole, so network managers must create scripts to log a particular user onto more than one server at a time. If the user or the server moves or if the access

rights to any server or service changes, the script must be changed. If the server moves and 50 user scripts must change, significant management effort is wasted. In most cases, the users will experience a time delay during which they cannot access the moved server—a serious matter if the server contains mission-critical applications!

Banyan has moved recently to modify VINES so that it will support rival Microsoft's LAN Manager NOS protocols and APIs. The two companies are swapping technical specifications for the NOS code so they can build future applications that will run on both NOSs. Banyan will be first out with a LAN Manager protocol stack in VINES, allowing its NOS to support mail, print, filing, and other

LAN Manager APIs (application programming interfaces), as well as the Named Pipes interprocess communications facility. The VINES StreetTalk facility will provide network management information on both VINES and LAN Manager nodes. The new version was expected to ship by late 1990.

---

## References

<sup>1</sup>A server is a software entity (with a generic or dedicated hardware platform) that delivers specific services to end-user devices. Refer to other DP reports for in-depth coverage of server technology.

<sup>2</sup>Frequently referred to as superservers.

<sup>3</sup>"LAN Manager" in this report refers to Microsoft's NOS, not an identically named IBM network management package.

<sup>4</sup>The original LAN Manager was codeveloped by 3Com for Microsoft. ■



# An Overview of PC-to-Host Communications Products

## In this report:

Products .....	8
Selection Guidelines .....	8

## Datapro Summary

Though not as active as it once was, the PC-to-host communications marketplace is still important to the growing number of companies that want access to information and applications residing on mainframes or minicomputers. In the past, companies such as DCA, Novell, and AST Research found themselves at the forefront as the PC-to-host marketplace enjoyed tremendous growth and expansion. These companies, particularly DCA, led the way with products ranging from simple 3270 terminal emulation products to remote emulation products to LAN gateways. Recently, though, the market has slowed. The dynamic growth of previous years has been replaced by a market now saturated with a wide variety of vendors and products. Many companies, unable to compete with the extensive product lines of vendors such as DCA and Novell, have found a home in niche markets such as micro-to-minicomputer links or Macintosh-to-host connectivity products.

## Technology Basics

PC-to-mainframe communications can be viewed as merely the substitution of PCs for dumb terminals. PCs were not designed for this role, however, and therefore have characteristics that must be accommodated through the technology and through careful user planning. To determine the best PC-to-host link for a particular situation, users must gain a basic understanding of the differences between mainframes and PCs and the elements that any communications system must employ.

Some fundamental operating characteristics are quite different. Mainframes and communications systems are both heavily involved with controlling a variety of concurrent tasks, whereas most of the 20 million PCs in use are designed to handle a single task at a time for a single user at a time. Most PC users have already encountered frustrations from the limitations of single-tasking systems. While many look with hope toward IBM's new multitasking PS/2-OS/2, its capabilities fall short of handling the complexities faced by PCs when connecting to large communications networks.

## Mainframe Communications

Mainframes have a different hardware architecture, usually with multiple supporting processors to expedite different I/O functions and multiprogramming, and so can efficiently handle large complex networks and databases. People who look at the fast-rising power of the processors in smaller, cheaper systems and predict the subsequent demise of the mainframe tend to forget these architectural differences that support various applications and greater system throughput. It may be cost effective to download certain jobs to a lower level processor (especially since the last generation's large system is this generation's midrange). Some complex jobs are always left over at the high end and new applications are becoming cost effective, as big mainframes can handle ever larger, more complex jobs.

Communications on a mainframe are usually handled by a separate, freestanding front-end processor (FEP) attached to an I/O channel. Small mainframes and most superminis substitute an integrated communications adapter capable of handling fewer lines. In IBM systems, the front end is tightly bound to the mainframe because the

telecommunications access method software ("VTAM" for SNA networks) resides in the mainframe central processor; the Network Control Program (NCP) that serves as the operating system for the front end must be generated by another mainframe program, the System Support Program (SSP). Some mainframe vendors have all this software in the front end, allowing more functional independence between the two processors.

The incoming communications lines attach to ports on boards that, in turn, are often organized into several groups or subsystems, depending on the system type, front-end model, and so forth. A single front end almost always handles a variety of line speeds and, less often, more than one line protocol and code. The front end, in turn, attaches to one of the mainframe's numerous general-purpose I/O channels that, on larger systems, are parallel-transfer multiplexer channels handled by a separate I/O processor. The front end may or may not share the channel with a subsystem of another type—disk or tape, for instance.

To provide configuration flexibility, mainframes have multiple I/O channels, each usually designed to attach up to eight subsystems with multiple devices per subsystem. Therefore, one channel can attach up to eight current IBM 3725 front ends, for instance, and the 3725 can conceivably attach up to 256 lines, each supporting multiple subsystem controllers that in turn attach 32 terminals in a cluster. Putting all this on one channel would create an impossible bottleneck, but the flexible attachment capabilities allow configuration adjustments for better performance. All large mainframes are designed to comfortably handle a number of front ends, and each physical front end can handle a number of logical networks.

These hardware elements form the backbone of the communications systems, but the real definition and management of the network reside in the software. Communications software on big systems is almost always in a strictly defined layer, separate from the application programs; earlier amalgamations of the two led to too much confusion and duplication of effort. PC communications usually require applications interfacing software added to the mainframe, as well as a program and/or board in the PC; users need to check whether they are expected to write the applications interfacing programs themselves if they are missing from a product. Most off-the-shelf products run on IBM or Digital systems, so this task is almost surely needed if the mainframe is from another vendor.

### PC Communications Hardware and Software

Most PCs cannot concurrently process an application and serve as a communications link to a host. A single-task computer can undertake only one function at a time—word processing, spreadsheet, database, or communications. The user must decide when to invoke the particular function, based on current needs. This type of serial work scheduling is normally suitable for human interactions; people's work habits are also typically single minded. The limitations of single tasking, however, are apparent when quick access to a spreadsheet program is needed in the middle of writing a large report on a word processor. To get it, a user would probably have to save his/her work, exit the word processor, and enter the spreadsheet program.

Communications interactions are quite different from the characteristically serial processing of most PC tasks. Communicating between a PC and the data center is not an end in itself, but rather an intermediary step in the handling of information. As such, it occurs as part of a job (from the user's point of view) and rarely as a job in itself.

Putting it another way, communications between a PC and the data center is naturally interactive. Single-task PCs, on the other hand, are serial oriented. Their "interactive" requests for communications service are therefore normally made outside the application, leaving the user with the problem of manually integrating the data received with local work or extracting and formatting data to send.

The most common communications implementation consists of a board with software support that allows rudimentary dumb terminal emulation. For example, the most basic 3270 emulation allows the incoming code to be displayed, but screen-by-screen printing must occur. Ways have evolved to allow file transfer, but not necessarily in the desired format. Of course, communications programs can be and have been written for the PC, adding multitasking and/or the specific code needed, so that integrating incoming code might be automatic; but this is difficult to implement on most existing systems due to the 640K-byte memory ceiling. Third-party vendors must be careful to find applications widespread enough to justify development costs.

### The Data Communications Network

The mainframe is a foreign architecture that requires compatibility adjustments. A PC user who wants to link with a mainframe soon discovers that he or she must interface to a separate networking world that requires a much more intimate knowledge of hardware. Regardless of the complexity of the PC-to-mainframe link, there are certain elements that must be considered and adjusted for before communications can occur at all.

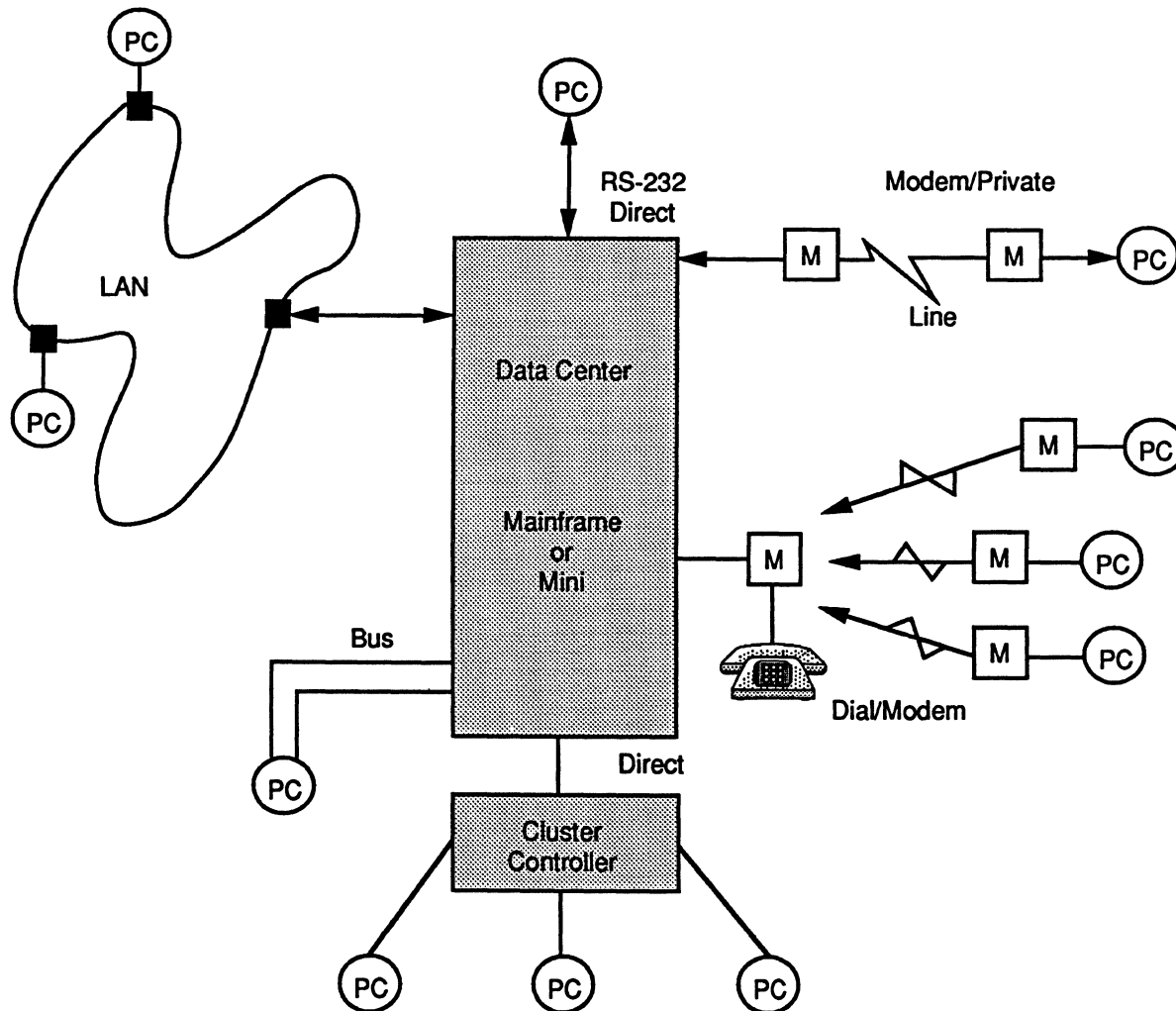
### Codes and Character Sets

The majority of current computers are fundamentally encoded in either the ASCII code, which uses six-, seven-, or eight-bit characters plus a parity bit, or EBCDIC, which uses eight-bit characters. IBM and IBM-compatible PCs, and all other PCs based on Intel microprocessors, use ASCII code, as do Digital VAX and 8000 series superminis and IBM Series/1. IBM System/36, System/38 minis, and System/370-compatible mainframes and superminis (3080, 3090, 4300, and 9370 product lines) all use EBCDIC. This means, ironically, that an IBM PC communicating with an IBM mini or mainframe requires code translation, whereas an IBM PC communicating with a Digital supermini does not. Admittedly, code translations are not complicated, except that special symbols in one character set that do not exist in the other are usually lost. Translation, however, must occur somewhere during the transmission—either at the outset in the PC, in a protocol converter or departmental node en route; in the receiving program in the front end; or in the destination application program.

### Protocols

*General Definition:* A protocol is a set of rules for formatting the transmission control codes and data; as such, it is the key to how the logical and physical networks interconnect. In all current protocols, headers and trailers are added to the data message to provide information on the source and destination application, the source and destination computers, network routing, validity checking, security codes, clocking bits, and so on. There may or may not be start/stop bits on individual characters. There are two classes of protocols: asynchronous and synchronous. Because protocols are added to the message to be transmitted

Figure 1.  
Hardware Connection Options



and then stripped away as the data reaches its destination, they are completely transparent to the end user. Protocols are an important compatibility consideration.

**Asynchronous Protocols:** Asynchronous protocols are lower speed connections with irregular timing, requiring start/stop bits to be added to each character after the session has been established. Session control information is rudimentary, and complex routing is not directly supported, although DECnet's sophistication and success shows there are various methods to get around this. Digital's VT52 and VT100, IBM's 3101 and 3160 series, and Teletype terminals all use asynchronous protocols with the ASCII code.

**Synchronous Protocols:** Synchronous protocols depend on finely tuned clocking to distinguish characters or bits that have been grouped together; this allows higher transmission speeds and better reliability (at the cost of more complex implementation logic). There are two types of synchronous protocols: byte oriented and bit oriented.

**Byte-Oriented Protocols:** Past implementation of byte-oriented protocols such as BSC have tended to be somewhat machine dependent, with certain characters having

explicit control functions; as a result, there are slightly different versions for the IBM 3270 interactive terminals, the 2780/3780 batch terminals, etc. This relative lack of transparency also means that different terminal types usually cannot be mixed on a line.

**Bit-Oriented Protocols:** Most bit-oriented protocols, such as IBM's SNA/SDLC and ISO's X.25/HDLC, on the other hand, are packet-switching bit-oriented protocols. They are designed to get around these limitations by defining the "data" as a completely unpredictable bit stream that is allowed to have any sequence of bits except one (the data delimiting markers), whereas control characters with specific functions are recognized primarily by their position before and after the delimiting markers. This system easily transmits any type of data or coding, including graphic material. It also provides great routing flexibility, since a new address, or any other new information, can be added by just attaching new headers/trailers and delimiters around both the message and its original headers/trailers, for an intermediate destination. This treats the original headers as if they were part of the "data" bit stream. At the intermediate destination, the new headers/trailers are stripped off, and the original ones emerge. Multiple "envelopes" can be used for routing to a number of nodes. Since the

control and data fields are firmly separated and the control code standardized across all products, multiple device types can share a leased line.

#### Attachment Mode

Data communications support in existing facilities can be classified according to the type of information exchange supported. Older batch terminal systems, such as the IBM 2780/3780, were designed to upload/download large files after the file had been created off-line on punched cards or data entry systems. Keyboard-oriented interactive terminals, such as the Digital VT100, IBM 3270, or IBM 5250, were designed to work on files that remained on the mainframe by sending or receiving data in smaller amounts, usually a screen at a time, in conversational mode, with some optional printing capabilities. PCs are far more likely to emulate existing interactive terminals and add batch file transfer capabilities by software extensions.

#### Attachment Method

The actual attachment hardware consists of a plugged-in circuit board with a cable connection either directly to the computer or to a modem. The most common interface for serial devices and communications is an RS-232-C plug with only a subset of the pins implemented (4 to 7 pins for asynchronous communications, 9 to 12 pins for synchronous communications). The IBM first-generation PC is standardized on this interface, but wide variations are permitted by this "standard." The wire leading from the board is connected to a modem, and the wire leading from the modem is connected to the telephone system.

Boards for terminal cluster device emulation expose other interfaces and require the appropriate cabling. In the case of 3278 emulation, a coaxial cable connects the PC to the communications cluster controller or an integrated controller on a low-end mainframe. In the case of 5251 emulation, a twinaxial cable connects the PC to the cluster controller or to an integrated controller on a System/36 or System/38.

PCs connected to local mainframes do not require modems; their cabling, therefore, can be a little different. If the computer is nearby (it must be within 50 feet to use the RS-232-C interfaces), asynchronous terminals can use a null modem cable for attachment. Synchronous terminals, however, need the clocking signals from the modem for proper transmission. They must use a different device with a clock, called a modem eliminator. A local device that is farther away—on another floor, perhaps—can be reached by a limited-distance/short-haul modem.

#### Facility Layout

PCs are rarely used close to the data center, so consideration must be given to the impact of the relative locations of the equipment. This is particularly important where the attachment will not use a conventional phone/modem connection. RS-232-C cable has a nominal limit of 50 feet and, even the extended-distance cable is rarely usable beyond 250 feet. While other interface methods such as RS-449 support greater distances, these may not be offered either at the PC or on the data center equipment. Multiplexers and concentrators can solve some of these problems and reduce the wiring load. Use of existing wiring can save money, but walls and ceilings filled with unidentified cables can also be difficult to sort out. Installing a cabling system, with coaxial or optical fiber backbone cables between floors, may therefore be cost effective.

#### Speed and CPU Port Capacity

The most popular PC interfaces are 1200 bps asynchronous interfaces, but there are a growing number of higher speed synchronous interfaces as well. Since front-end processors handle a variety of data rates, it is rarely a problem to interface a line of the right speed to the mainframe. Since many PCs are occasional mainframe users, however, it may be quite cost effective to investigate multiplexers, concentrators, and/or departmental systems that allow a higher speed line on the mainframe. This would be particularly true if the number of spare hardware ports on the mainframe is limited, or if adding new ports requires expensive configuration expansions.

#### Terminal Emulation versus Integrated Links

As we have seen, the networks on the mainframe host are often so large and complex that changes must be carefully architected. This network characteristic provides the first limit to the choices available to the PC. Most users find the easiest way to connect their PCs to the data center is to have the PC emulate the dumb terminal that is already supported by the application(s) they want to access and, hence, is already integrated into the data center network. This seems straightforward enough, but, in reality, there are a number of qualifications that can have repercussions if the user wants to expand the system's capabilities beyond the basic query and data entry functions of dumb terminals or if more than one computer in incompatible networks must be accessed.

The evolution of the PC link beyond mere terminal emulation has proceeded cautiously. The first step was extending emulation so that entire files could be downloaded, instead of having them printed out and then rekeyed into the PC if the PC user wanted to work on them locally. This did not have a direct effect on the mainframe, but certain activities, such as uploading files, generating queries that extract only portions of a file to be downloaded, storing "virtual diskettes" (files formatted in PC formats on the host), and reformatting files so that they can be directly loaded into PC applications, move the PC more and more toward integrated cooperative processing with the host.

#### Terminal Emulation

The key questions for basic applications are the type of terminal to emulate and the configuration of the connection. IBM's System/3X and AS/400, which compete with Digital for the departmental computer market, attach the 5250 synchronous terminal series. Digital systems operate best with terminals emulating the asynchronous VT52 or VT100 terminals or their descendants. IBM S/370-compatible mainframes and superminis, which encompass three quarters of the large mainframe market, are almost always served by the interactive display terminals of the synchronous 3270 family. In a few cases it is important not to overlook the "almost," because IBM mainframes (including the 4300 and 9370 supermini versions) can operate under several different operating systems, and not everyone participates fully in SNA. For example, IBM's IX/370 UNIX system, even though it runs on a mainframe under VM, supports only asynchronous ASCII terminals such as the 3101 and, thus, must be connected to an auxiliary Series/1 as a networking front end. Some IBM users may also find products that emulate the batch workstations, such as the 2780 and 3780, suit their operating environment, but such devices are less common than the interactive terminals.



Once the type of terminal has been defined, or a range of acceptable devices has been identified, the issue of connection mode can be considered. A surprising variety of choices exists. Using as an example the emulation of a 3270 clustered system (3274 communications controller attaching up to thirty-two 3278 local displays) on a standard IBM PC, we find that the user would adopt one of six basic attachment methods.

**Terminal Emulation Only:** Emulate only the 3278/3279 and attach the PC to the 3274 as if it were one of the displays. This involves adding a board and any necessary software to the PC, cabling the board to the 3274 with a coaxial cable, and reconfiguring the 37XX front end and 3274 addressing and control microcode (as is usual when a terminal is added). This is one of the simplest methods because it essentially consists of cabling the PC into an existing communications system. The method is also less expensive, since there is no modem, and a higher speed communications line is usually shared by a number of terminals.

**Terminal and Controller Emulation:** Emulate both the 3278/3279 and the 3274 with a board that emulates the combined logic of both the terminal and controller. In its most common and basic form, this method involves adding a board and its associated software to the PC, cabling the board to a synchronous modem, interfacing the modem to the appropriate data communications lines, and adjusting the software on the remote host so that it recognizes that a new remote 3270 system has been added to the network.

**Attach to Front-End Processor:** Emulate both the 3278 and the 3274 as in the second method, but this time attach the PC to the front end on a host at the same site, so there is no transmission over a communications line. Again, the host must be adjusted to recognize that a new 3270 system has been added to the network. Since the 3270 board uses BSC or SDLC protocols that require a clock (located in the modem) for synchronization, locally attached systems must be cabled to either a modem eliminator cable (for short distances) or a limited-distance/short-haul modem (for longer distances, between floors, etc.). Existing telephone wiring can handle everything, but some users prefer coaxial cable or optical fiber between floors in a large building for performance reasons.

**Interfacing Programs:** Emulate a totally different terminal (an asynchronous terminal, for example) and allow all adjustments of code, protocol, syntax, etc. to be made by a special set of interfacing programs on both the PC and host.

**Departmental System:** Have the PC attached to a local "departmental computer" that is, in turn, attached to the front end. In this case the departmental system acts both as a file server and communications processor for local PCs. The PCs use emulation boards consistent with the departmental system (asynchronous VT100 for VAX, or synchronous 5250 for IBM System/36, for example), and then the departmental system takes care of communicating compatibly with the mainframe. Departmental systems range widely in size and support complex wide area networks (WANs) and local area networks (LANs) in their own right, so they are not discussed at length in this report.

**Attach to LAN:** The PC may be attached to a LAN, and the LAN can in turn be attached to the mainframe (or to a departmental system). There are a number of LANs on the market, including Digital's Ethernet and IBM's Token-Ring. Ethernet users may find that attachment to the Digital host is an attractive option, providing that DECnet support is available in the data center. In cases where Ethernet is already used or where a large number of PCs will be emulating Digital terminals, Ethernet may be an economical connection option. Several vendors supply PC links to Ethernet, and these products support high data rates. Since Ethernet attachment is not as common as the more traditional RS-232-C or modem attachment, product selection based on features will be restricted.

#### File Upload/Download, Virtual Diskettes, and Other Extensions

The first IBM PC was delivered in 1981; the first terminal emulation board, in 1982; and the first file transfer to the PC product using interactive terminal emulation, in 1983. The market for interactive terminal emulation blossomed immediately, and its primary limitations (from the PC user's point of view) were felt at once. The user not only wanted to query a file and view the data but also needed to retain whole files and manipulate them at the PC site.

This was a problem for systems using 3270 or ASCII terminal emulation, because these terminals were screen oriented; data was transferred in screen-sized blocks, usually 1,920 characters at a time. It was a step forward just to initiate a program allowing the host to download a fill-in-the-blank screen format that could be stored locally and receive back only the new entry records. The 3270 procedure was to download the screen format again for every new screen to be filled out and then strip it of incoming records—thus nearly doubling the total traffic, as well as the size of upload record.

The answer for 3270 emulators was to provide a set of packages that divided files into 1,920-character segments, allowing the file to be downloaded to the PC or uploaded to the host. Very few "pure" emulators presently exist; most also have file download capabilities as well.

It shortly became obvious that simply dumping an enormous mainframe file into a PC created a lot of work, although it was still better than rekeying from a printout. The mainframe file needed to be structured, edited, and reformatted before the PC could use it, and if the data were to be used by two different programs (Lotus 1-2-3 and dBASE, for example), more than data format was needed. In some cases, users wanted to upload files as well as download them. Flexibility was needed in interfacing to different mainframe file formats and different databases, for automatic generation of JCL code and queries in the right format, and so on. One popular approach pioneered by Micro-Tempus' Tempus-Link was to create separate file extracts in PC formats (virtual diskettes) for PC users during the overnight batch updates.

As the market continued to evolve, a variety of features were offered to users in independent packages, in various combinations. Some merely expanded communications flexibility, but many added to processing flexibility as well. The following are some of the most important broad categories.

**Data/File Manipulation:** Included in this category are data search and extraction with conditional retrieval, data compression, data format restructuring, and automatic query generation.

**Virtual Diskettes:** In addition to providing easy-access data extracts, as previously mentioned, virtual diskettes allow the PC to use the mainframe as secondary storage, as backup, and as a file server by uploading files in PC formats instead of translating them into mainframe formats. Sometimes provision is made for mainframe access of the data as well. This capability is frequently combined with transparent file sharing, with the mainframe acting as a file server for the PCs.

**Communications Extensions:** Features in this category include auto dial, auto logon, unattended operations, transparent access to network resources, data compression, multiple sessions, emulation of a variety of terminals on one board, remote operation of other PCs, and remote troubleshooting.

**Ease of Interface** includes menu-driven interfaces, on-line help, macros, and prompts.

**Script Processing and Security Features** includes single/multilevel passwords, transmission encryption, and script languages for programmability.

### Integrated Processing

Integrated processing can be viewed from two perspectives: the recognition and use of the PC's unique characteristics and the integration of PC-link software with host software so that they operate as more of a unit.

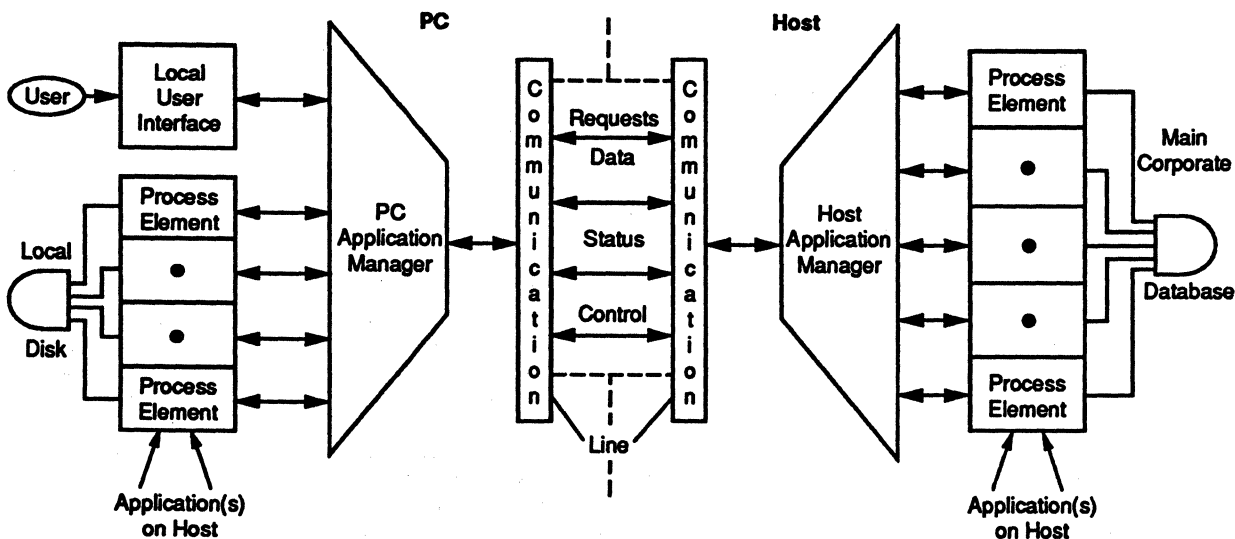
Some of the impediments to more integrated processing come from the PC's very nature. PC systems that support a continuously active communications function and "coresident" communications programs may permit the operator to switch keying and display between local programs and emulation of a data center terminal. This allows the operator to interlace local and remote functions, but only by acting as a common element in two dialogs—one with the local application on the PC and the other with the data center application. Software can be added to the PC to provide these capabilities, but the 640K-byte memory ceiling can present a problem.

In order for the PC's processing and disk storage capabilities to be recognized directly by the mainframe, terminal emulation must be abandoned because interactive terminals typically have printers but not disks or user-programmable controllers. Instead, the PC must be addressed as a processor or a node. This is what makes the advent of LU6.2 protocols for Type 2.1 units so exciting to IBM SNA network users, because the Type 2.1 specification is for a programmable terminal that can participate in a peer-to-peer network.

Packages for integrated processing vary according to the degree to which the PC and host software are linked. In a tightly integrated package, the PC user makes requests without knowing where the data or resources needed to carry them out actually reside. This type of application limits the ability of the PC user to operate with nonintegrated software and prevents the integrated files from being used with any other software package. Loosely integrated applications generally provide only a facility to download data from the host, translate it into a usable format for local processing with many different software packages, and upload it to the host to replace or update the database. Highly integrated packages are easier for users to handle because they present a consistent, request-oriented operating environment; however, such packages are almost always functionally limited. Choosing a level of integration involves measuring the degree to which the package corresponds with the ideal model of user interaction developed in the previous phase. This must then be balanced against any loss of functionality or compatibility with existing software that is associated with the package. Figures 2 and 3 illustrate these differences.

While the processing elements of the applications in the host and in the PC may be generally similar between the two architectures, those of the integrated application must be designed to operate on requests that are generated internally by the software and not through the action of a terminal operator. In other words, they are automatic, not manual. Each element in an integrated application must be modular in structure because it may perform only a small part of a total task that may span several structures.

Figure 2.  
A Tightly Coupled Integrated Application



In general, the systems with the widest range of compatible integrated applications are those where the PC vendor provides proper technical facilities for mainframe communications. Software vendors, given a consistent and preferred hardware environment, tend to build to it. The IBM host communications support, for example, is the preferred attachment method for many of the PC's third-party integrated application software. Vendors also agree on complementary packages.

Data validation and control in an integrated environment may be totally within the package's capability, making data handling at the PC as safe as or safer than the same process at a mainframe terminal. The potential for this level of validation exists with integrated packages, but the promise may not be realized in practice. As the level of integration decreases, from the ideal tight coupling shown in Figure 2 to the looser structure shown in Figure 3, so does the capability of the application to manage the data flow and to ensure that proper validation rules are applied at each update.

One area of concern is the ability to perform local data manipulations on data resident in the PC, either as a normal practice for loosely integrated applications or in tightly integrated applications when the communications link is down. This may be a useful feature from the standpoint of immediate production requirements, but its availability opens the door to divergence of the local and central databases. Tightly integrated applications that provide almost transparent host access and full distribution of files and computer resources are likely to be restricted to a single application, but they offer the highest level of control over the information resources of the company. Loose structures may be applied to more PC and mainframe software so the processing flexibility provided is high, but the potential for data corruption through poor control on the PC is significantly greater. Users must evaluate the benefit of wide application flexibility against the potential loss of control.

## Operational Issues in PC-to-Host Links

### User Interface

The user interface, which comprises the instructions and procedures that a user is expected to understand in order to use a piece of software, is notorious for its lack of standardization in the PC world. As previously mentioned, vendors have at least attempted to make these various interfaces user friendly, usually by means of menu-driven interactions.

A PC acting as a mainframe terminal, however, is normally governed by logic that emulates the functioning of the terminal's less friendly user interface, as well as its technical characteristics. This may involve more than one set of operating rules that must be memorized, if there are no menus or prompts as guides, and each set of rules can differ if more than one application or more than one mainframe is accessed. A New York bank PC operator, switching between 3270 emulation to an IBM host and a local spreadsheet package, was heard to remark, on sitting down at the computer after a break, "Let's see . . . who am I now?"

### Local Data Integrity Problems

First-generation PCs are single-user systems and can function quite readily with simple file handling and security procedures. This simplicity is part of the reason for their cost-effectiveness. When two or more users of separate

freestanding systems want to work on the same file, or if multiple copies of the same file are to be distributed around a network, there can be significant difficulties in maintaining file integrity.

User departments and workgroups normally lack the DP department's editing and control expertise. Therefore, they may be unaware of the proper procedures to ensure data reliability and consistency. Their primary orientation is toward the end-user task, not toward data processing techniques. It may be perfectly obvious to local department supervisors that an expense report date cannot be in the future, but it may not occur to them to incorporate the necessary coding into the programs to properly handle this data in the file. As a result, reports that display expense data in current-month form may fail to show the item at all, and such an item is very likely to be in the wrong place when it eventually does appear.

Lack of a properly organized central collection point for data used by a number of different freestanding systems results in multiple "versions" of information in the PC files of several departments, or even in several systems in the same department. This is no problem with what-if or other isolated applications, but it causes complications if a merged update is needed. For example, it can be disastrous if attempts are made to collect and use the information to update mainframe files at a later date.

### PC Links and Mainframe Data Integrity

At first it would seem that linking PCs to the mainframe (or to a departmental system) would solve this problem, since all of the PCs would be working on centralized files as if they were the mainframe's own terminals. Since PCs are distributed systems and are rarely scheduled and controlled in the same way that the mainframe's own terminals are, however, they do not work the same way. In fact, the most significant problems in the use of terminal emulators may be in the area of data control, at least initially.

When a PC captures a segment of a mainframe database and moves it to local storage, that data becomes subject to manipulation outside the application framework intended to validate it. In many cases, the movement of data to a PC—its editing and return to the host—will bypass host editing or validation functions. Even if the data is sent to the PC by an application program and returned by that same program, there is a good chance that some of the changes made while the data was "remote" will not be fully edited on its return. The host database is thus gradually computed.

The problem is magnified if several PCs access the same information. A major insurance company had an application where file segments were sent to PCs via a batch terminal emulator and processed locally. There was no protection against several PC users concurrently calling for the *same file segment*, and when the data was returned to the file later, each user's version wrote over the previous version.

### Data Integrity in Distributed Databases

Ideally, corporations would like PCs and data center equipment to relate to each other in a way that would make the exact location of the data and the processing power being harnessed for any particular user task to be transparent to the user. This implies distributed databases, distribution of function, and cataloging of locations. The key element, the catalog, exists on all processors in the network in some schemes, and only on some processors (departmental systems, mainframes) in other schemes. Either

method might be workable on small networks, but current catalog types would grow to an unmanageable size on large networks, resulting in performance problems and enormous overhead.

Data integrity problems can be very complex when databases are distributed around different parts of a network, especially when several transactions require that data be transferred between several computers. This requirement has also slowed the development of this type of database, which is still in its infancy.

A commonly cited example of the potential data integrity problems that everyone can relate to is that of a user who has bank accounts in several cities. The user might want to use a terminal in Philadelphia to transfer funds from a Boston account to a New York account and then authorize multiple payments from the New York account to designated companies. It would cause a data integrity problem if the networked system debited the Boston account but failed to credit the New York account. The second level of transactions, which might appear to have been authorized against insufficient funds, compounds the complication.

To carry the example further into a multiuser situation, if this were a joint account the user shared with a spouse, the other person could initiate the same set of transactions. This time the transfer could be performed successfully, but if the amount transferred is more than enough to cover the first set of transactions but not enough to pay double for all the transactions, the result now could be two debits against the Boston account, one credit to the New York account, some payments made twice, some payments made once, and some payments authorized against insufficient funds. If the data integrity error had not occurred, the mistake would have been easy to recognize, but the result of the error is quite confusing.

### Reducing the Risks

Some steps can be taken to reduce the risk of corrupting data center files through the use of PCs emulating terminals.

- Limit "out of data center" updates. If PC copies of files are used only for read access, there is no danger of poor edit control producing bad data.
- Review update applications to ensure that the data quality measures taken by the data center system are also applied to the PC.
- Provide a "staging" area where PC files that have been updated are returned for review and validation, prior to being accepted into the main database.
- Initiate a form of access control on data that can be loaded into PCs to prevent multiple users from requesting the same data elements at the same time.

### Products

PC-to-host products comprise software packages or hardware/software combinations that offer, at the very least, terminal emulation and file transfer. For example, simply by emulating a 3270 terminal (the terminal most used in accessing IBM mainframe computers), a PC can communicate with a mainframe. Other functions, of course, are usually added, and PC-to-host products often support features such as file translation and conversion, unattended operation, data compression, virtual disks, and Applications Program Interfaces (APIs). Connections to the host are most often established in three ways: directly from the

PC via coaxial or twinaxial cable, remotely via a modem, or through a gateway connection to a LAN. Direct, coax attachment of the PC to the host is the oldest and remains the most common method in place, although remote and LAN connections are now preferred.

*Add-In Boards:* DCA's IRMA board was the first add-in board to offer a coaxial connection to IBM mainframes, and it quickly established itself as the product against which all emulation boards would be measured. The boards plug into expansion slots in the microcomputer and provide a coaxial (and in some cases a twinaxial) connection to controllers, mainframes, or minicomputers. This segment of the market has shown limited growth recently, although DCA has begun marketing the IRMA 3 Convertible emulation board—the most recent version of its original IRMA product—which is the only board that can be configured by the user to function with both the traditional Personal Computer Architecture (PCA), as well as the newer Micro Channel Architecture (MCA).

*Software:* In their most basic form, PC-to-host software packages will provide terminal emulation and file transfer capabilities (though most go considerably beyond this point). Generally, the software is used in conjunction with an emulation board, as with the IRMA products or Novell's NetWare 5250 line. One development enjoying increased popularity is that of software-only links such as PACKET/PC's line of software-based products. These links consist of microcomputer and host software components, thus eliminating the need for emulation boards for each connected PC.

### Selection Guidelines

Implementing a PC-to-mainframe link is difficult because it involves expertise from several areas, overlaps responsibilities from several managers, and implies changes in the corporate data flow. For these reasons, it is almost mandatory that such a link be designed by a committee. If a committee is unworkable for any reason, data center management must still get involved, regardless of how elementary the connection will be. The data center should provide information on the current configuration, hardware, software support, application programs, and security measures.

It is advisable to evaluate an application in terms of both terminal emulation and integrated application. One way to do this is to "walk through" the interactions associated with the application to discover what the PC user will experience. Potential users will most likely need the input of the data center professionals to help with the way in which the terminal emulator will interact with the data center equipment. It is a good idea to do three walk-throughs: one for interactive terminal emulation, one for batch terminal emulation, and one for integrated applications. The purpose is not to be too specific to any package, but to see whether the general interaction is acceptable to both the PC users and the data center.

### Cost Justifying an Acquisition

#### PC Justification

Cost justification can be a single-step process, where both the PC and its communications features are considered as a single package for an application that requires communications. Because of the size of the current installed base,

however, it is more common for communications to be added to an existing system as the second step in a two-step procurement process.

In some types of applications, studies have shown an inexplicable leap in user productivity when a system can provide subsecond terminal response time. The gains are more than those accounted for by the increase in machine speed and have led to conjectures about the nature of attention, memory, and mental functioning in creative design tasks. Subsecond response time is more commonly obtained from dedicated PCs or from terminals attached to local computers than from terminals attached to mainframes. These increases and decreases in productivity should also be factored into cost comparisons for some applications.

Thus, the justification for obtaining PCs for high-level management is rarely based on simple cost savings at a local site or relevant to a particular application. This is not a serious problem because there are fewer PCs involved with high-level managers than there are with middle management, data entry, programming, and other jobs that require frequent use of the PC.

As PC use travels down the corporate hierarchy, cost justification is more and more important; the numbers involved are so much greater. Justification for the PC is likely to be expressed in two basic ways. The PC will be automating a particular, relatively controlled set of applications (sometimes called the mainframe's "hidden" backlog) that were previously done manually and were unlikely to ever have been done on the mainframe. Or, the PC will automate a needed function that would normally have been done on the mainframe, but is amenable to the PC instead. Usually the former is cost justified in terms of local needs and manual processing costs, while the latter is cost justified in terms of a comparison of equipment processing and operating costs.

Office automation, for example, using special-purpose word processors and/or data entry, was already under way before the general-purpose PC began to proliferate in the business environment. It was already becoming clear that secretarial productivity could be increased by providing local screen formatting and data entry functions on terminals attached to either a mainframe or a local special-purpose system (or by using local special-purpose terminals). The flexibility, improved response time, and greater accuracy of an automated system can tip the scales in favor of automation even if costs are only at a break-even point. One of the alternatives was a terminal attached to a mainframe, but since the mainframe software was developing very slowly, this application could be considered part of the mainframe's "hidden backlog" of potential applications that were constantly being set aside as impractical because of the importance of more familiar types of work.

### PCs versus Terminals

The PC has particularly altered the price/performance relationships on which the decisions for clerical automation have been based. When general-purpose PCs with word processing packages are compared with special-purpose word processors, the comparable costs and greater flexibility of the ubiquitous PCs have inclined most users to choose this solution over office systems based on dedicated specialized equipment. Economies of scale have meant that a PC with a word processing package can successfully compete with word processing equipment specifically developed for the office automation market. Packaged PC software of all kinds can be purchased from a

retail store and modified if necessary, rather than developed at considerable time and expense by corporate data processing organizations.

Depending on the terminal's flexibility, the PC configuration being compared, and the manufacturer, the price of PCs based on current microprocessor technologies can be comparable to, or even less expensive than, traditional dumb terminals for data center mainframe applications. This is a constantly shifting relationship due to the effect of price wars in both segments of the marketplace. It is especially true of applications that could be shifted entirely to the PC. When the costs of communicating with and processing on the mainframe are calculated into the price of the dumb terminal, there is no question that the cost of mainframe processing is considerably greater than performing the same task on a PC. The trade-offs are not nearly so clear cut when the terminals are attached to a local departmental mini or supermini, however, because these systems tend to be terminal oriented and cost effective.

Cost justification of communications equipment can bring into play the same range of factors as the justification for the original PCs, but since the local financial outlay for adding communications boards and supporting software is less than the cost of the original system, the hidden financial implications of the addition are more likely to be overlooked. The Pandora's box of linking PCs and mainframes without sufficient controls has already been intimated by the discussion of data integrity. The cost of those controls depends highly on the application. One new user of LANs stated that the first year's increase in local management costs was three times that of the cost of the LAN hardware itself; certainly numbers like these would be likely for new cooperative processing applications where data integrity disciplines need to be established where none existed before.

### Limitations of Justification by Cost Alone

The clarity of locally defined price/performance benefits over previously existing solutions and the "free" power increase can be a deceptive cloud covering the issues of great importance to the competitive/financial performance of the company as a whole. One price for the productivity gains of PCs is often paid in the area of information resource management. A corporation's operating information base is itself a significant asset, and management information systems rely on proper maintenance of the basic operating data to provide summaries of current activities and to project future trends. Freestanding PCs can trap vital information before it reaches the data center, making it unavailable to corporate MIS, upper management, or other areas of the organization that need it to operate more competitively.

On the other hand, local managers frequently find that localizing computer power provides greater potential to increase efficiency, because there is less need to conform to standards that reflect irrelevant applications. Furthermore, they have more power over what is processed when; they can thus fine-tune their own operations better and avoid irritating delays that they were previously powerless to resolve.

Here the question of the relationship of the parts to the whole becomes tricky. A department may maintain its own budget more efficiently on a local PC, but is that gain in efficiency worth the loss of the information to the business as a whole? A PC that is linked to a mainframe so that key operations are performed cooperatively usually implies

mainframe access to the PC data, at least to some extent. This is a key argument that should not be lost when considering price and the probable negative impact of improperly controlled PCs on mainframe data integrity.

### A Checklist of Concerns

As management meets to plan for new PC-to-mainframe links, it is helpful to compile a checklist of questions from several different points of view to ensure nothing has been overlooked. The following core questions provide a springboard for expanded checklists, which should be tailored to the company's individual situation.

#### Basic Choices

A user faced with linking PCs to the data center is confronted by choices in several areas:

*Operating Strategy:* The PC can emulate a terminal to an existing application on the mainframe, with or without support for direct transfer of files. It can also link with the mainframe in a joint application supported by custom software in both places.

*Mode of Attachment:* The PC can connect directly via data cables (usually an RS-232-C interface), through a modem, via coaxial cable to a cluster controller, over a local area network, directly to the computer bus, or in other mutually supportable ways.

*Product Source:* The necessary equipment and software can be acquired from a single source or multiple sources; those sources can be computer vendors, modem vendors, third-party hardware or software suppliers, or system integrators.

*Information Management and Control:* Figure 4 shows some of the forms of PC data interaction that can take place and the impact on data center files.

#### Terminal Emulation

Questions important to consider when selecting any terminal emulator are as follows:

*Emulation and Protocol Support:* Does the product emulate a terminal type and protocol that are already supported at the data center or that can be supported easily?

*Attachment:* Will the method of attaching the product to the data center equipment be satisfactory, given the location of the systems and the types of wiring, etc., which are available?

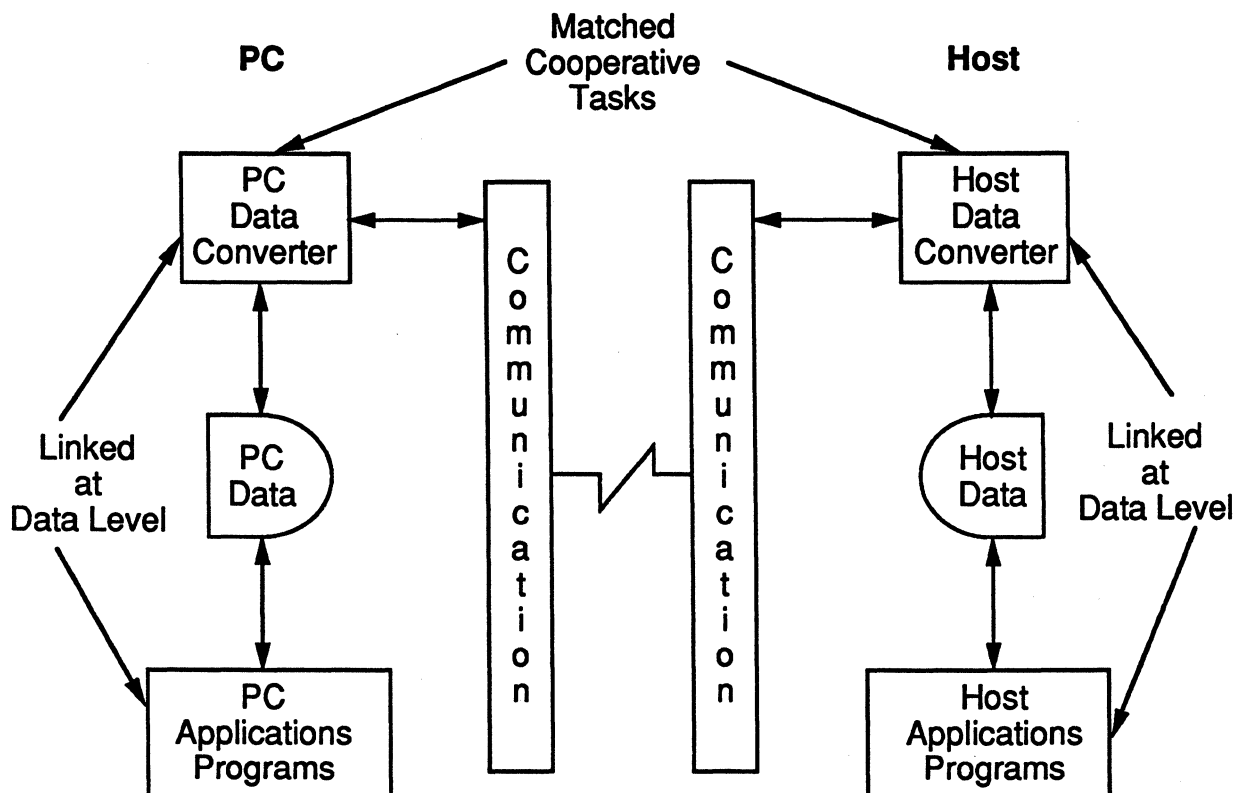
*Hardware/Software Compatibility:* Is the product compatible with any special hardware or software that is already in use on the system?

*Operational Mode:* Is the mode of operation of the product suitable for the current PC operator's level and for the type of work which is expected to be done with the product?

*Operating Procedures:* Are the operating procedures so different from those associated with the PC's local operation that they will be difficult to complete properly?

*Environment:* Can the combined environment created by the product between the PC and the host be controlled

Figure 3.  
A Loosely Coupled Integrated Application



from the viewpoint of access security, data integrity, and information resource management?

*Memory Requirements:* How much memory does the communications board require? The more complex the operation, the greater the amount of memory needed. On first-generation IBM PCs, with their 640K-byte memory ceiling, packages requiring 150K bytes are more likely to prevent a large application from running than one requiring 80K bytes. This problem will be resolved with OS/2 and the new PS/2 systems.

### **Integrated Applications**

Evaluating an integrated application can be approached in phases.

*Phase One:* During the first phase, the user identifies "target applications" for such systems and indicates the host and PC software (if any) which are already in place in the application areas.

*Phase Two:* The second phase requires the construction of an "ideal access" scenario, where the system's operation is described as the user would like to see it work. These phases provide an idea of the current environment for the task, the investment in it, and the way in which the task should operate (economic constraints aside).

*Phase Three:* The third phase of evaluation is the actual investigation of integrated application alternatives. These alternatives may be identified based on competitive survey publications, advertising, or contact with software vendors.

Technical communications support selection with integrated application software is usually a matter of following the supplier's requirements. Most of these packages will not support wide varieties of communications protocols, modems, or custom interface boards, so selection of hardware in advance of selecting the integrated package could create a problem. There is also a danger that two separate integrated applications may require different hardware support. Users should prioritize their needs for such software and watch for incompatibilities in supporting hardware.

### **If Several Products Fill the Bill**

If several products are generally available that meet the needs of the targeted applications, the following guidelines may help narrow the choice:

*Board Functionality:* Select products that pack several facilities onto one board, to conserve slot space inside the system unit and avoid having to add an external box for more slots later. Some vendors supply both asynchronous and synchronous communications on the same board, for example.

*Enhancement:* Select products that enhance the capabilities of the device emulated without requiring host support for those enhancements. Integral printing support for terminals that cannot normally drive a satellite printer is an example of this.

*Modem Interface:* Select communications products with an interface to an external modem over one with similar features but using an integral modem. Modem technology

is changing very rapidly. Furthermore, the modularity gained allows both the PC and the modem to be swapped around as workgroup needs change.

*Expansion:* PC users will want to grow in unexpected directions and are not likely to remain at the level of the target applications. If the PC vendor supplies a package that meets individual needs and price requirements, give it preference over comparable products from third parties. The chances of the vendor's own product tracking future changes in the PC line are better than those of a third-party product.

*Vendor Support:* Avoid products that rely on the combination of third-party software from one vendor and hardware from another third-party vendor.

*Interface Boards:* If a product contains an interface board that plugs into the PC directly, buy from a company that supplies many types of such boards over one that supplies only the board used in the product. Broad familiarity with the PC's bus architecture will translate into fewer compatibility problems with other board-level products.

*Warranties:* Check the warranties; there can be some surprising differences.

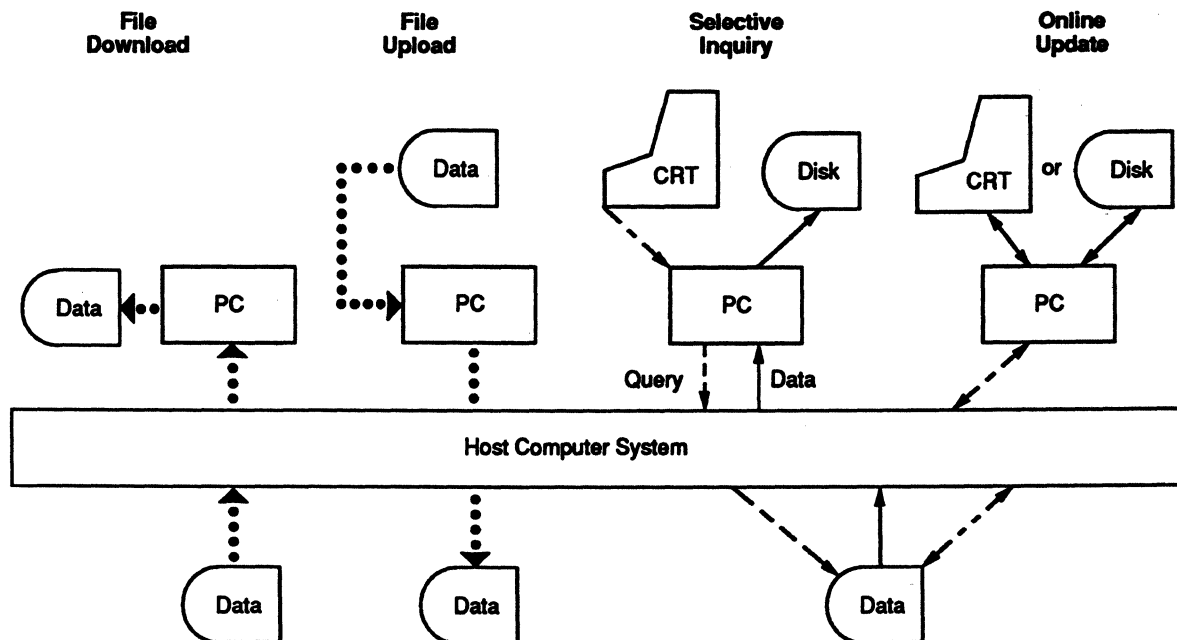
### **Security, Data Integrity, and Management**

Provision of a microcomputer-to-host connection affects security in a number of ways. Most host access techniques use terminal emulation, allowing the user to take advantage of such standard host security as password protection and usage logs. Any security provisions used with terminal access may be used with microcomputer access, provided that only terminal emulation is permitted. Packages that allow unhindered uploading from the microcomputer to any host files will, of course, create genuine hazards, but this is a terminal control problem. The real threats lie in the dramatic proliferation of microcomputers. Terminal proliferation becomes a security problem in situations where limiting physical access to terminals has played a part in the security system. This situation is rather uncommon, but it does happen. The cure is adequate security via logons and passwords, thus placing security at the individual, rather than the terminal, level.

Local microcomputer storage presents a more serious problem. Highly confidential material, such as salary information, new product descriptions, and customer files, may be downloaded by authorized personnel. This information can later be tapped by unauthorized users if it is stored on disk. Disks may be stolen or copied, or the user may fail to end the local application, leaving part of the file resident in RAM. The only solution to this problem is adequate physical security for the microcomputer, with a provision for secure storage of disks. Because this form of security is not at the host level and is not easily implemented in software, many organizations overlook it. With the microcomputer-to-host connection, the risk of microcomputer vulnerability is magnified.

The use of microcomputers, and their growing familiarity within organizations, also poses a problem independent of the microcomputer-to-host connection. Increased computer literacy leads, inevitably, to increased vulnerability of data processing centers. As more people know more

Figure 4.  
File Interactions with Remote PCs



about how to enter a system, the potential for abuse increases. Connections established through telecommunications must be carefully guarded, and any non-terminal connection should be examined to ensure that security measures are adequate.

The following are some questions to help in evaluating the data integrity, security, and information resource management implications of a PC link.

**Access Control:** Are requests for data made from the PC subject to satisfactory levels of access control, such as password control? PC access controls should be at least as strong as those controlling terminal access.

**PC Access:** Can similar control be provided at the PC for operator access to data once the data is downloaded? It is pointless to protect data in the data center and leave it available to all the world once it is loaded onto a PC.

**Outstanding Data:** Can several copies of the data be outstanding in different PCs at the same time? Can this be controlled in any way at the host end? Multiple copies of the same file are invitations to information corruption. If a file cannot be protected from access by several users at once, it may be better to require that the file remain in the data center and be accessed only for use and not for storage at the PC.

**Update Controls:** Can the return of data to the mainframe result in the loss of any update controls? This situation may arise from a lack of control over the update process or from the inability to provide all the transaction editing at

the PC due to lack of other key data files. In any case, relaxing the validation rules may result in a lowering of data quality standards, and this should be specifically understood and approved by management.

**Out-of-File Data:** How long will information be "out of file"—resident in a newer version on the PC than in the main database? Long intervals when the data center does not have current data are highly undesirable, so if data is to be held on the PC for more than a day or so, the application should be reviewed. The data center should check the possible impact of "out-of-file" data on key reports that are run periodically for corporate management or external distribution; conflicts may indicate that PC production schedules will need adjustment for the period near the report cycle.

**New Users:** How is the impact of new PC users on the mainframe to be controlled and budgeted? If, as PC users are added, overall system response time suddenly goes down, how can the data center locate problem areas? How are problems to be controlled? Is time on the mainframe charged to the local department?

**Responsibility:** Who is responsible for what and to whom? If PC access to mainframe files does result in problems with data integrity, security, mainframe performance, etc., who is the person at the remote site responsible for remedying this? Does the reporting structure make it likely that he or she will favor local interests over central mainframe interests? Does that person have enough time to study problems and come up with solutions and enough power to enforce those solutions, which could involve restrictions on local users in middle and upper management? ■



# An Overview of Asynchronous Communications Software

## In this report:

Asynchronous Transmission.....	2
Matching Products with Applications.....	3

## Datapro Summary

Communications software can be used for a variety of applications. Unlike applications programs (e.g., word processing and spreadsheets), communications programs provide tools to accomplish specific tasks. Each communications solution works best only for certain tasks, and users should make their selection based on a careful definition of their needs. Asynchronous communications software for the desktop computer provides basic communications between two PCs.

## Technology Basics

Asynchronous communications involves two computers, two software programs, two modems, and the telephone line between them. Its complexities stem from coordinating this system. To establish a connection to a remote system, the PC requires communications software and a physical connection to the remote system. The physical connection consists of an asynchronous modem sending data over telephone lines. Asynchronous modems can be internal or external; internal modems are add-in boards, while external modems are stand-alone devices cable-connected to the standard RS-232-C serial interface. A null modem cable can connect local systems directly. At its most basic, the communications software accesses the serial port, dials the phone, formats data for transmission, and decodes data received. For communications to occur, certain parameters of the sending and receiving systems must match.

### Physical Interface

The interface is the physical link between data circuit-terminating equipment (DCE)

and data terminal equipment (DTE). Modems are DCEs; microcomputers are DTEs. Most asynchronous communications use the RS-232-C interface. Interface standards define the electrical, mechanical, and physical specifications for connecting data processing equipment. The RS-232-C interface uses a pin-per-function approach to define circuit functions and their corresponding pin assignments. Each pin is assigned a particular task; e.g., RS-232-C pin 8 performs carrier detection. A complete breakdown of pin assignments is as follows:

**Table 1. RS-232-C Pin Assignments**

Pin Number	Description
1	Protective ground
2	Transmitted data
3	Received data
4	Request to send
5	Clear to send
6	Data set ready
7	Signal ground
8	Received line signal detector
9	Reserved for data set testing
10	Reserved for data set testing

—By *Dave Hickey*  
Staff Writer

**Table 1. RS-232-C Pin Assignments  
(Continued)**

Pin Number	Description
11	Unassigned
12	Secondary received line signal detector
13	Secondary clear to send
14	Secondary transmitted data
15	Transmission signal element timing
16	Secondary received data
17	Receiver signal element timing
18	Unassigned
19	Secondary request to send
20	Data terminal ready
21	Signal quality detector
22	Ring indicator
23	Data signal rate selector
24	Transmitted signal element timing
25	Unassigned

The RS-232-C interface is used for both asynchronous and synchronous serial transmission at data rates up to 20K bps, in half- or full-duplex mode. It is also known as the serial port.

Whether internal or external, most asynchronous modems used with PCs are smart modems or Hayes-compatible modems. Smart modems can use commands from the communications software to automatically answer the phone, dial a phone number, and hang up the connection. Hayes-compatible modems, sometimes called AT-compatible modems, use the de facto standard AT command set, developed by Hayes Microcomputer Products, to control modem functions. All modem commands are preceded by AT; for example, the command to dial is ATDT5554444, and the command to hang up is ATH.

### Asynchronous Transmission

Asynchronous transmission refers to data transmission in spurts—delays are possible between bytes of data. The other form of data transmission is synchronous. In synchronous transmission, a group of characters is sent in a continuous bit stream; data transfer is controlled by a timing device, called the clock, initiated at the sending device. Microcomputers normally transmit asynchronously. This works well with irregular data input rates, such as those associated with interactive communications. Synchronous transmission is more efficient, but it is also more expensive because it requires additional hardware, either in the modem or internally installed in the PC. PC-to-PC and

PC-to-Digital Equipment Corp. minicomputer communications is primarily asynchronous. PC-to-IBM mainframe system communications is primarily synchronous.

### Serial Data Transfer

Asynchronous communications uses serial data transfer. Transmission is bit by bit rather than parallel. To transmit serially, parallel codes (data characters) must be converted to a serial bit stream to travel over the communications line. Therefore, the serial bit stream must be broken up into individual characters, from five to eight bits in length. These characters are defined by adding start and stop bits (see Figure 1).

### Data Codes

All data on a computer exists in coded format. A code is a standard format for ordering binary digits (bits), i.e., 1s and 0s, for transmission. The most common codes in use today are the American Standard Code for Information Interchange (ASCII) and the Extended Binary-Coded Decimal Interchange Code (EBCDIC). All microcomputers are ASCII based. ASCII defines 128 characters; codes 32 to 127 are standard printable characters such as ?, A, B, and c. The codes 0 to 31 are characters that control line feed, vertical tab, and backspacing. Extended ASCII includes codes 128 to 256; these are not standardized. A text file uses only printable ASCII codes plus four of the ASCII control codes: carriage return (CR), line feed (LF), form feed (FF), and horizontal tab (HT).

All characters are represented by seven bits, binary 1s and 0s; an eighth bit can be added for parity checking. Most communications programs allow the user to select character length, set stop and start bits, and set parity. These settings must be the same on both the sending and receiving computers. One of the most commonly used settings between PCs is eight bits, one start bit, one stop bit, and no parity. Each online service has a specific required setting.

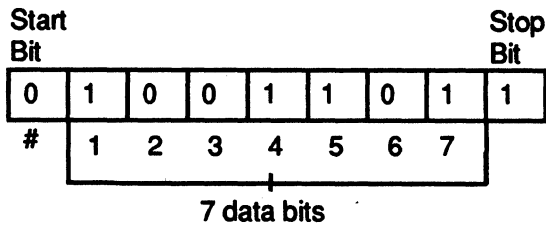
EBCDIC defines 256 characters, each represented by eight bits. All IBM networks use the EBCDIC code, and it is important in PC-to-host communications.

### Error Controls/Protocols

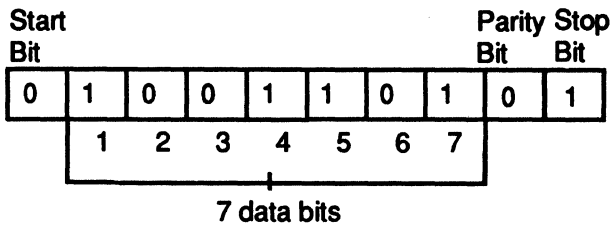
The most critical element in transferring data between two systems is to ensure integrity. Telephone lines are an inefficient medium for transferring digital data. A communications line carrying a serial datastream is subject to interference from storms, cross talk, echoes, and other lines. It can also be subject to delays, especially across satellite networks. This interference destroys data integrity and propagates errors. There is no way to prevent error transmission, especially over long distances.

Protocols, however, do allow error correction. A protocol is a standard set of procedures for establishing and controlling communications. Protocols group data into envelopes, called frames or blocks. These envelopes comprise start and stop bits, error detection sequences, and methods for retransmitting blocks of data containing errors. Transferring text files is possible using only the start and stop bits, because the ASCII codes are transparent and can be passed through the communications system. Exchanging binary files (e.g., some word processing, spreadsheet, or database files), however, requires a protocol because binary files contain nonstandard bit patterns. The same protocol must be implemented at both the sending and receiving computers.

Figure 1.  
The ASCII Character Y



**The ASCII Character Y With Parity Bit Added.**



Note: Parity was set to even. This means that the total number of 1s will be even. The character Y includes 4 1s: therefore, the parity bit is set to 0.

ASCII character codes representing the letter Y, both with and without parity bit added.

Many protocols divide the data to be transmitted into blocks. Each block is then checked for its correct transmission; a positive acknowledgment (ACK) is sent if it is correct, and a negative acknowledgment (NAK) and request for retransmission is sent if it is incorrect. The size of each data block depends on the protocol. Xmodem defines 256K bytes to a block, but some versions allow the block size to be adjusted up to 1,024K bytes. The more bytes per block, the more efficient the file transfer will be. The difference in transfer time can be considerable. For example, take a 5,000K-byte file. At 1 minute per block, 256K blocks will take about 20 minutes; 1,024K blocks will take about 5 minutes. High block sizes work best over relatively clean lines and good connections. The exact method of error detection depends on the protocol being used. Many employ some variation of cyclic redundancy checking (CRC).

Parity is the simplest form of error detection. With parity, a bit is appended to each character, and the receiving system verifies the number of bits in each character according to the parity setting. For example, if odd parity is selected, the parity bit is set to 1 or 0 so that the total number of 1s, excluding the stop and start bits, is always odd. Parity can be set to odd, even, or none. If none is selected, no parity bit is added (see Figure 1). Different online services require different parity settings. The parity setting must be the same at both the receiving and sending computers. Cyclic redundancy checking is a more powerful error-checking technique that employs polynomially generated check characters. With CRC, the entire block of data is considered to be one long binary number divided by another standard binary number. After the quotient is discarded, the remainder is transmitted as one- or two-block check characters that are checked at the other end.

Asynchronous communications packages offer both proprietary and public domain protocols. Since the same protocol must be implemented on the sending and the receiving system, proprietary protocols, such as Communications Research Group's Blast, must be running on both systems. Commonly used access public domain protocols, such as xmodem and Kermit, allow file transfer with a wide variety of remote systems. Xmodem remains the most popular and widely implemented protocol. Kermit, a system developed at Columbia University, is increasingly popular because it runs on microcomputers, minicomputers, and mainframes. X.PC is an asynchronous protocol from Tymnet that supports a maximum of 15 simultaneous sessions.

**Transmission Modes**

There are three transmission modes: simplex, half duplex, and full duplex. Simplex transmission is in one direction only. Half duplex transmits in one direction at a time. Full duplex transmission occurs in both directions at once. Full duplex is the most efficient, since data can be sent in one direction while ACKs and NAKs can be sent in the other. Some protocols are full duplex, e.g., Kermit and X.PC; xmodem is half duplex. Users should be careful to distinguish full- and half-duplex transmission from full- and half-duplex terminal modes. These are also known as local echo. This setting determines whether characters typed at the local keyboard are echoed to the local screen.

**Matching Products with Applications**

The various functions offered by each package determine its suitability for a particular application. For example, the user who wants to exchange files with another PC or access an online service or public BBS needs a basic communications package with file transfer and error-correcting protocols. A basic communications package also should support the industry-standard Hayes AT command set, which controls intelligent modems.

In addition, many users, especially novices, prefer communications packages that are easy to use. Certain features make a package simpler to learn and use. One of the most important of these is a Help facility. Help facilities can be simple (an explanation of keys and commands) or sophisticated (context-sensitive text explanations of program functions). The user interface also affects the package's ease of use. Menus can step the user through system configuration and spell out options such as modem types and speeds; this can be reassuring for users unfamiliar with exact types and names.

Command-driven programs save time for more experienced users. Many packages offer preset logons to online services, a text editor, and printer support, along with user-defined communications parameters that can be customized for specific configurations. Most basic communications packages support telephone directories, and many offer such common terminal emulation as TTY and Digital VT100/200. Some features such as macro support and disk access make communications easier but are unnecessary. Macro support allows the user to put a series of operations into a single key sequence—for example, the logon to CompuServe or a user's account number. Disk access makes it easier to locate files for transfer.

Applications requiring host access must use communications packages that support specific types of terminal emulation. If the application requires file transfer, the

package should offer protocol support. Proprietary protocols are more commonly used with host access packages aimed at specific environments, since the user often controls both ends of the exchange. The public domain Kermit protocol, however, is widely implemented for general host access. Host/remote access functions and unattended execution features, which allow the user to exchange data with the host during off-peak hours and from remote sites, are frequently valued by users.

These packages generally offer a more limited range of user-defined parameters and calling/answering functions that match the targeted host/terminal environment. Packages that concentrate on host access and terminal emulation may not support printer access, editing functions, or telephone number directories.

Packages that allow programmers to integrate communications into an application support sophisticated programming functions and added flexibility. Packages that offer a wide variety of user-defined parameters contain added flexibility for meeting specific requirements. Some packages support script files and offer programmable control over parameters, calls to DOS, and calls to other programs or languages.

Security is a separate issue. Some packages offer single-level password access; others offer multiple-level password access. Some packages implement security only for remote access or unattended execution. Many terminal emulation or host access packages depend on the security provided by the host rather than adding their own. Security remains a side issue in communications; most packages do not offer it. ■

# Document Management Software for Networks

## In this report:

File Naming and Location.....	2
Searching for a Document .....	2
Full-Text Indexing .....	2
Controlling Multiple Versions of Files.....	3
System Security .....	3
Storage and Reporting.....	3
The Programs One-on-One.....	4
Accessing a DMS Within a File .....	4
Criteria for Choice .....	4

## Datapro Summary

Files on a network can run rampant, with the network manager losing track of files within directories and subdirectories. A new genre of software is emerging due to its ability to name, save, search for, and archive data files. Document management systems (DMS) are saving users and managers hours of frustration and unproductive time.

Plowing through an extensive list of directories, subdirectories, and files in search of a single online document can be as daunting as looking for a needle in a haystack. Hours are lost. Frustration mounts. Money is wasted. More document-management troubles brew when someone retrieves an old version of a file, or two people unknowingly edit different versions of the same file. While your goal is to avoid these problems and develop and enforce meaningful file names and directory structures, DOS's limitations make it difficult, sometimes impossible.

A new category of software is evolving to address these issues. File management or document management systems (DMSes) name, save, and archive data files of any type. In addition, through a variety of search capabilities, DMSes help you locate any file on the system. A file-management program expands file-location power not just among subdirectories in one drive but across multiple file servers and even between networks, in effect turning the contents of a hard disk into a fully indexed intellectual database.

This Datapro report is a reprint of "Fields of Files" by Ruth Halpern, pp. 48-62, from *LAN Technology*, Volume 7, Number 9, September 1991. Copyright © 1991 by M&T Publishing, Inc. Reprinted with permission.

Whether your local area network supports a law firm, a bank, or an engineering company, the more files your organization generates per month, the more you need file management software.

This report presents an overview of the key features common to most file management programs. It also provides guidelines on how to decide which file-management program will meet your organization's needs, focusing on three software packages: CMS/Data Corp.'s PC DOCS 3.0, SoftSolutions Technology Corp.'s SoftSolution 2.0, and Interpreter Inc.'s Document Administrator 2.2. Because the field is still so new, the file management programs described in this report are continually adding new features. Both PC DOCS and Document Administrator recently released updates to their products that were not available at the time this report went to press. The core techniques, however, remain the same.

## Basic Features

An efficient file-management program is like an intelligent, highly protective file clerk. Each time you create a file, the DMS insists that you create a document profile for it, so that the DMS can track that document for all time. With a DMS on your network, you are assured of always being able to find any file that was created on the system, regardless of the skill level of the user who created it. Improved document control, however, is accompanied by a loss of flexibility. In order for the DMS to do its

job, users must not perform any standard DOS file-management commands such as copying, renaming, moving, or deleting files, tasks controlled by the DMS.

A document profile is a descriptive cover sheet that is attached to the file. The document profile usually stores the file's author, descriptive name, and type, the customer and project to which it pertains, and other information. The fields in the profile vary from one program to another. Some fields are optional, a few of them require user-supplied information, and the rest are filled in automatically by the DMS.

Filling in the required fields of a document profile is made easier with lookup tables, which list all the possible valid entries for that field. In many DMSes, you can import lookup tables from outside sources; for example, you can import the bindery from your network operating system for the list of users, or import data from a billing program. As an additional shortcut, most lookup tables offer an option to quickly position the cursor on the correct field entry by using the Tab, Enter, or arrow keys. These tools help new users get familiar with the DMS.

One of the most important methods of managing files with a DMS is correctly setting up and using the document- or file-type field. The document-type field indicates whether a file is a memo, a report, a spreadsheet, or a tax form. In addition, document type determines how the file is managed by the DMS: whether it's indexed by profile only or by the file's contents as well; whether it's archived or deleted; and how long it stays on the system before being marked for archival or deletion. Thus, through document type you can define different "shelf lives" for different document types, just like in a library, where books and periodicals are kept on the shelves for different lengths of time. Cover letters could be set for deletion 30 days after they are last retrieved, while proposals could stay on the system for 365 days—or forever. You can develop your own list of file types, determining indexing, archive policy, and shelf life. This list will evolve over time as you become more familiar with your organization's document management requirements.

---

## File Naming and Location

When a file profile is created, the DMS automatically assigns it a unique number (1 to 99,999,999) as its DOS filename, and automatically determines what subdirectory the file will be stored in. This means users don't need any knowledge of DOS file naming and subdirectory conventions. With some DMSes, such as SoftSolution, the system administrator defines whether files are stored by document type, author, or customer and project number; with others, such as Document Administrator, files are always located by "file cabinet" (equivalent to a subdirectory) and author. In any case, once the system is set up, subdirectory management is completely controlled by the file manager software.

Through a feature called automatic location update, a file is automatically moved to a new subdirectory if the contents of the location-determining field are changed. For instance, in a DMS that is set up to organize files by author, when a different author is assigned to an existing profile, the file itself is automatically moved to that author's directory.

If your network has been running without a DMS for a while, you have a significant volume of files that need to be imported into the DMS. Rather than creating profiles one at a time for each existing file, most DMSes have a mass

import feature that lets you generate partially completed profiles listing whatever information is known about the files. For example, if your files are grouped by author, you can import them that way, automatically specifying the contents of the author field for each group of documents. Users can complete the rest of the profile when they are ready to work on an imported document.

Once a profile has been completed, a DMS automatically launches the correct application to create that file. This means that users only need to know which file they want to work on, not which software program created it.

The system administrator can control which users have access to which software by designating security groups. For instance, one group, called WP DEPT, may only have access to Microsoft Word, while another group, called SYS ADMIN, may have access to every program on the network. You can also set a default application for each security group or individual.

An additional application-related feature common to most DMSes is the ability to launch an application with two or more files simultaneously. This is useful for executing a mail merge, retrieving two documents simultaneously for editing, or comparing two versions of a document with a third-party program such as CompuLaw's DocuLiner or Jurisoft's CompareRite.

---

## Searching for a Document

Document profiles are stored in a system similar to a library's card catalog. Just as a card catalog helps you find books by author, title, or subject, a DMS can locate a file by any criteria—author, subject, creation date, type of file, or contents—even if the file itself has been archived off the server to tape or disk. Thus, the profile becomes a permanent "card" in the DMS's database.

With a DMS on your network, you don't need to know a file's subdirectory or its unique DOS filename to find it. For example, in a commercial enterprise, using the file manager's search function to locate a document allows you to specify as little as the author and the type of file (such as a letter, spreadsheet, invoice, or memo), or simply the month it was created and the customer for whom it was done. With these few fields of information, a DMS will locate and display a list of all the files that meet your search criteria, letting you select the one you had in mind. With many DMSes, you can view or preview a file outside the application that created it. It's likely, however, that the previewed document will not have all the formatting that would be visible in the actual application.

---

## Full-Text Indexing

In addition to searching for files by their profiles, many DMSes have full-text indexing. With this powerful feature, files are scanned for all key words (words other than commonly used words such as "the" and "and"). These indexed words are stored in a separate database file. Using a full-text search, you can locate all files that contain, for instance, the words "fraud" and "false representation." Or, with the SoftSolution full-text search screen, you can combine a document profile search with a full-text search, looking for every document by author Tina Evans that contains the words "dog bite" and "fraud." PC DOCS has a proximity-based full text search, so you can specify that the words be within a certain number of lines or paragraphs of each other. Because the full-text index is stored

separately from the file itself, you can use a full-text search to find even those documents that have been archived off the network.

The full-text index of a file ranges between 5% to 15% of the size of the actual file, depending on the DMS. You will need to allocate sufficient space on the server to store this expanding index. Remember: not all file types need to be full-text indexed. To minimize mass-storage overload, use this feature judiciously.

### Controlling Multiple Versions of Files

DMSes have a number of protections against one of the biggest problems of sharing files on a network: two people editing two different copies of a file at the same time. As in many network environments, only one user at a time is allowed to retrieve a file. However, with a DMS, once a file is retrieved, the name, and in some DMSes the extension, of the person using that file is listed when someone else tries to retrieve it. This makes it easy to contact the person using the file and negotiate who should get to work with it first. A file that is already in use can be copied using a DMS; however, a file should only be copied if it is to be revised for a completely different project.

In addition to an ordinary copy of a document, most DMSes offer a feature called "version control," which duplicates the entire file (and its document profile) and stores it under the same file number, with a different extension. Thus, the first version might be 1442.1, the second 1442.2, and so on, with the most current version displayed first. Since they share the same file number, they are clearly identified as related versions. In PC DOCS and Document Administrator, as many as 99 versions and 26 sub-versions of a file can be created, leading to file names such as 1442.89B.

In a further effort to prevent multiple versions of the same document from being edited simultaneously, most DMSes offer a "check out" feature. When a file is checked out, no one else can retrieve or edit that version. The profile can be displayed and the file can be copied, but users will clearly see that they are not working on the original version. Thus, a manager can check a file out to a floppy disk, take it on the road, and be sure that no one else will edit it while he or she is gone. When the manager returns, he or she can check the document back in to the system, and everyone will be able to access it again. You can also check a file out without removing it from the file server as a way of preventing anyone else from modifying it.

### System Security

DMSes address the issue of network security and file-access rights both on the basis of user or security group rights and on a document-by-document basis. PC DOCS incorporates NetWare's security group definitions for each user, while SoftSolution requires creation of a separate system of user groups and rights. Document Administrator assigns file-access rights by giving users access to specific file cabinets. In any case, users' security rights determine what areas of the DMS they can use, as well as what files they can retrieve. A file's creator can also designate a security level for that specific file, so that only designated users or groups can access it. The combination of a user's and a document's security levels determine who can view, retrieve, copy, rename, or delete that file.

To prevent the copy, rename, and delete file-management functions from being done in DOS, some

## For More Information

PC DOCS .....	\$295/server, \$295/workstation	Orem, UT 84058 801-226-6000
CMS/Data Corp.		
124 Marriott Drive		Document
Park Centre, Suite 200		Administrator .....
Tallahassee, FL 32301		server and 10 users,
904-942-DOCS		\$150/each additional user
800-933-DOCS		Interpreter Inc.
		11455 West 48th Ave.
SoftSolution ..	\$1,495/server, \$195/workstation	Wheat Ridge, CO 80033
SoftSolutions Technology		303-431-8991
Corp.		800-232-4687
ParkView Plaza		
625 S. State St.		

DMSes, such as Document Administrator, automatically lock all non-supervisors out of DOS; others allow DOS lockout at the supervisor's discretion. PC DOCS automatically applies NetWare-level security to subdirectories, so that PC DOCS can store files in NetWare-locked directories, and the files remain protected through both PC DOCS and DOS. The security of the DMS database itself (the collection of information on all the files on the system) is protected through a number of reconstruction utilities that rehabilitate damaged data. In the ideal world it wouldn't be necessary to reconstruct the DMS's database, but it is vital to have powerful utilities available to perform reconstruction if necessary.

### Storage and Reporting

DMS programs assist with archival and deletion by identifying files that are ripe for removal from the system and moving them into a designated area on the hard drive. From there, they can be summarized into reports, deleted, or archived using your existing archival system and media.

The profiles and full-text indexes of archived files are kept in the DMS, so any standard search will locate both archived and active files. The profile of an archived file will direct users to the date and media of archival, so that the archived file can be restored onto the system for further use. In general, the profiles and full-text indexes of deleted files are removed from the system. This ensures that the server does not continue to fill up endlessly.

All DMSes can generate various reports on the files and users in the system. For instance, reports on all documents produced by an author can be sorted by customer, by document type, or by creation date. These reports can be printed to the screen, to a file, or directly to the printer. Once a report format has been designed, it can be stored and reused by anyone on the system.

One type of report that many organizations are interested in is a billing report to track time for work on specific document types, or for pages printed on a laser printer. PC DOCS can bill either by time spent in a document (accounting for inactive periods as well) or by number of pages printed. In contrast, SoftSolution only tracks bill-

able hours, and it offers no mechanism for recognizing when the keyboard is not in use.

---

### The Programs One-on-One

In addition to the shared features noted, each DMS has its own unique approach to document management. For example, because SoftSolution can run on NetWare, Banyan Systems Inc.'s VINES, and smaller networks, as well as with DOS, UNIX, and other systems, it is primarily distinguished by the simplicity of its user interface and its ability to work with a broad variety of applications and networks. For ease of use, a single entry screen is used to create, search for, copy, mark, and modify document profiles. Only when users move to advanced document-management operations, such as full-text searching and launching multiple documents, will they encounter additional screens. This simplicity means minimal user training is necessary and quick searches are possible, since the user doesn't have to make decisions about how to locate a document before beginning the search.

PC DOCS offers quite a few high-powered file-location features that make it especially useful for managing large volumes of related files. The key words feature gives users a lookup table of key words, similar to the lookup tables for author and document type, that help categorize documents. A key word need not actually be present in the body of the text; it is selected in the document profile as a way of grouping a file into an information-organization system. Thus, there may be many documents of the "contract" type on the system, but only certain ones would contain the key word "buyer" in the profile. The keyword field can be incorporated into both profile and full-text searches.

PC DOCS generates a document activity log that lists every date and time when a document was retrieved, checked out, or checked in, and by whom. This can be useful for examining a document's history to study work flow and editing responsibility.

Document Administrator, the newest entry into the DMS field (Version 2.2 was released in June 1991), proposes a more strictly limited method of file access through its department-related file cabinets. A user in one department and cabinet can be restricted to search for documents only in that cabinet; in order to search in other cabinets, the system administrator must grant special access rights. This approach to file management is especially effective in organizations where separate departments handle largely unrelated types of documents, or where high levels of security and little exchange are required. Document Administrator's cabinet organization is intended to limit the vulnerability of the profile and full-text databases by distributing them one per cabinet. This technique also speeds up searches by limiting them to smaller subsets of the overall data on the network.

To aid in handling on-going projects, Document Administrator has a "tickler" feature that flags document types for further action. For instance, a sales-letter document type can be defined that automatically notifies the author when a specific number of days has passed and further action is required.

---

### Accessing a DMS Within a File

There is one vital file management feature that is available only in Document Administrator 2.2: the ability to perform a DMS search, while you're in the middle of a file, to locate a second file. For instance, if you are working on a proposal and you want to copy three paragraphs from a

document you wrote last week, you can jump to Document Administrator using a predefined "hot key," locate the second document, and copy it into your local directory. From there, you can return to the application, retrieve the second document, and copy out the relevant paragraphs. This procedure requires familiarity with the file management capabilities of the application (for instance, WordPerfect's List Files), in addition to Document Administrator, since Document Administrator copies the file to the local directory but doesn't automatically retrieve it.

To work with multiple files using SoftSolution, you have to exit the application altogether, use SoftSolution to mark both files, and then relaunch the application with the two documents loaded into two document windows. This process is cumbersome, since it requires interrupting the work flow and exiting the document; however, it does preclude the need to do any file retrieval from within the application.

In PC DOCS, as long as the new WordPerfect document has never been saved, if you press either F5 (List Files) or Shift-F10 (Retrieve), the PC DOCS search function will be activated and you can locate the second document. However, you will have to go through the complicated process of transferring the new document into Doc 2 and loading the existing document into Doc 1 before you can copy the text from one to the other. And then, before saving the new document, you must transfer it back to Doc 1, since PC DOCS 3.0 can't manage documents in Doc 2. Also, if the document has already been saved, the only way to do a second search is to exit the document and initiate the search function.

---

### Criteria for Choice

As you must with all software programs, when choosing a DMS you must face certain tradeoffs over power, speed, flexibility and ease of use. To determine which of these is more important, consider your network and your organization. What is the volume of files on the network? What will it be at its peak? What percent of your files will be full-text indexed? How sophisticated are your users? What kinds of reports will you need to generate? Do you plan to bill customers for file-creation time? This section outlines some of the specific areas to consider when evaluating a DMS.

While a system administrator with a strong technical background may have no trouble figuring out a DMS, it's essential to consider each DMS from your users' perspective. What is the user interface like? Is there an on-screen template displaying the necessary function keys and cursor commands? In PC DOCS and Document Administrator, there is always a key template on screen. In SoftSolution, no function key menu is displayed, but the function keys and cursor commands emulate WordPerfect.

Consider the main screen and submenus. They should be clearly marked, so that users always know exactly where they are and what action they're performing. The PC DOCS interface, which emulates NetWare's format of overlapping windows, can be disorienting in this respect.

Will users perceive the DMS as a time-saver or as an inconvenient intrusion into the file-creation process? Contributing to the answer to this question is whether profiles are created as a front end or a back end. For users accustomed to programs, such as WordPerfect, in which files are not named until after they're created, a back-end profiling system that doesn't appear until a document is typed may seem less intrusive. For users of other systems who have



always had to name a file before typing it, a front-end approach will not come as a great change.

Most DMSes can be set up to operate in either way, but in general they emphasize one approach or the other. Since SoftSolution and Document Administrator have the capability to act as menuing systems, they are necessarily front-end programs, though completion of a document profile can be postponed until after the file is created. PC DOCS works best as a back-end system, after a file is created, since it is so tightly intermeshed with WordPerfect. Moreover, PC DOCS lets users create several files in a row, postponing document-profile creation to a more convenient time; this may be seen as a disadvantage by others on the network who need access to the newly created documents.

All the DMSes discussed in this report offer profile and full-text searches. Still, within these categories there may be different ways to search for a file. What's the minimum amount of information required to perform a search? How easy is it to read, understand, and manipulate the list of located files? Ideally, the DMS search procedure is so intuitive that a novice can locate a file easily without any DOS-level information about it.

### Customization

The document profile can be customized in a variety of ways in different programs. For instance, Document Administrator lets you define the entire profile, including field names, required fields, and number of fields differently for every user on the system. Thus, the word processing staff might be asked to track author and typist, while managers might only need to track author and subject. In SoftSolution and PC DOCS, on the other hand, there are only two customizable fields on the profile, which may be used to designate, say, doctor and patient, or project and responsibility codes.

Although a DMS must have total control over file management once it's installed, you may want more or less control over the form of the installation itself. How much control do you have over where files are stored? Can you control how your hard disk is organized, or is the subdirectory structure dictated by the DMS? With Document Administrator, you must accept their system of cabinets and subdirectories; with SoftSolution, you can store work according to any of the five required fields in the profile, or according to different criteria for each defined entry in each field. (Thus, author Tina Evans might have her work organized in her directory, except when creating letters, which are all stored together in a \LETTERS directory).

### DMS as Menu

Most DMSes today are designed to run and manage files from many different applications. This is vital if you want a DMS that manages every file on the network, which means PC DOCS will not be suitable for you. PC DOCS 3.0 works only with WordPerfect on a NetWare network that is also running WordPerfect Office or Shell. Release 4.0, which was due to be released in July 1991, can manage a broader range of applications but is still limited to run only under NetWare because it is dependent on NetWare's Btrieve program.

A DMS program that can launch multiple applications can actually serve as a DOS menuing system, forcing users to create profiles for every file they create regardless of application and automatically loading the appropriate application for each file. If you are currently using NetWare menus to help users launch applications, you may opt to switch to a DMS start-up screen that shows each user a list of those applications they are permitted to use. This menu

approach forcibly links the function of document management with file creation for every application.

Each DMS program handles application launches differently. SoftSolution uses batch files, macros, and keyboard reprogramming to change the way each application runs when it's loaded from SoftSolution. In contrast, PC DOCS remains RAM-resident and uses a keystroke capture system to intercept users' keystrokes and direct them either to the application or to PC DOCS, as appropriate. Document Administrator can be customized to show users only those applications to which they have access.

### Speed and Volume

How will the speed of your workstations and servers affect the DMS's performance, especially when the number of documents on the system tops 10,000? How much does speed change as more users and files are added to the system? Speed and disk requirements are partly determined by the search algorithm the DMS uses. Some searches work at the same speed regardless of the number of files on the system, while others show a marked degradation as file volume increases. The interaction between the DMS and the network software will determine how the DMS affects the overall load on the file server.

When purchasing programs that provide full-text indexing, evaluate how that indexing is performed. PC DOCS makes it part of the general load on the file server, and indexing runs with no perceptible slowdown. SoftSolution, on the other hand, recommends a dedicated indexer of 386 speed or faster to take the burden of indexing off the server and ensure that files are indexed on the fly, as soon as they are created. This dedicated indexer cannot be used as an ordinary workstation; all of its time is devoted to processing new documents.

Remember, too, that a DMS fits into a specific niche in the network environment. It must interact with DOS, the network operating system and each of the applications whose files it manages, without interfering with any of their requirements. It must also be adaptable to upgrades and new releases with a minimum of readjustment. Therefore, you need to consider the software you already have in place to decide which DMS will mesh best with your system. PC DOCS is closely interdependent with both NetWare and WordPerfect; this makes it easy to access NetWare 3.x's security functions, but raises the issue of how PC DOCS would be affected by upgrades in either of those products.

Because DMS programs interact with so many levels of the system, they need to be installed with great care. To ensure that SoftSolution and PC DOCS are installed correctly, both companies have training programs to teach their resellers correct installation techniques. If the DMS you select must be installed by an authorized reseller, make sure that you have a reseller in your region who is available to service the system during its early phases. If adjustments and modifications are needed to the system, you will want someone close at hand to provide them. Also, consider having the DMS system administrator and at least one other sophisticated user attend the installation training, to bring the DMS installation and maintenance skills in-house.

When you feel overwhelmed by the technical and political issues that installing a DMS will inevitably raise in your organization, remember what you will be avoiding: all those frustrating hours combing through directories, wandering around the network, looking for a file whose cryptic name you might not even recognize. ■



# An Overview of Microcomputer Operating Systems

## In this report:

Products .....	2
Choosing an Operating System .....	6

## Datapro Summary

An operating system provides basic system functions for the microprocessor, the user, the applications software, and the peripherals. Services provided by an operating system range from the simple application loading offered by MS-DOS to the sophisticated messaging and communications facilities supported by UNIX. The choice of an operating system determines the functional limits of the system hardware and the applications which run on that platform. No one right choice will apply to everyone, however. Advantages and disadvantages exist for each of the four major alternatives—MS-DOS, OS/2, UNIX, and the Macintosh. This report discusses the features of each alternative and outlines the factors that should be weighed carefully when choosing an operating system.

## Technology Basics

### New Developments

Typically, system software changes slowly, but the past year has been pivotal for microcomputer operating systems. The introduction of Windows 3.0, further delays in the release of Apple's System 7.0, and questions about DOS, Windows, OS/2, and the relationship between IBM and Microsoft have lent an air of suspense to the operating systems market. Add UNIX, in all its many flavors, and with all its many standards and consortiums, and this year promises to be important for operating system technology.

### The Influence of 80386/80486 Microprocessors

Intel 80386 and 80486 microprocessors in PC-compatible microcomputers and servers have caused many developers and end users to rethink their computer system strategy. This is because an 80386 or 80486 file server in a distributed processing environment can provide the power once only available from a larger system. The 80386/

80486 provides an efficient hardware solution for multitasking and offers a virtual 8086 mode that allows a computer to function as if it were multiple single-tasking MS-DOS systems, and its introduction has created room for innovation in the operating system market.

To take advantage of the features available through the 80386 and 80486 microprocessors, users need software designed to use them, and MS-DOS, in an unmodified form, cannot support them. The 80386/80486, therefore, has sparked renewed interest in microcomputer operating systems. Users have a range of choices for operating systems on the 80386/80486 platform, choices with far reaching implications in today's competitive business environment.

The 80486 microprocessor, however, has not had as much of an impact on operating system technology as the 80386. Because the 80486 offers complete backward compatibility with the 80386 and because its major improvements are in performance (such as on-board 80387 math co-processor and RISC design), we do not consider the 80486 to be as much of an influence on operating system technology at this time.

—By Karen J. Offermann  
Associate Editor

## Products

### MS-DOS

The single-user, single-tasking MS-DOS operating system (also known as PC-DOS) was introduced in 1981 for the IBM Personal Computer. MS-DOS has since been updated several times, and Version 5.0 of the operating system was released in June of this year. (It will also be released by IBM as IBM PC-DOS 5.0.) In a widely recognized binary standard, MS-DOS links closely with the Intel 8086, 80286, 80386, and 80486 microprocessors (a binary standard couples a processor with an operating system at the object code level; MS-DOS applications, for example, can be run on any MS-DOS-compatible computer).

The 5.0 release adds memory management facilities that permit the loading of device drivers, TSRs, and the operating system kernel into high memory, reducing the operating system impact to 15K bytes on an 80286 or higher machine. Version 5.0 is ROM executable, permits 2.88M bytes of diskette formatting, and supports task-switching among active programs.

### Advantages

*Large Base of Users and Applications.* Introduced in 1981 as the operating system of choice for IBM's Personal Computer, MS-DOS has become the microcomputer industry's most popular operating system, with more than 60 million installations.

More specialized software is available for MS-DOS than has been written for any other operating system. Over 35,000 commercial and public domain applications have been released for MS-DOS, ranging from many language compilers and utilities to word processing, spreadsheets, databases, and graphics packages.

*Easily Supported.* MS-DOS can generally be installed by even a novice, and, once installed, rarely requires significant customization.

*Runs on Inexpensive Hardware.* MS-DOS will run on any IBM PC/XT/AT or compatible, giving the user a wide choice of hardware. An initial purchase is inexpensive—for approximately \$1,000, an 8088-based system can be bought with a 40M-byte hard disk and 640K bytes of RAM.

*Windows 3.0 Changes Everything.* Windows 3.0, introduced in 1990, renewed interest in the DOS operating system. Many analysts predicted the demise of DOS after the 1987 introduction of OS/2. Thus far, because few end-user applications are available, many users have found no compelling reason to make the switch to OS/2, and most OS/2 installations are distributed environments where OS/2 is resident on the server.

The features in Windows 3.0 have provided DOS users with a bridge from single-tasking DOS (with its inherent memory restrictions) to a network-compatible, multitasking environment with extended memory support. Windows 3.0 has sold beyond Microsoft's own projections and promises to extend the life span of the vast majority of existing DOS installations for the foreseeable future. In addition, Microsoft has announced plans to develop a 32-bit version of Windows.

## Table 1. OS/2 2.0 and OS/2 3.0 Comparison

OS/2 2.0	OS/2 3.0
Written by IBM, licensed to Microsoft, will be released by both, due for release in late '91	Written by Microsoft, licensed to IBM, will be released by both, due for release in '92-'93 timeframe
Supports 32-bit Intel processors	Supports 32-bit Intel and RISC processors
Runs DOS applications	Runs DOS applications
Runs Windows 2.0 and 3.0 applications unmodified	Runs Windows 2.0 and 3.0 applications unmodified
Runs OS/2 applications	Runs OS/2 applications
	Symmetric multiprocessor support, fault tolerant, new technology kernel
	Code is 99% portable, only processor-specific code is not

### Disadvantages

*Single-tasking, in Real Mode Only.* Since MS-DOS applications write directly to the physical devices in a system, they bypass the operating system. Under MS-DOS, only one application can run at a time, since simultaneous multiple attempts to access devices will result in a system crash.

*Limited to 640K Bytes of RAM.* This constraint is built into the operating system and is not the result of any inherent hardware limitation.

*File System Limitations.* Older versions of MS-DOS limited logical file and partition size to 32M bytes. However, Version 4.01 of MS-DOS allows processing of 32-bit logical sectors, increasing the maximum disk size to 2G bytes. Users who do not upgrade are restricted by the old file system's limitations. 4.01 is backward compatible with files and partitions created under previous versions of MS-DOS.

*Limited Life Span.* The announcement of the OS/2 operating system in April 1987 caused speculation that MS-DOS support and enhancements would be gradually phased out and that IBM and Microsoft would abandon MS-DOS in favor of OS/2. This phase-out has not taken place. DOS 4.01 addresses some basic problems—providing support for Lotus/Intel/Microsoft (LIM) extended memory and files and partitions larger than 32M bytes. However, industry support for 4.01 has not been overwhelming. Some vendors have been slow to provide 4.01 versions, citing incremental performance improvements.

MS-DOS 4.01 allows more functionality than any previous version of MS-DOS and reaffirms Microsoft's commitment to the MS-DOS environment. It does not, however, overcome the restrictions that OS/2 was designed to

avoid. MS-DOS simply does not have the architecture required to access the functions currently available through the 80286, 80386, and 80486 architectures. However, some Windows 3.0 features will likely influence the extension of MS-DOS to a certain extent, to provide support for the protected mode functions of the 80286, 80386, and 80486 microprocessors. It is therefore reasonable to expect that a new implementation of MS-DOS, without Windows 3.0, is a *safe*, but still limited investment.

## OS/2

OS/2 will be radically changed by two announced upgrades—Versions 2.0 and 3.0.

OS/2 2.0, developed by IBM, is due for release early in the fourth quarter this year and is expected to have a sizable impact on the personal computing operating systems market. OS/2 2.0 will fully support the 32-bit Intel 80386 and 80486 processors. It will also be able to run DOS, Windows 2.0, Windows 3.0, OS/2 1.3, and OS/2 2.0 applications simultaneously, overcoming many of the DOS compatibility problems that hampered earlier versions of OS/2.

Version 3.0, a superset of 2.0, will also support the MIPS RISC processor. OS/2 3.0 is fault-tolerant, provides a transaction-based recoverable file system, disk mirroring, and support for symmetric multiprocessors. In addition to running DOS, Windows, and OS/2 applications, it will also run POSIX-compliant UNIX applications.

It is clear that OS/2 will be linked with the Intel 80286 and 80386/80486 and MIPS RISC processors in a binary standard as strongly as MS-DOS, since it provides a natural progression from MS-DOS. IBM and Microsoft's plans for linking future versions of Windows and OS/2 make this inevitable.

### Advantages

**Multitasking.** OS/2 supports preemptive multitasking, which means the operating system shares the CPU among several running applications. Version 3.0 will support symmetric multiprocessing. Symmetric multiprocessing is the division of all processing requirements across all processor resources available.

**Large Addressable Memory.** The memory constraints imposed by the MS-DOS operating system are not an issue under OS/2. The 80286 version of OS/2 can access up to 16M bytes of real RAM and, through swapping, 1G byte of virtual memory. OS/2 2.0 will allow access of 1G byte of real RAM and up to 1T byte of virtual memory on the 80386/80486 processors.

**MS-DOS Compatibility.** Versions 1.2 and 1.3 of the OS/2 operating system have the capability to provide limited MS-DOS processing in a *DOS compatibility box*. The DOS compatibility box provides an emulation of MS-DOS 3.3 in a 640K-byte address space; this allows the user to run only one MS-DOS application in the foreground while OS/2 applications continue processing as background tasks. OS/2 2.0 overcomes this limitation and permits DOS multitasking with greater memory access.

**Interprocess Communications.** The IPC facilities provided by OS/2 allow running programs access to the services of other running programs. Data and commands can be passed from one program to another without concern about file formats or program design. In the simplest

terms, OS/2 applications talk to one another, sharing data and initiating other programs.

The Dynamic Data Exchange (DDE) component of the Presentation Manager allows dynamic links between running programs. In this scenario, a program can broadcast a request for certain information. If another program can provide the requested information, an acknowledgment is returned to the requestor, thus initiating a dialog. Future implementations of DDE will likely provide broadcast services across networks, leading the microcomputer into the realm of distributed processing.

**80386 Processing Support.** Unlike previous releases of OS/2, the 32-bit version, which is currently available as an SDK, provides developers who are accustomed to programming in large system environments with a strong incentive to write applications for the operating system. We expect, therefore, to see Version 2.0 of OS/2 spawn a new class of powerful microcomputer applications, completely unlike DOS software. Microsoft has also stated that it plans to port OS/2 to other microprocessors as the market demands. Current plans are to port OS/2 to the MIPS RISC chip, which will provide significant speed and performance benefits over many currently available processors.

**Power for Developers.** OS/2 2.0 is a powerful development environment. Multitasking has been improved in the 2.0 release, allowing 4,000 threads in the system (up from 512), and unlimited threads per process, as compared to 53 in Version 1.2. And, 80387 floating-point emulation has been built into the system.

### Disadvantages

**Immaturity.** OS/2 is still primarily a developer's, rather than an end user's, platform. However, a new class of OS/2 network-based applications is emerging, acting as servers on a LAN to OS/2- and Windows-based client workstations. Microsoft reports that there are over 130 of these server applications available for OS/2. Distributed processing is expected to be the trump card for the OS/2 operating system, but until these facilities are complete, OS/2 1.3 may be only a marginal improvement over MS-DOS and Windows 3.0. We expect that the release of Version 2.0 (and later 3.0) will provide a tremendous incentive for implementation, with DOS, DOS Windows, and OS/2 support.

**Lack of Available Applications.** About 400 applications are available for OS/2 Presentation Manager. Many are developer's tools, and others are server-based applications. Compared to the wide range of programs available for MS-DOS, and the growing number available for Windows 3.0, OS/2 has not yet become an applications-rich platform.

**Cannot Use the 80386/80486—Yet.** The first implementation of OS/2 was optimized for the 80286 processor. It will run on the 80386/80486, but only by using the chip as a fast 80286. OS/2 Version 2.0 eliminates this problem, but is not due for release until later this year.

**Cost.** Running OS/2 can be expensive. A fast 80386 system with 4M bytes of RAM and a fast hard disk is recommended. Server-based applications require even more hardware.

## UNIX

The premiere UNIX microcomputer operating system is UNIX System V/386, developed by Microsoft in conjunction with AT&T and Interactive Systems, and sold through The Santa Cruz Operation (SCO).

The success of this operating system in the 80386/80486 market has encouraged the emergence of more competitors within the past couple of years. These competitors provide the same microcomputer support as SCO UNIX System V/386, but are less expensive.

The growing interest in UNIX on microcomputers is currently encouraging the development of off-the-shelf UNIX applications. In fact, to stay competitive in the microcomputer market, many UNIX operating systems vendors have identified applications development as a primary development area in making UNIX a successful contender against DOS.

Standards bodies are currently deciding on UNIX standards to make it a truly portable open system environment. Some consistency between the various versions is being achieved through UNIX mergers—System V Release 4.0 incorporated four of the most popular versions: System V, SunOS, Berkeley, and Xenix; OSF/1 is combining the Mach kernel with AIX. In addition, major areas of computer development are underway to address the need for realtime capabilities, greater network support, and more efficient handling of large multiprocessor machines. Clearly, the future of UNIX will depend on technological developments as well as the standards bodies and the cooperation of major UNIX vendors through trade groups such as UNIX International, OSF, and X/Open.

### UNIX System V

UNIX System V was created by AT&T over 20 years ago. The company's latest development efforts resulted in AT&T UNIX System V Release 4 (SVR4) in 1989, a combination of four of the most popular UNIX variations: AT&T UNIX System V, Sun Microsystems SunOS, Berkeley BSD, and Xenix.

UNIX System V Release 4 provides expanded capabilities to its core services in four areas: streams I/O enhancements, file system improvements, third-generation memory management, and realtime processing.

- *Streams*—Streams I/O enhancements such as improved buffer allocation and faster data servicing provide the modularity and portability for all I/O processing done inside the kernel.
- *File System*—In addition to the virtual file system released with SVR4, the operating system supports the Berkeley Fast File Systems, the Remote File System (RFS), the Network File System (NFS) from Sun Microsystems, as well as the PROC, FIFOPS, and SPECFS file systems.
- *Memory Management*—SVR4 uses a third-generation memory management scheme based on the virtual memory system provided in the SunOS operating system.
- *Realtime*—AT&T has added user-controlled process scheduling, higher resolution timers, and enhanced memory-locking.

### UNIX on Workstations

No discussion of UNIX on high-end PCs would be complete without mentioning the competition from engineering workstations offered by vendors such as Sun Microsystems and Hewlett-Packard. Workstations based on the

Motorola 68000 family first became available in the early 1980s, but at that time the high-performance systems were too expensive for all but the most specialized engineering applications. The high-end RISC-based UNIX workstations are still \$25,000 and up, but low-end workstations now start as low as \$5,000, making these systems an attractive alternative to high-end PCs.

### Advantages

*Both Multiuser and Multitasking.* One of UNIX' strongest features is its well-defined multiuser, multitasking capability, where the operating system shares the processor among a number of running tasks and among a number of users. In the past, when CPUs were extremely expensive, processor time slicing among users was perceived as a cost benefit. Current pricing and attitude support the conviction that hardware is cheap, while people are the more valuable resource.

*Development Platform.* The UNIX operating system is written in the C language, allowing source-level compatibility over a variety of microcomputer hardware platforms. This source-level compatibility enhances UNIX as a development platform. UNIX provides the programmer with a mature, full-featured development environment. A wealth of utilities is available for UNIX, allowing the programmer to customize the configuration for optimum performance.

*Large Addressable Memory.* UNIX offers the user a significant gain in addressable memory over the MS-DOS environment, allowing complete access to the facilities provided by the hardware environment.

*Full Support for the 80386 and 80486.* Among the numerous versions of UNIX are several that are specific to the Intel 80386 and 80486 microprocessors. AT&T, SCO, and Interactive Systems all offer full 32-bit implementations of UNIX on an 80386 platform; OS/2, in its current form, is limited to a 16-bit implementation, regardless of processor platform.

*MS-DOS Processing.* Most 80386 UNIX versions support programming extensions allowing MS-DOS processing. These include VP/ix from Interactive Systems Corporation and DOS Merge from Locus Computing. These solutions provide the same access to MS-DOS multitasking as Microsoft Windows, but allow the use of UNIX multiuser facilities at the same time. The current OS/2 release allows only limited MS-DOS processing, running one MS-DOS application at a time.

### Disadvantages

*Applications.* The large base of MS-DOS business applications that are the cornerstone of most microcomputer business environments are simply not available for the UNIX operating system. Although UNIX versions of traditional MS-DOS applications continue to emerge, most businesses that are dependent on off-the-shelf software find that there are not enough application offerings to make a move to UNIX cost-effective.

*Incompatible Versions of UNIX.* There are many different versions of UNIX on the market, including several supporting the Intel architecture. Most of these implementations are currently incompatible with each other. The incompatibilities are known limitations in the UNIX community, however, and are being addressed by standards bodies.

*Initial Hardware Can Be Expensive.* Start-up hardware costs for the Intel-based UNIX environment are expensive—they are roughly comparable to those for OS/2.

### The Macintosh

The Apple Macintosh, based on the Motorola 68000 processor, was a pioneer in implementing the graphical user interface. The Macintosh is a contender in the corporate desktop market, even though it has been slow to penetrate this environment. Apple has responded by aggressively promoting the Macintosh for desktop publishing, targeting government sales, upgrading MultiFinder to allow more than one open application, and introducing A/UX (the Apple version of the UNIX operating system).

The slow decline of MS-DOS, the staggered release schedule for OS/2, and the current lack of a viable UNIX standard have all contributed to a confused, misdirected atmosphere in the microcomputer market. The lack of a clearly superior operating system, and Apple's clear superiority in the user interface area, has left an opportunity for Apple to enter the office automation environment, where once it would have been excluded from consideration. Users may conclude, however, that in some ways, the Macintosh offers no functional gain over MS-DOS systems running Windows 3.0.

A new version of the Macintosh operating system, System 7, which was released in 1991. System 7 will provide enhanced features including virtual memory, 32-bit addressing (up to 4 gigabytes), an InterApplication Communication Architecture (IAC) application-to-application communications framework, outline fonts (developed in cooperation with Microsoft), remote database access (through Apple's Data Access Language), and a new Finder. (See comparison columns for System 7.0 specs.)

### Advantages

*Easy to Use.* The Apple Macintosh was a pioneer in implementing a graphical user interface with a windowing environment, pull-down menus, and a mouse. The Macintosh interface has become a de facto standard, as seen by its replication in Windows 3.0, the Presentation Manager, Hewlett-Packard's NewWave, and Sun Microsystems' Open Look.

*Requires Minimal Training.* The Apple family of computers is remarkably easy for a novice to learn; users can become productive quickly. A Peat Marwick study claims that training users on the Macintosh costs around half what it costs for character-based systems. This, more than anything, explains the overwhelming interest in graphical interfaces.

*High-Quality Graphics.* Even in many predominantly IBM shops, the Macintosh has been the preferred system for graphics and desktop publishing. Apple has built graphics primitives—the building blocks of any graphics application—into the system ROM of the Macintosh;

Windows and other platforms must rely on RAM-based graphics primitives. Its superior graphics capabilities have established the Macintosh as the graphics standard in the microcomputer arena.

### Disadvantages

*Lack of Presence in the Corporate Market.* Many corporate computing departments have their roots in mainframe systems and entered the microcomputer world as a logical outgrowth of their previous implementations of IBM systems. The corporate microcomputing environment is overwhelmingly populated with Intel-based hardware. The connectivity and compatibility considerations that arise when evaluating the Macintosh can make the system less attractive to MIS directors with a history of purchasing from vendors such as IBM and Digital Equipment.

*Lack of True Multitasking.* Finder is a single-tasking environment, and even MultiFinder allows only one active application at a time (it allows a number of applications to be open at once; only the foreground application is given processor time). The MultiFinder feature that Apple likes to call *multitasking* is actually nothing more than rudimentary print spooling, available in the MS-DOS environment for years. Apple's new System 7 will not provide true multitasking either. Another flaw of MultiFinder is that contiguous memory space must be provided for a running application.

*No Built-In MS-DOS Compatibility.* Under MultiFinder, no facilities are provided for processing MS-DOS applications. MS-DOS is specific to the Intel family—the Motorola microprocessors used by the Macintosh cannot support MS-DOS. And, without an MS-DOS emulation program, it is not possible to process MS-DOS applications on the Macintosh platform. This incompatibility is created by the underlying hardware platform as well as by the operating system itself; it virtually eliminates the Macintosh from consideration in an environment where it is necessary to remain backward compatible with an existing base of MS-DOS applications.

*Longevity.* Apple is in a pivotal position. The transition from the current MultiFinder operating system to a true multitasking system will be necessary if the Macintosh is to become a serious contender in the corporate microcomputer environment. Continued delays in the release of System 7.0 have done little to increase corporate confidence in Apple.

The development and implementation of such an operating system can be a costly and time-consuming process, as Microsoft and IBM are finding with OS/2. There have been no announcements to indicate that a true multitasking version of MultiFinder is on its way in the near term. Without a multitasking operating system, however, Apple cannot compete against OS/2 and UNIX in the corporate arena.

*Cost.* Apple's hardware has contributed to its lack of acceptance in a PC-compatible commodities market, where buyers are used to a vast selection of discounted systems. Recently, however, an increasing number of third-party developers are offering Macintosh-compatible products. Even Apple is offering low-cost systems to entice users to purchase the Macintosh.

---

## Choosing an Operating System

1. Define the application need, choose an operating system. The applications that a user wishes to run may well be the most important consideration when choosing an operating system. One method of choosing an application is to determine which manual records must be automated. Accounting (accounts receivable/payable), employee records, and business correspondence are primary targets for office automation.
2. Is there an existing application base to support? The base of existing applications is key when considering a second operating system. Look for systems that permit the user to run multiple DOS applications with access to more than 640K bytes of memory.
3. Let the number of users and method of use determine whether to choose a single-user or multiuser system. If only one employee will use the system, or if several employees will use the system in a serial fashion, then a single-user system can be considered. A multitasking system is the preferred choice. If a number of employees will require simultaneous access to a system, a multiuser or a networked environment should be given strong consideration. In our opinion, single-user, single-tasking operating systems should only be considered for home use. A cautionary note: most users tend to underestimate their system needs and potential usage. It is prudent to approximately double the initial estimates to determine current needs, then double them again to project future processing needs.
4. Consider extending your current operating system. Windows 3.0 from Microsoft and DESQview from Quarterdeck are shells designed to overlay the MS-DOS operating system and provide applications management services in a single-user, multitasking environment. Either of these solutions will provide the user with immediate access to MS-DOS multitasking capabilities.
5. Do you have an existing system? Users with an investment in existing hardware should consider two issues: whether the existing hardware can (or should) be implemented in the new configuration, and what inherent limitations are presented by the existing hardware relative to operating systems support. Users who wish to maintain an existing base of 8086/8088 machines are limited to the MS-DOS operating system. Users with 80386 machines have a broader range of choices, including MS-DOS, OS/2, and UNIX.
6. Have you planned for future hardware acquisitions? Future hardware acquisitions will more than likely be influenced by the operating system already in place. Users planning to make major hardware purchases should at least be aware of the compatibility issues surrounding their potential choices.

---

## Conclusion

Three critical issues confront managers responsible for microcomputer use in their organizations:

- Operating systems must support the business applications that their users require.
- They must do so while keeping costs of development, training, and maintenance as low as possible.
- They must also weave these applications across multiple hardware platforms in distributed computing environments that will become increasingly decentralized in the future.

Each of the four operating systems offers a different response to these issues and is thus better fitted to fill different technical and organizational needs. ■



# Multiprocessing Network Operating Systems

## In this report:

Is MP in Your Future?.....	2
Why More Is Better .....	2
Tale of Two Architectures .....	3
Multiprocessing Glossary.....	3
The Road Ahead.....	5
Product Specifics .....	6

## Datapro Summary

Multiprocessing is an architecture that has been historically used by high-end host operating systems; now, however, the technology has trickled down to the personal computer local area network environment. This report explains why multiprocessing is the next step for PC LANs, and how it works. It provides definitions of high-tech multiprocessing terminology, and focuses on numerous multiprocessing network operating systems that are commercially available.

Although application developers and MIS managers are still battling with client-server issues, this high-end architecture is no longer the state of the art. The network operating system and "superserver" vendors have thrown the next punch: multiprocessing.

Multiprocessing network operating systems (NOSs) and servers tantalize MS managers with promises of greatly improved performance for a relatively small cash outlay. But will the leap to multiprocessing NOSs lead MS over the edge rather than to a higher performance plane?

Multiprocessing is the next step in NOS technology for those users who have made the move to client-server systems, but found that their existing hardware lacks sufficient

horsepower. Multiprocessing allows a NOS and server to support more users, even workers who place a high demand on the system. Multiprocessing will grow in importance as more applications are consolidated onto a single server, making multiprocessing an integral part of distributed computing.

Today, it's a small group of MS managers that can realize the potential of multiprocessing. "Multiprocessing NOSs are a brand-new market," says Tom Wood, industry analyst at Business Research Group (BRG, Newton, Mass.). "Multiprocessing is still embryonic in any kind of implementation."

Today, Banyan (Westboro, Mass.) and Microsoft (Redmond, Wash.) ship multiprocessing versions of their NOSs. The other major NOS players, Novell (Provo, Utah) and IBM (Armonk, N.Y.), as well as Unix Systems Labs (Morristown, N.J.), have announced that multiprocessing is in their futures.

This Datapro report is a reprint of "Make the Leap—Multiprocessing NOSs Are the Next Step," by Patricia Schnaidt, pp. 38-42, from *LAN Magazine*, Volume 6, Number 2, February 1991. Copyright © 1991 by Miller Freeman Publications. Reprinted with permission.

## Is MP in Your Future?

Today, the majority of networks are still used for file and printer sharing. That's not to say that networks aren't becoming an essential part of companies' communication systems. The average number of nodes per network is growing. More sophisticated applications are being developed and installed on production networks. For example, workgroup software, combining the functions of e-mail, conferencing, and scheduling software, is attempting to automate communication among workers.

Multiprocessing NOSs and servers are for those high-end networks. As more demanding applications are installed on networks, existing hardware and software simply won't be able to keep up. Just as the requirements for memory and disk space have risen, so has the need for processing power. Even a 486 may be hard-pressed to keep up with the demands of a distributed computing network. In such situations, the performance of the server's CPU will limit overall productivity. The logical step is to add processors.

Such powerful servers, boasting 486 CPUs, several hundred megabytes of memory and gigabytes of disk storage, have the muscle to serve an enterprisewide network's needs. Products are available from a number of vendors, including Compaq, Tricord, Parallan, and NetFrame.

Servers with multiple CPUs enable the consolidation of services. Instead of having a database server machine, a file server machine, a mainframe gateway machine, and a communication server machine, a superserver has the power to run all four types of software without performance degradation.

"In these cases, multiprocessing servers are the next wave of server technology," says John Dunkle, vice president of WorkGroup Technologies, a market research firm in Hampton, N.H. "In that technology, the multiprocessing NOS is required to enable the multiprocessing architecture."

From the ground, multiprocessing NOSs are simply taking advantage of the multiprocessing hardware. But from the vantage point of the one making the leap, it's a complex step.

## Why More Is Better

Networking is particularly well-suited to multiprocessing, since servers run at the same time as many functionally independent applications, such as file and print service, an SNA gateway, and a database server.

"The typical guy who needs multiprocessing needs a lot of power," says Wood. "The NOS is running a big-ticket, mission-critical application that they probably wrote inhouse. The NOS is running a lot of other things, and there's probably a gateway. They're probably looking to expand the network, and they need the power to handle all that."

Other potential users of multiprocessing NOSs are looking to downsize applications from mainframes or minicomputers. Or they are building a distributed computing network.

Immediate benefits of multiprocessing are the ability to run applications more quickly and support more users. Banyan says multiprocessing has improved its server performance and throughput by 1.5 to 2 times when compared to a single-processor server. Banyan says VINES SMP supports 50 to 100 percent more users. Microsoft says multiprocessing provides a significant performance edge, but has no specific figures.

"Multiprocessing is the most inexpensive way to add capacity. With it, you can have a single-server network that handles more clients. As soon as people are stressing their system, multiprocessing makes a lot of sense," says Drew Freeman, product manager at Microsoft.

Multiprocessing is—and must be—invisible to users. Existing applications run unchanged on a multiprocessing NOS. Just as the operating system and network operating system hide the hardware details from the users, the multiprocessing NOS shoulders the burden of handling multiple CPUs.

Architecting the NOS—rather than the application—to take advantage of the multiprocessing hardware delivers the highest performance gains. Even if the application is unchanged, performance improvement will be seen.

"You don't want the application to have to be rewritten or recompiled. Multiprocessing should be supported by the NOS. Whether you add two or  $n$  processors, the multiprocessing should be hidden," says Dunkle.

Nevertheless, if applications were rewritten to take advantage of multiprocessing, they would run more efficiently than multiprocessing-ignorant applications.

Multiprocessing can accommodate a network's growth. Multiprocessing servers and NOSs provide a "scalable" solution. Users can start with a small system and expand it within the same architecture as demands increase. Instead of buying a PS/2 Model 95 and demoting it to a workstation when you need the power of an AS/400, a scalable solution would enable you to add additional processors to an existing server, as you can with the Compaq Systempro. The multiprocessing NOS then takes advantage of the hardware's power. Your company's investment in equipment is preserved. Additional users and more demanding applications no longer mean more servers.

The consolidation of application servers simplifies network management. Fewer physical servers mean fewer parts to manage, maintain, and troubleshoot. (Of course, when the server fails in this situation, the potential for disaster is greater.)

Multiprocessing does not increase the network manager's responsibilities. Additional processors in the server (or the software to accommodate them) does not add to the network manager's burden. Nor does it increase the risk of downtime.

---

### Tale of Two Architectures

Multiprocessing NOSs provide higher performance simply because they can perform two tasks in parallel. "Ordinary" servers and NOSs are single-processing; one CPU can process one task at a time. For example, while a server is indexing a database, a user might request a file from the disk. The server must temporarily suspend the indexing while it serves the higher-priority I/O request. The indexing resumes immediately afterward. Because this task-switching occurs so rapidly, Network users usually don't realize they are sharing a single CPU. Because of the high demand for the server's CPU, network users may notice slower response time.

A two-CPU server can process two tasks at once. But you cannot assume that two processors will double the throughput. There is no linear growth between the number of processors and performance, because, although multiprocessing systems provide increased computational power, they

---

## Multiprocessing Glossary

### Single-Processor

A computer that provides one CPU to execute applications. The vast majority of systems are single-processing.

### Multiprocessor

A computer that contains two or more CPUs that share common memory and peripherals to execute multiple programs simultaneously. This is the state-of-the-art in PC LANs.

### Asymmetric Multiprocessing

The operating system runs on one processor. Application(s) run on the other processor(s). Microsoft is shipping an asymmetric version of LAN Manager.

### Symmetric Multiprocessing

All processors have the same capabilities; they can all run the same applications and network operating system (NOS) code, access the same memory and I/O devices, and can handle interrupts. All tasks are put into a common queue. A scheduler distributes tasks on a first-come, first-served basis. Banyan is shipping a symmetric multiprocessing version of VINES.

### Multithreading

Threads share the same address space and apparently execute concurrently. This allows faster communication between threads and avoids duplication of data. (A thread is the control unit most basic to CPU utilization.)

Multiprocessing does not imply multithreading. A multiprocessor can execute multiple, unrelated programs simultaneously. Threads may be executed truly simultaneously on a multiprocessor.

### Tightly Coupled

Both processors sit on the same microprocessor bus and access a common memory, enabling them to exchange information very quickly. All processors share access to I/O channels, control units, and devices. The entire system is controlled by one NOS. Compaq's Systempro is tightly coupled.

### Loosely Coupled

The multiple processors have much more autonomy. They do not share the same bus; each is likely to run its own operating system. A local area network is a loosely coupled environment.

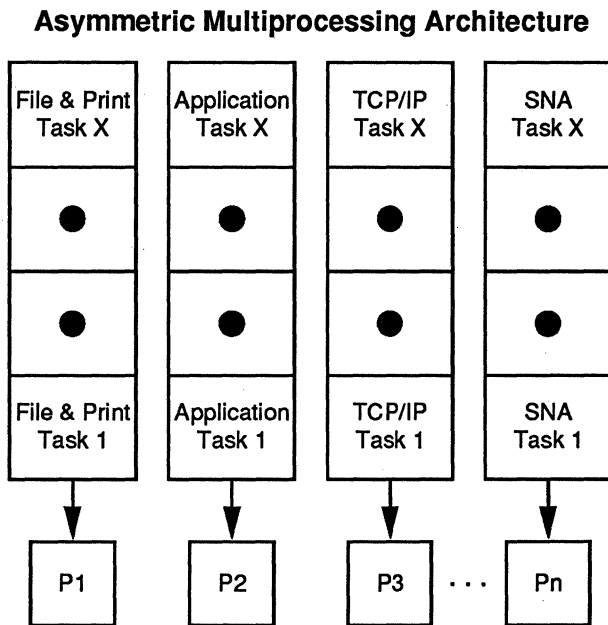
---

do not necessarily provide equal gains in I/O. Contributing to the overhead are interprocess communication, synchronization, and whether the NOS uses a symmetric or asymmetric architecture.

An *asymmetric*, sometimes called *master-slave*, *multiprocessing system*, is the simpler form (see Figure 1). The operating system runs on one processor, which can be considered the master. Application(s) run on the other processor(s). For example, in a dual-processor system, Processor 1 gets all the file tasks and Processor 2 gets all the application tasks.

Asymmetric implementations are best suited for repetitive applications or applications with predictable loads, since processors can be designed specifically for a task. One processor can be de-

Figure 1.  
Asymmetric Multiprocessing Architecture



In an asymmetric multiprocessing system, the operating system runs on one processor. Application(s) run on the other processor(s). For example, in a dual-processor system, Processor 1 gets all the file tasks and Processor 2 gets all the application tasks. Bottlenecks are common in asymmetric systems when operating system usage increases. For example, if the network doesn't have a lot of application traffic but has a great deal of file requests, Processor 1 will have a long list of jobs waiting to be serviced, while Processor 2 will sit idle.

signed for I/O, the other for compute power. However, networks are neither characterized by repetitive tasks nor predictable loads.

Also, bottlenecks are common in asymmetric systems when operating system usage increases. For example, if the network doesn't have a lot of

application traffic but has a great deal of file requests, Processor 1 will have a long list of jobs waiting to be serviced, while Processor 2 sits idle.

Asymmetric systems are less fault tolerant than symmetric multiprocessing systems. If the master processor fails, the entire system fails. Asymmetric systems are suitable for only a small number of processors; they will not support systems with tens of processors, which will be necessary in very large systems.

*Symmetric multiprocessing* is the more sophisticated architecture (see Figure 2). All processors have the same capabilities; they can run the same applications and NOS code, access the same memory and I/O devices, and can handle interrupts.

All tasks are put into a common queue. A scheduler distributes tasks on a first-come, first-served basis as processors become available. When a task is queued, the scheduler checks which processor is available and assigns the task to it. The scheduler looks to the queue again and assigns the next tasks to the next available processor.

For example, as a database is sorted, it requires data from the disk. An interrupt is sent to the CPU to gain access to the disk. With symmetric multiprocessing, that task can be processed in parallel with the current task.

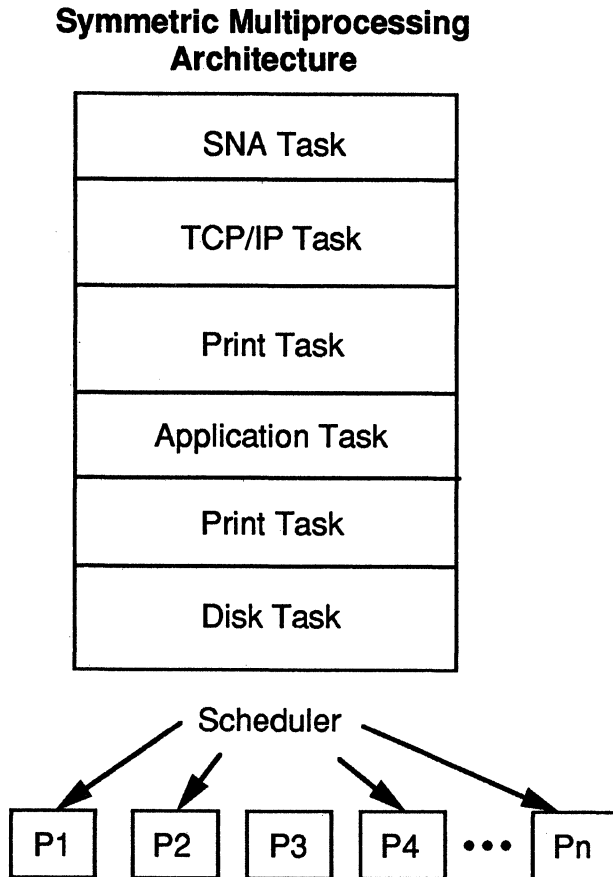
Symmetric multiprocessing provides load-balancing across multiple processors, providing more consistent throughput and scalability than asymmetric multiprocessing.

In its earliest stages, the symmetric multiprocessing's advantage over the asymmetric architecture is slight. But that will change as multiprocessing develops further.

Multiprocessing NOSs must divide network services and applications into parallel streams of code to run concurrently on different processors. The key is maintaining data integrity when multiple processors are simultaneously trying to access the same data or locate unused memory.

Software *locks* are used to achieve this delicate balance of synchronization. Locks are established around critical program code paths. A lock ensures that only one process at a time can execute critical code or make changes. In principle, locks within the operating system execution are similar to record locking at the network level.

Figure 2.  
Symmetric Multiprocessing Architecture



*In symmetric multiprocessing, all processors have the same capabilities; they can all run the same applications and network operating systems code, access the same memory and I/O devices, and can handle interrupts. All tasks are put into a common queue.*

System performance is greatly affected by vendors' locking strategies. The size and number of critical code areas affect performance.

"Granularity is really a fine balancing act," says Dan Rasmussen, the Banyan VINES SMP product marketing manager. "It is very much an implement-test-implement-test cycle to get maximum performance. We spent a lot of time optimizing the granularity."

## Levels of Concurrency

When discussing concurrency of operations in multiprocessors, one must distinguish among the levels at which the concurrency is implemented.

**Job** is the highest level and consists of one or more tasks.

**Task** is a unit of scheduling to be assigned to one or more processors and consists of one or more processes.

**Process** is a collection of program instructions, executed on one processor. It is an indivisible unit with respect to processor allocation.

**Instruction** is a simple unit of execution at the lowest level. (An instruction can also be decomposed into a number of micro-instructions, however.)

—*Multiprocessors*, by Daniel Tabak. Published by Prentice-Hall.

### The Road Ahead

Although existing applications will run faster on multiprocessing NOSs, multiprocessing-aware applications would run even faster. It's like the early days of LANs. You were grateful if the application didn't crash the network. You were ecstatic if the application supported record and file locking.

"As long as the multiprocessing NOS looks like a single-processing NOS, then there's no difference to the application. That's not to say you can't make better use of it," says Business Research Group's Wood.

The road to multiprocessing-aware applications will be long and difficult. Application developers are still reeling from trying to develop client-server applications. Front-end applications that were promised two years ago are only starting to come to market today. Multiprocessing adds another layer of complexity.

To take full advantage of multiple processors, applications must be rewritten to accommodate them. In multiprocessing, all tasks do not have to be executed linearly; some programs have routines that can be executed in parallel.

"ISVs [Independent Software Vendors] are extremely pragmatic. They target the belly of the

market," says Dunkle. Multiprocessing and client-server software developers face many of the same issues: What is the best time schedule to introduce such applications? Which operating system should I develop for? Which network operating system or protocol should I develop for? What are the remote procedure calls that I should use? What are the application programming interfaces that I should write to? None of these issues is resolved today.

Although multiprocessing NOSs per se don't add administrative burden, client-server systems, as a whole, pose new management challenges. A major problem in developing and managing client-server applications is that an upgrade to network operating systems requires an upgrade to the application.

Multiprocessing NOSs may also face competition from minicomputers and mainframes. "Companies that have minis and mainframes are reconverting them to be mainframe servers. It may be cheaper than buying a high-end multiprocessing NOS with an Intel-based server," says Wood. "It may stunt the growth of the multiprocessing NOS. MIS managers' budgets are limited, and they must make better use of what they already bought. Client-server allows you to do that."

Downsizing applications from host computers presents its own set of problems. Currently, it is very difficult to manage and secure data that can be moved from a LAN server to a mainframe and back. Management tools for client-server systems are scarce.

"If I have interoperability between a server and a server, and a mainframe and a server, do I really want that information on the mainframe to be changed [on the server] if I don't control the server environment? No. Today you can boost the server performance to be comparable to a minicomputer, but you have to be sure you have the tools to manage it," says WorkGroup Technologies' Dunkle.

---

## Product Specifics

Introduced in September 1990, Banyan's VINES SMP was the first multiprocessing NOS to ship. VINES SMP was designed for networks with file I/O- and CPU-bound applications on a single server, says Banyan's Rasmussen. "The product was also designed to downsize from minicomputer and mainframes to networks."

VINES SMP currently supports the dual-processor Compaq Systempro. It will be able to support up to eight processors in the future. Although it runs only on the Compaq server today, Banyan officials say it would take minimal effort to port VINES SMP to other platforms. VINES SMP is priced at \$13,995. An upgrade kit to make a single-processor system multiprocessing costs \$7,995.

Microsoft debuted its multiprocessing solution in December 1990, becoming the second vendor to ship a multiprocessing NOS. Because the LAN Manager Multiprocessor Server Pak is built on OS/2, Microsoft could implement only an asymmetrical architecture. Banyan was able to build a symmetrical implementation because VINES is based on Unix.

"You get a pretty significant boost with applications such as SQL," says Microsoft's Freeman.

LAN Manager Multiprocessor runs on a dual-processor Compaq Systempro. One processor delivers network file service; the other runs server applications. For example, you could run LAN Manager file services on one processor and SQL Server on the other. Like VINES SMP, LAN Manager will eventually run on other hardware platforms as well. The Multiprocessor Server Pak, priced at \$2,495 per server, is an add-on to an existing LAN Manager network. A five-user license of LAN Manager costs \$995.

Unix Systems Labs is working on a core version of a symmetric multiprocessing Unix that is slated for delivery in 1992. The Unix System V Roadmap calls for a fully parallelized, secure multiprocessing Unix System V that is compatible with X/Open's Portability Guide, Issue 3, supports tens of processors with user-level multiprocessing access, and offers a full multiprocessing API. To preserve users' investments in application software, multiprocessing Unix must be compatible with early versions of Unix System V. Unix System Labs is shooting for a close to linear improvement in performance as processors are added.

Novell has announced NetWare multiprocessing guidelines for hardware manufacturers and has pledged it will package a version of NetWare 386 with full multiprocessing capabilities "as the application needs of our customers demand them." No dates have been announced.

---

### **The Cutting Edge**

Multiprocessing promises to run applications much faster and support more users for relatively only a little more money. The server companies are investing in multiprocessing. So are operating system companies. The application vendors are looking at multiprocessing; some are acting already. But installing a multiprocessing system today puts you in a very small, elite group just beyond beta tester. Implementing brand new technology may give you a significant edge over your competition. But it's a risk.

The decision to go with multiprocessing (or even client-server computing) or not comes down to cost and need. "A lot of this is a cost-benefit analysis. If MIS can convert the minicomputer [to a LAN server] and still have the same confidence level, then they will do it. But then, it may seem the more evolutionary way than to go out and buy a multiprocessing system," says BRG's Wood. ■





# An Overview of Multiuser DOS Systems

## In this report:

Multiuser DOS Operating Systems.....	2
Multiuser Serial Port Boards.....	4
Multiuser Graphics Adapters.....	5
Turnkey Multiuser Systems.....	6
Vendors.....	6

## Datapro Summary

In what might seem like an indication that we have come full circle, there are some new multiuser systems on the market. But these are not the multiuser systems of the past with 32 terminals attached to a box the size of a refrigerator, running some sort of user-unfriendly proprietary operating system. No, these are multiuser systems for the 1990s—cheap, compact, easy to install, connected by fiber optic cables, DOS based, and hosted by a PC. What? DOS based? PC host? That is right. Granted, the PC host must be a 386 or better, and the DOS user interface runs on top of a powerful multitasking operating system you may never have heard of, but these setups can be faster than most LANs and, perhaps best of all, they can cost as little as half as much.

## Not Just an Entry-Level Solution

Touted by the vendors as an alternative to a LAN, it might seem that these systems are aimed strictly at the entry-level, small business environment, but a closer look will reveal that many offer interconnectivity to local area networks as an option. These systems could be extremely cost-effective if used to supplement an existing LAN in departmental settings where only occasional access to company-wide data is required. Remote sites could realize cost benefits from such an installation as well.

While multiuser systems for microcomputers have been around almost since the beginning—some may recall Digital Research's MP/M, the multiuser version of the early CP/M operating system—the advent of 386-based PCs made multiuser computing on a PC platform a much more serious option. For the first time, a microcomputer was available that offered hardware support for the kinds of operating system features that are a must for true

multitasking, multiuser operation. The 80386's 32-bit bus, its capability to address memory up to 4 gigabytes, its "virtual machine" mode, paging, and virtual memory, were all features that users had come to expect in minicomputer operating systems by the mid-1970s.

Several companies banded together in the summer of 1990 to form the Multiuser DOS Federation. Some of the firms had, for some years, been producing multiuser operating systems that delivered acceptable performance for a small number of users attached to an 80286-based PC. Others manufactured serial cards and graphics adapters that allowed multiple stations to connect to a single PC.

Some of these companies offer full turnkey systems, while others provide only software or hardware components that allow users to build their own systems. One of the advantages of these multiuser systems is that they allow existing standalone PCs to be used as terminals on the multiuser system, as well as supporting inexpensive "dumb" terminals. In particular, they can breathe new life into older, 8088-based machines that might otherwise be gathering dust in a storeroom.

Some of the operating system vendors, most notably Theos Software and Bluebird

—By *John Krick*  
Associate Editor

Systems, provide the kind of software development systems that one might expect to be bundled with a full-size minicomputer. Basic, Pascal, Cobol, and C language packages are available.

A software developer can create a multiuser DOS operating system in several ways:

- An existing multitasking, multiuser operating system not originally intended to emulate DOS can be modified to support DOS emulation, usually through an add-on software module. This is the path taken by Bluebird Systems with its SuperDOS with PC-Connect, and by Theos Software, in its THEOS+DOS product.
- An entirely new multiuser operating system kernel that behaves like DOS and provides DOS emulation at each workstation can be written. Alloy 386/MultiWare, Digital Research DR Multiuser DOS, and S & H Computer Systems' TSX-32 are examples of this approach.
- A standalone multitasking addition to standard MS-DOS can be turned into the basis of a multiuser system. This is how IGC VM/386 MultiUser and StarPath Systems' Vmos/MultiUser were created.

All of these strategies for creating a multiuser DOS system use the features of the 80386 and higher processors in similar ways. The virtual 8086 mode that divides the processor clock among several virtual 8086 systems, and provides each of those systems with their own memory, is fundamental to the operation of these systems. The capabilities of individual systems vary widely, however, especially in the number of users supported. Some systems support as few as 8, others as many as 64. Users can build text-only, monochrome systems with terminals connected to standard serial ports over copper wire, or, under some of these systems, they can opt for VGA graphics systems with each workstation linked to the host by fiber optic connections.

It should also be pointed out that multiuser DOS systems are not the only way to attach multiple users to a 386 or larger Intel-based system. Many vendors, including some of those described here, offer or support UNIX systems that provide color X-terminal operation.

## Multiuser DOS Operating Systems

### Alloy Computer Products, Inc.

Alloy Computer Products' 386/MultiWare allows up to 21 users, including the host, to attach terminals or PCs to a 386 or 386SX computer. 386/MultiWare uses the virtual machine mode of the 386 processor and allows each virtual 8086 created to have up to 4MB dedicated to its own use. Each user may have up to eight DOS tasks running simultaneously. The Alloy multiuser operating system is available in three versions—MW386E, an entry-level version that supports up to 5 users; the standard version, MW386, that supports up to 21 users; and the third, and newest, version, 386/MultiWare EZ. 386/MultiWare EZ supports three users without adding additional hardware to the host PC. The standard COM1 and COM2 ports are used to support the second and third workstations.

386/MultiWare includes LINK-PC, terminal emulation software that allows an existing PC to act as a terminal attached to the host system. Alloy also offers MAC-ATTACH terminal emulation software that allows Apple Macintosh computers to connect to a 386/MultiWare host.

Two important software options set 386/MultiWare apart from the pack in connectivity. Alloy's NetWare Con-

nectivity software module allows a 386/MultiWare system to connect to a node on a Novell NetWare local area network. This means that users with NetWare LANs already installed can economically expand the number of users attached to the LAN and are able to share LAN resources. Only a single Ethernet or Arcnet adapter card in the 386/MultiWare host is required to add the full complement of 21 MultiWare users to the LAN. MultiWare users become fully privileged NetWare users, able to access NetWare's electronic mail and any gateways attached to the network. Up to eight NetWare and DOS tasks can run simultaneously.

The second important connectivity option offered for 386/MultiWare is Sangoma Technologies' ClusterComm MultiWare, which allows users of a MultiWare system to perform 3270 or 5250 terminal emulation for attachment to IBM mainframes and midrange computers. ClusterComm supports both SDLC (for leased-line and dial-up connections) and X.25 packet-switching networks.

Alloy also sells what it refers to as IMPs, Intelligent MultiPort cards, in two-user and eight-user versions. Each IMP card is available in models for the ISA bus and for the Micro Channel Architecture (MCA) bus used in the PS/2 product line. The IMP2 and IMP2/PS two-user cards support monochrome Hercules Graphics or CGA text. The IMP8 and IMP8/PS can attach eight text terminals or four text and four Hercules Graphics. Users with PCs can run in CGA emulation mode with Alloy's Link-PC terminal emulation software.

### Bluebird Systems

Bluebird has offered a non-MS-DOS multiuser system, called SuperDOS, for some time, and has only added DOS connectivity comparatively recently. SuperDOS has had a wide variety of applications written for it, many aimed at vertical markets and sold through VARs that specialize in particular vertical market application categories. SuperDOS on a 286-based IBM PS/2 Model 30 can support 18 users. On the PS/2 Model 80, it allows up to 66 users to share resources. SuperDOS can also run on all types of IBM PC compatibles from 8086- to 80486-based machines.

PC-Connect, based on Microsoft Windows, allows the attachment of an IBM PC or compatible to a SuperDOS host machine over a serial connection. In this configuration, the PC can run as many as six SuperDOS tasks, each in its own window, and as many DOS applications as the PC's memory will allow. While the PC with its DOS applications and the SuperDOS host remain two separate environments, the PC-Connect Windows interface does allow users to cut and paste between the two. Data from a SuperDOS application can be easily transferred to a DOS application in this way, and vice versa.

Bluebird offers a rich set of application development tools including compilers for C, RM-COBOL-85, Data General- and Wang-compatible BASICs, and Pascal. A Basic-to-80x86 assembler cross-compiler is also offered that allows applications developed for SuperDOS to be ported to MS-DOS-based microcomputers.

Bluebird also offers a LAN interconnection option for SuperDOS host systems called SuperLAN. It allows 255 SuperDOS microcomputers to be connected over 2.5M bps Arcnet hardware. Applications, peripherals, and other resources residing on any SuperDOS system can be accessed by users attached to any of the network nodes. Each node can support the full complement of 66 users each, and up to 80 disk drives can be distributed over the network.

**Concurrent Controls, Inc. (CCI)**

Concurrent Controls' 386-DOS is based on Digital Research's DR Multiuser DOS. It can connect up to 67 workstations to an 80386 or 80486 host and uses the virtual machine mode of the processor. Any workstation can run up to 16 programs concurrently. 386-DOS implements disk caching to speed I/O operations. 386-DOS includes Smartscreen, a 386-DOS extension that supports graphics workstations with multiuser graphics display adapters. Running on top of 386-DOS, it allows monochrome Hercules graphics-capable monitors and color EGA monitors to be used as terminals. Up to 256 of these host systems can be connected using CCI-Net, allowing any user on any host to share files, disk storage, and printers attached to any host. Remote workstation support is included with CCI multiuser systems so that users off-site can dial into the host system.

**Digital Research, Inc.**

DR Multiuser DOS replaces DR's earlier Concurrent DOS 386 product and shares features with DR DOS 6.0, DR's standalone operating system. DR Multiuser DOS was designed to be DOS compatible from the ground up, and does not build on an existing multitasking operating system and add DOS support on top. DR Multiuser DOS supports a maximum of 64 users. Three users can be supported just by attaching serial terminals to the COM1 and COM2 ports.

DR Multiuser DOS includes PCTERM terminal emulation software that allows PCs to act as DR Multiuser DOS terminals. DR Multiuser DOS supports color text, CGA graphics, and the Microsoft Windows 3.0 environment. Each user station can have as many as eight DOS sessions running at one time. Users can hot key between DOS sessions, but only the foreground session will accept keyboard input. Applications that do not require keyboard input, such as spreadsheet recalculation, database sort, or file transfer via telecommunications, can run in the background. All sessions can run in graphics mode.

DR Multiuser DOS supports applications that use the Lotus-Intel-Microsoft Expanded Memory Specification (LIM EMS). DR Multiuser DOS allocates conventional memory and LIM EMS memory dynamically. Rather than having a certain amount of memory preallocated to it, a session is not allocated memory until an application is started.

**IGC, Inc.**

IGC takes a unique approach to achieving multiuser connectivity. The company's initial product, VM/386, is an add-on to DOS. VM/386 uses the virtual machine mode of the 80386 microprocessor to create a multitasking environment on top of a standard copy of DOS. VM/386 requires DOS 3.0 or later, allows data sharing among running tasks, provides file locking, and makes use of the DOS command.com file to allow use of traditional DOS commands.

VM/386 MultiUser runs on top of VM/386, creating a multiuser, multitasking operating system that can support up to 32 users. VM/386 allows a user to restart an individual virtual machine without affecting the operation of any of the other users' virtual machines. VM/386 Multiuser supports graphics when used with Advanced Micro Research, SunRiver, and Viewport International graphics adapters.

VM/386 MultiUser Starter allows a single 386 or 386SX computer to host two monochrome text-only terminals as well as supporting the monitor attached to the

host machine. It includes a print spooler, and local printers are supported at the terminals. VM/386 Multiuser requires 4MB for a three-user system.

VM/386 and VM/386 MultiUser allow network access with an add-on option called NetPak. NetPak runs in one of the virtual machines, giving that machine access to the network server, all of the files stored on it, electronic mail, and the network-attached printers and other peripherals. The other virtual machines can access the network's resources through a NetPak utility called the Network Distributor. NetPak requires a network interface card and driver software. VM/386 supports most PC networks, including Novell NetWare.

**S & H Computer Systems, Inc.**

S & H Computer Systems' TSX-32 is a multitasking, multiuser operating system for use on 386- and 486-based computers. TSX-32 executes programs in the 80X86's protected mode using 32-bit addressing and demand paging. TSX-32 also uses a proprietary job scheduling algorithm S & H calls Adaptive Scheduling, which allows flexible response to external factors such as I/O activity and completion, and each individual job's execution time. Jobs can be assigned higher priorities by administrators to override the normal interactive scheduling priorities. Each user connection to the host uses the virtual 8086 mode so that each terminal emulates a dedicated PC. DOS support in TSX-32 is not a program that rides on top of the multitasking OS. Instead, it is an integrated part of the TSX-32 kernel. Each user can have up to 10 programs running at one time.

TSX-TERM is a terminal emulator that allows a PC to connect directly to a TSX-32 system and retain its local processing capability, while acting as a TSX system terminal as well.

MessageNet is electronic mail for the TSX-32 system. It includes browse screens, forwarding and reply functions, reminder messages, distribution lists, and folders for message storage.

TSX-Net is an option that allows Ethernet networking between TSX host systems. A serial networking scheme is a standard feature of TSX-32. The TSX-32 network server software works with either attachment scheme to allow users on one system to access files and resources that reside on a different TSX-32 system.

**The Software Link, Inc. (TSL)**

PC-MOS is a multitasking, multiuser operating system that lets users run multiple DOS tasks simultaneously, switching between programs with a "hot key." PC-MOS is available in 5-, 9-, and 25-user versions. It can address up to 4GB of RAM and can allocate up to 676K to each task, depending on configuration.

PC EmuLink is terminal emulation software that allows a PC to connect to a PC-MOS host. It requires 64K of RAM in the PC and connects either directly to the host with a null modem cable for speeds up to 38.4K bps, or remotely over a dial-up connection. PC EmuLink's 25-line display supports either monochrome or CGA color text and graphics.

PC-MOS Gateway to Novell NetWare allows the PC-MOS multiuser system to communicate with a Novell NetWare LAN. It allows up to 16 users to share a single network adapter card connection, reducing the hardware expense usually involved with adding new users to the LAN.

TSL distributes the MaXtation SH-4 manufactured by Maxpeed Corp. of Foster City, CA. The MaXtation SH-4

is component set that includes a four-port Intelligent Workstation Controller, and a Maxpeed Hercules Workstation Unit that is attached to the controller by an eight-conductor phone cord with RJ-45 connectors. The MaXtation SH-4 Workstation unit includes a mouse interface, a parallel printer port, a beeper, and an activity LED. Up to four MaXtation Intelligent Workstation Controllers can be installed in a PC, to support up to 16 workstations.

#### StarPath Systems, Inc.

StarPath Systems developed the Vmos/3 operating system to provide a DOS-compatible multitasking operating system for PC use. Vmos/3 implements demand paging to allocate RAM to programs only as needed, and virtual paging to allow hard disk storage space to simulate RAM storage. The Lotus-Intel-Microsoft/Expanded Memory Specification (LIM/EMS) is used to provide up to 32MB for each program.

Vmos/MultiUser (Vmos/MU) runs under Vmos/3 and allows as many workstations, each with its own DOS prompt, as the hardware is capable of supporting. Users may run multiple DOS sessions at one time. Individual sessions can be rebooted without affecting any other session. Up to 768KB are available for each session depending on the type of video in use. Vmos/MU requires at least 2MB of memory in the host, and the vendor recommends 4MB for optimum performance.

#### Theos Software Corp.

THEOS+DOS runs on top of THEOS 386, the THEOS multiuser system, which, like Bluebird's SuperDOS, is a proprietary multiuser operating system. THEOS+DOS requires a licensed copy of DOS 3.1 or higher for each DOS user, and network versions of DOS applications are recommended by the manufacturer to ensure proper file sharing and file locking capabilities. THEOS+DOS supports up to 256 simultaneous tasks and up to 30 users. 952K of RAM is allocated to each DOS user from up to 128MB of total memory. THEOS+DOS can support up to 26 hard disks, each with a capacity of up to 4.29GB, for a total storage capacity of up to 111.6GB.

Users of THEOS 386 or THEOS+DOS require another product, THEO+TERM, in order to have multiple sessions at their terminal. THEO+TERM allows up to eight active sessions per terminal that the user can "hot key" between.

ScanTerm is Theos' product that allows existing PCs to act as THEOS+DOS or THEOS 386 terminals, while retaining their standalone computing capability.

THEO+GRAFX provides up to 16 users with VGA, EGA, or CGA graphics. It works with the Theos' TG-4 multiuser graphics adapter and controller box that Theos has OEM'ed from DigiBoard. The TG-4 is similar to the DigiBoard MV/4 described below. THEO+GRAFX also works with the SunRiver Fiber Optic Workstation.

## Multiuser Serial Port Boards

#### Arnet Corp.

Arnet's newest product, the ClusterPort/S, is a workgroup concentrator that can support up to 128 ports. Available in versions for PC AT, Micro Channel, and EISA machines, the ClusterPort/S system includes a single host adapter card; an external power supply; and an RS-422 interface that can attach up to eight 16-port concentrator boxes, for the full complement of 128 ports. The concentrator boxes are available with DB-25 or RJ-45 connectors. A host PC

can be equipped with as many as four ClusterPort/S systems giving it the capability to support up to 512 users.

Arnet's SmartPort Plus line of intelligent serial port boards is available in models with 8, 16, 24, and 32 ports. Based on a 10MHz 80C186 processor, the SmartPort Plus boards feature 64K of RAM onboard. Onboard RAM can be upgraded to a total of 512K. Port expansion kits for each board are available as well, in 8-, 16-, and 24-port increments, so any SmartPort Plus board can be upgraded to the full complement of 32 ports. Arnet also offers the SmartPort Plus Micro Channel for the IBM PS/2 Models 50 and above. The SmartPort Plus Micro Channel is available in 8-, 16-, 24-, and 32-port versions.

Arnet also makes the SmartPort board, originally known, in its eight-port version, as OctaPort. The SmartPort board is available in four- and eight-port versions. It can be attached to terminals or PCs using Arnet's four-line Quadracable or eight-line Octacable or an optional external box which is available with either DB-25 connectors or RJ-45 modular plugs.

MultiPort is Arnet's name for its line of standard serial port expansion cards, available in four- and eight-port models. As with the SmartPort board, the MultiPort can be ordered with the Octacable, or with an RJ-45 modular, or DB-25 equipped connection box. A Micro Channel Architecture model of the MultiPort 8 is also available, with the same three connection options.

Finally, Arnet offers a low-priced, two-port board called TwinPort.

#### Control Corp.

Control offers its Hostess four- and eight-port models for the IBM PC AT bus. They feature one 16450 UART chip per port. Four-port models are field upgradable to eight ports. Hostess/MC four- and eight-port models for the Micro Channel Architecture bus are similar to the AT bus models.

Hostess 550 buffered 4-, 8-, and 16-port models use 16550 UARTs. The four-port model is a half-slot card that can be upgraded to support eight ports with the addition of a plug-in daughter board. The Hostess 550 16-port model is a full-slot card. The Hostess 550/MC for the Micro Channel Architecture also supports 16 devices. Up to four Hostess 550 controllers can be installed in a single system.

Smart Hostess cards feature an onboard 80186 processor and up to 512K of RAM. The Smart Hostess cards are available in four-port and eight-port models, and the four-port version can be upgraded to eight ports with the addition of a plug-in module. All ports on either the four- or eight-port version can support synchronous communications.

The Hostess i Series cards are Control's top-of-the-line models. Available in models for the PC AT bus (Hostess i), the PS/2 Micro Channel Architecture (Hostess i/MC), and the EISA bus (Hostess i/E), these boards use a 10MHz 80286 processor (i/MC), or a 12MHz NEC V53 processor (i, i/E). All three cards are available in 8- and 16-user models, and any 8-user model can be upgraded to 16 users with an add-on card. All of the cards feature 128K of dual-ported RAM. The Hostess i/MC can be upgraded to 512K of onboard RAM, the Hostess i and the Hostess i/E to 2MB.

#### DigiBoard

DigiBoard offers a unique product in its top-of-the-line DigiCHANNEL C/CON-16. The C/CON-16 is a cluster controller capable of supporting up to 128 users. The C/CON-16 consists of two components—the C/X Adapter

card that occupies a single host slot and is available in PC AT, Micro Channel Architecture (MCA), and EISA models; and an external concentrator box to which the terminals are attached. The adapter card is based on a 10MHz 80186 and has 128K of dual-ported RAM onboard. It has two full-duplex RS-422 synchronous ports that provide a data transmission rate of up to 1.2M bps. The concentrator box is also based on an 80186, but this processor is running at 16MHz. The concentrator also has 128K of RAM and sixteen 16C550 UART chips. An LED diagnostic display and a row of LEDs that indicate handshaking and transmission status like those found on a modem are provided. Transmission speeds of up to 38.4K bps are supported.

The DigiCHANNEL C/CON-16e is a less expensive version of the concentrator box, without the LED diagnostic display. It uses sixteen 16C450 UARTs and supports transmission speeds up to 19.2K bps.

DigiBoard makes several intelligent serial port boards. The DigiCHANNEL PC/Xi, which comes in 8- or 16-port models, is based on a 12.5MHz 80186 processor and the processor can be upgraded to 16MHz. It has 128K of onboard RAM standard, and can be upgraded to 256K or 512K. In addition to up to 16 asynchronous serial ports, the PC/Xi card can also be equipped with an optional synchronous port.

The DigiCHANNEL MC/Xi is for the IBM PS/2 Micro Channel Architecture. It comes in 4-, 8-, or 16-port models, is based on the 12.5MHz 80186, and has 256K of RAM onboard.

The DigiCHANNEL PC/Xe is available in 4-, 8-, and 16-port models. The 4- and 8-port versions are based on an 8MHz 80186; the 16-port model has a 10MHz 80186. It has 64K of onboard RAM.

The DigiCHANNEL COM/Xi, available in four- or eight-port models, is based on a 10MHz 80188 processor and has 256K of RAM onboard.

DigiBoard also makes a line of standard serial port boards. The DigiCHANNEL PC/X for the PC AT bus, and the DigiCHANNEL MC/X for the Micro Channel Architecture, are both available in 4-, 8-, and 16-port models.

DigiBoard also makes a multiuser graphics adapter. The DigiCHANNEL MV/4 provides four 640 x 480 color VGA channels for multiuser systems. Up to four of these cards can be installed in the same system for a total of 16 VGA graphics channels. Terminals can be located up to 200 feet from the host. An adapter box at the terminal allows the connection of a monitor, a keyboard, and a mouse to a single cable attached to the MV/4 card in the host.

### Star Gate Technologies, Inc.

Star Gate offers the Advanced Communication Link (ACL) series of intelligent serial port expansion cards. The ACL II+ is an eight-port board based on a 16MHz 80188 and is available with 16K or 64K of dual-ported RAM onboard. The ACL IIR+ is a 16MHz, 80188-based board with eight RJ-12 connectors that support low-cost telephone wire connections.

The ACL 16+ is a 16-port card for the AT bus that features two 80C186 processors. Each processor supports eight ports, and each has its own 64K of onboard RAM. The ACL MC and ACL MC16 fit the IBM Micro Channel Architecture bus and are built around 10MHz 80C186 microprocessor.

The PLUS 8 is an eight-port serial expansion board for the IBM PC XT and AT buses. A four-port version of the card is also offered, and it is upgradable to eight ports. The PLUS 8MC is an eight-port serial board for the IBM PS/2 Micro Channel bus.

## Multiuser Graphics Adapters

### Advance Micro Research, Inc.

The UnTerminal VNA (Video Network Adapter) is a full-size card for the IBM PC AT bus that supports monochrome text and Hercules-compatible graphics. Used in an 80386 host running in virtual processor mode, the VNA board provides a dedicated video display memory buffer and PC keyboard interface for each virtual terminal. The basic VNA card supports one workstation. Up to three additional VNA User Modules can be added to support a total of four workstations per card. VNA User Modules are piggyback cards that plug into connectors on the VNA card. Up to four VNA cards may reside in a single system, for a total of 16 workstations.

The VNA system also requires a Translation Unit and an Interface Unit. Since the video and keyboard signals are multiplexed over the cable, the Translation Unit is required to demultiplex the two at the host end. The Interface Unit splits video, keyboard, and I/O signals at the workstation end. The Translation Unit provides a 37-pin female adapter for parallel or serial I/O support for each workstation. The Interface Unit can be optionally equipped with a 25-pin female connector to attach a local parallel printer.

The UnTerminal VNA Plus (Video Network Adapter Plus) is similar in most respects to the original VNA adapter, but supports up to eight workstations. The base VNA Plus card supports four workstations, and four VNA Plus User Modules that each support a single workstation can be added.

The UnTerminal EGNA (Enhanced Graphics Network Adapter) provides up to four users with CGA or EGA graphics. An EGNA user module is added to the basic EGNA card to support each individual workstation. The EGNA card supports 640 x 350 alphanumeric color text and all-point addressable graphics. Like the other AMR cards, the EGNA card requires the external Translation Unit and Interface Unit at the host and workstation end, respectively.

The UnTerminal VGNA Plus (Video Graphics Network Adapter) connects four standalone VGA monitors and AT keyboards to an 80386 host PC. The VGNA Plus card supports a single monitor and keyboard, and additional stations are accommodated by adding up to three VGNA Plus User Modules. The User Modules are piggyback boards that attach to the VGNA Plus card. The VGNA Plus card and each User Module connect to a VGNA Plus Interface Unit that demultiplexes video, keyboard, and I/O signals. The Interface Unit also supports the attachment of a mouse and a printer. In contrast to the other AMR cards, only a single external box, the Interface Unit, is required with the VGNA system.

The UnTerminal Connect Card (UCC) replaces the Interface Unit when a PC XT or PC AT, instead of a terminal, is connected to an 80386 host PC equipped with a Video Network Adapter (VNA).

### SunRiver Corp.

SunRiver offers a broad line of products that enable text terminal and graphics workstation attachment to multiuser DOS systems via both fiber optic and shielded twisted-pair cabling.

SunRiver's Fiber Optic Station attaches to a 32M bps duplex fiber optic link. The Fiber Optic Station is offered in versions that support Hercules, EGA, or VGA graphics,

and each version includes a 101-key IBM PC AT-compatible keyboard. The Fiber Optic Station supports Digital Research DR Multiuser DOS, Virtual Systems Quick Connect 386, IGC VM/386 MultiUser, TSL PC-MOS386, Alloy 386/MultiWare, and Theos THEO+DOS.

LightAdapter is the name SunRiver gives to its fiber optic card designed for use in 80386 or 80486 host systems. AT, EISA, and Micro Channel models of the LightAdapter card are available. Each card supports four workstation attachments.

The PC LightCard allows a PC to act as graphics workstation when connected to a multiuser host. The card fits the IBM PC XT or AT bus and is connected to the host's LightAdapter card via fiber optic cables, and data is transmitted at 32M bps in full duplex. The PC LightCard comes in two versions—one works with EGA, CGA, and MGA monochrome; the other supports VGA. Both versions include a file transfer utility and are equipped with a serial port.

#### **Viewport International, Inc.**

The VPT System 1000 is a four-user monochrome text and Hercules graphics system that consists of a host controller card that plugs into an IBM PC AT bus slot, and a station controller for each monitor. The host and station controller modules are connected by a single twisted-pair cable at distances of up to 250 feet. The station controller module features a serial port that can be used to attach a mouse, a modem, or a local printer. Resolution is 720 x 350.

VPT System 2000 is similar to the VPT 1000 described above but supports two EGA or EGA+ users. EGA resolution of 640 x 480 is supported in both digital and analog modes. The station controller module has two serial ports available.

ViewPort's VGA product, the System 3000, is available in two versions—the VPT System 3000-2 card is a two-user VGA system, and the VPT System 3000-4, a four-user VGA system.

### **Turnkey Multiuser Systems**

#### **Star Gate Technologies, Inc.**

Star Gate offers the Star Light Graphics Display System based on IGC's VM/386 software and Star Gate's fiber interface board and workstation module. The system can support up to 32 workstations featuring VGA graphics. Each workstation module is connected by a ring of fiber optic cabling to an interface card in the host PC. Running at a transmission speed of 125M bps, the fiber optic connection provides performance that contributes greatly to the user's illusion of a dedicated machine of his or her own. The system's VGA display capability supports Microsoft Windows 3.0 with a mouse. Putting a network interface card in the PC host will allow users to access the resources of a LAN, providing a low-cost way to increase the number of users on a network without adding a network node for each workstation.

A typical PC LAN using 386 PCs with VGA graphics can cost at least \$3,000 per workstation. A multiuser graphics system can cost from \$1,800 to \$5,400 per station. The StarLight system costs under \$1,500 per workstation.

## Vendors

#### **Advance Micro Research, Inc.**

2045 Corporation Court  
San Jose, CA 95131 (408) 456-9400

#### **Alloy Computer Products, Inc.**

One Brigham Street  
Marlborough, MA 01752 (508) 481-8500

#### **Arnet Corp.**

618 Grassmere Park Drive #6  
Nashville, TN 37211 (615) 834-8000, (800) 366-8844

#### **Bluebird Systems**

5900 LaPlace Court  
Carlsbad, CA 92008 (619) 438-2220, (800) 346-8232

#### **Control Corp.**

2675 Patton Road  
St. Paul, MN 55113 (612) 631-7654, (800) 926-6876

#### **Concurrent Controls, Inc.**

880 Dubuque Avenue  
South San Francisco, CA 94080 (415) 873-6240

#### **DigiBoard**

6400 Flying Cloud Drive  
Eden Prairie, MN 55344 (612) 943-9020, (800) 344-4287

#### **Digital Research, Inc.**

70 Garden Court  
Box DRI  
Monterey, CA 93942 (408) 649-3896

#### **IGC, Inc.**

1740 Technology Drive  
San Jose, CA 95110 (408) 441-0366, (800) 458-9108

#### **S&H Computer Systems, Inc.**

1027 17th Avenue South  
Nashville, TN 37212 (615) 327-3670

#### **The Software Link, Inc.**

3577 Parkway Lane  
Norcross, GA 30092 (404) 448-5465, (800) 451-5465

#### **Star Gate Technologies, Inc.**

29300 Aurora Road  
Solon, OH 44139 (216) 349-1860

#### **StarPath Systems, Inc.**

4700 S. Hagadorn Road  
East Lansing, MI 48823 (517) 332-1137, (800) 456-8667

#### **SunRiver Corp.**

11500 Metric Boulevard  
Suite 150  
Austin, TX 78758 (512) 835-8001

#### **Theos Software Corp.**

1777 Botelho Drive  
Suite 360  
Walnut Creek, CA 94596-5022 (510) 935-1118

#### **Viewport International, Inc.**

4800 Great America Parkway  
Suite 410  
Santa Clara, CA 95054 (408) 748-8500 ■

# An Overview of Object-Oriented Programming

## In this report:

The Benefits of OOP .....	4
Popular Misconceptions .....	6
OOP Methodology .....	6

## Datapro Summary

As microcomputer environments increase in complexity, the programming models associated with them have followed suit. System services to communications networks, SQL servers, and other links to other environments are becoming commonplace. Programmers are struggling to keep up with the rapid pace of hardware innovation while maintaining the integrity and functionality of code. When working in a traditional structured environment, a change to a variable or subroutine can invalidate the entire program. Object-oriented programming (OOP) promises to relieve some of the hardship associated with maintenance and upgrades to existing code while providing a methodology and a set of tools for creating new programs which can be more easily developed, managed, and maintained.

## The Growing Interest in OOP

Consider the growing complexity of today's microcomputer environments and the wealth of new features and facilities they provide. Imagine a programmer trying to implement a Windows 3.0 version of a database management program which implements an SQL engine, dynamic data exchange, interprocess communication, communications links, and provides all the other features commonly associated with a DBMS program. The task is overwhelming.

These complex problems do not lend themselves to a simple solution. Because it would be crippling for a programmer to attempt to deal with all these issues simultaneously, a modular approach is almost required.

More and more programmers are using object-oriented programming to help resolve complex problems. OOP provides an avenue for programmers to break programs into manageable pieces while still dealing with the pieces as parts of a whole.

—By Karen J. Offermann  
Associate Editor

## What Is Object-Oriented Programming?

OOP combines programming tools and methodology. An object represents a collection of data and behaviors, and an object contains within itself all the information about its nature and functions.

The most common definition of object-oriented programming includes three key characteristics: encapsulation, inheritance, and dynamic binding. These factors differentiate OOP from the classic Von Neuman programming model, in which data has a type and a structure, is distinct from the program code, and is processed sequentially. The concept of a *framework* is also discussed here. Although a framework is not a requirement of object-oriented programming, it represents one of the key benefits of the technology.

### • Encapsulation

Encapsulation joins procedures and data so that users see only the procedures. The data itself is hidden from view. Encapsulation isolates the complexity of both the data and the internal workings of the object. Only the procedures (methods) of the object are "visible" to the outside world for use as in the way conventional programs use a handle.

## Planning for OOP Implementation: What a Manager Needs to Know

### Interview with Jason Matthews, President, Genesis Development Corporation

We would like to thank Jason Matthews, President, Genesis Development Corporation, for graciously sharing his insights on planning for the implementation of object-oriented technology in the corporate environment. A former MIS director for Bell Atlantic, Matthews has also worked for Morrissey Associates, Whitewater Group, and the NCR Cooperation project.

**Datapro:** Can you provide a background of your introduction to object technology?

**Matthew:** I was introduced to the technology when I went to work for a company called Morrissey Associates in Chicago. At the time, we [Morrissey] were utilizing **Whitewater's Actor**, which was, in its day, literally the best interactive object-oriented language on the planet. As far as I was concerned, it had more commercial applicability than anything [else available]. At the time, the only real other interactive development language was Smalltalk. We started

making Actor do things that Whitewater didn't know that it could do.

**Datapro:** How do you identify projects that lend themselves to an object-oriented approach?

**Matthews:** It really depends on what the company wants. But if it wants to embrace object technology, the worst thing a company can do is to embrace it in a big way initially. What it needs to do is to identify what I call a pet project. The project is really irrelevant whether it's CEO driven or DP manager driven. You need to find a noncritical project. One [the project] of significant interest to the decision makers in the company, but not one of critical complexity to the organization is needed. So, if it does fail you don't hurt the organization. You staff it [the project] accordingly and you build this subsystem or this entire system (as the case may be) utilizing the object-oriented technology. I would not say that OOPS is real useful for things like interrupt-driven clocks. Jobs of medium to large complexity lend themselves to it

[OOPS] better, if only from the standpoint that you really start beginning to achieve the economies of scale and reuse of code on medium to large projects. A lot of small projects are needed to build up a storehouse of code that is easily reused and built in a generic enough fashion to be useful to the enterprise.

If you were to build a small scale project, there is typically not enough genericity in what you're doing to provide a good enough abstraction for a set of objects. You're going to have to go out and create a whole bunch of common [or foundation objects] things like dictionaries. You're going to have to create all those [foundation objects] unless they come with the language you're using. That's an awful lot of work. C++, as a for instance, doesn't really come with a lot of foundation classes. Zortech has just released its latest, which is getting better. By and large, there wasn't a huge set of foundation classes that were truly robust. You could break them easily by doing things that were just a little bit odd. So, with C++, you had less benefit at the initial starting point than you would with something like Object1, or Smalltalk V/PM, or even Smalltalk-80. Those come with an enormous number of classes that are very well tested.

**Datapro:** In the long run, does C++ pay off over Smalltalk or Object1? Does something that

is fully object oriented have as much flexibility?

**Matthews:** Well, certainly something that is a late-bound, interpreted language will have much more flexibility than a nonlatebound interpreted language. You also lose speed, a lot of performance (there are other measures of performance than speed). Since you start out with less overhead in C++, sometimes, the code is smaller than in an interpreted language. If you take Actor, Object1 or Smalltalk, the whole environment goes along with those things to a great degree. Although you can strip out a lot of it, you still have a pretty substantial initial code hit. On a medium-sized project, it's sort of break-even. On a large project, you really start to gain the usefulness of all that code. All of the projects Genesis has been involved in are fairly large projects, all involve tens of thousands of line of code in Actor, Object1, or C++. The NCR project is more than 300K lines of C++ code. I can't think of a late-bound OOP language that would work as well for this project. We're building a platform, and that would be very difficult in Actor, Object1 or Smalltalk. These are either monolithic or semi-monolithic systems. That is—the systems think they are the environment, they don't participate in the environment.

**Datapro:** What advice do you have for corporate users who

### • Inheritance

Inheritance passes attributes to dependent objects, called *descendants*, or receives attributes from objects, called *ancestors*, on which they depend. For example: the superclass *automobiles* includes all four-wheeled passenger vehicles with internal combustion engines, the subclass *sports cars* inherits all the properties of automobiles and adds new characteristics of its own, such as a manual transmission and greater than 120 horsepower. The subclass *Porsche* inherits all the characteristics of *automobiles* and of *sports cars*, and adds new characteristics of its own.

### • Dynamic Binding

Dynamic binding is the process in which all linking occurs at program execution time; therefore all objects are defined at runtime, and their functions depend on the

circumstances at the time the program executes. Dynamic binding is also known as late binding. As an example, in *dynamic binding*, the instructions might state: *when you reach the point in the program where a lunch is called for, go to the refrigerator and get anything that will make a noontime meal.* The content of the *lunch* will depend on what is available and its state at the time the instruction is executed.

### • Framework

Framework is another emergent term in the world of object-oriented programming. A framework is a mature, tested library of reusable code which builds something else *about* that system. A framework is targeted to a particular application or application environment. The framework does not contain the actual application code, but it *does* contain code that enables development of the



are not familiar with OOP technology who want to investigate it?

**Matthews:** Identify the project. It is important to choose one that is of a non-critical nature initially. The reason is quite simply, you have a learning curve to get over. My experience has been that the single largest problem in getting this technology into large corporations has not been from the DP manager down, but from the CEO down to the CIO.

I typically recommend that once somebody decides to investigate it (OOP); they get a knowledgeable—and I underline the word *knowledgable*—consulting house to come in and help them identify a reasonable size project that isn't going to require an enormous amount of time or money to get done. Once that project is identified, I recommend that the company cordon off a corner of the development shop [lab] and put a couple of the corporate people in there. The consulting house brings in a couple of their object-oriented gurus to work with the company's people. These gurus, with the assistance of the company's people, go through the process of designing and developing the system to whatever point the client wants the system developed. Some companies stop at the prototype stage. Other times, a project will be taken through to conclusion [deliverable product].

On first projects, I believe that people should utilize an interactive development language for a number of reasons. First off, the language includes dictionaries; [second] these environments are very forgiving; and the third (and most important) thing is that you gain the purest viewpoint of object technology because they are latebound—everything in the system is an object.

**Datapro:** *The popular press still seems to believe that OOP is an immature technology, that it is only good for developers, for people who have worked as developers.*

**Matthews:** My experience is that the longer you've been in software development, the more difficult it will be to grasp object technology (unless you are one of the few who has a completely open mind to new technology).

If people are concerned that the technology doesn't bring much more into their business, or that you need to be a real pro to use it, the best thing for them to do is to talk with other companies that have utilized object technology.

**Datapro:** *Are there any misconceptions about object technology that you would like to address?*

**Matthews:** Yes, it is not the panacea for software development. But, you no longer

look at problem space/solution space. You look at problem space. If you understand the problem, you also understand the solution because you just defined it. You don't think of things in the traditional way. But you still have the initial learning curve. There's no doubt about it—this is a new paradigm. You have to train people to use new languages from both a syntactical and design viewpoint.

Where you really gain the benefits is not in Project 1, maybe not in Project 2, but Project 3 and on—you're going to gain some real benefit from the technology. You're going to be able to reuse the code you've developed, not just small modules, but entire, very complex systems. You'll be able to reuse or modify this code very easily and very quickly. And, this code is fully tested and robust, because you tested it right the first time.

So it [OOP] is not going to solve all your development problems today. I think it will solve your development problems of tomorrow. I think the days of developing complex systems in procedural languages are drawing to a close quite rapidly. The amount of maintenance and support spent on those things is astronomical and a lot of that problem goes away with OOP. But, you won't see it in the front end, you'll see it later. Management needs to be schooled

to get away from the instant gratification concept. You need to realize that object technology is a long-term investment. It's not a short-term investment with a quick payoff.

One thing people need to keep in mind when considering object technology is that code you are developing is a company asset. You are developing a reusable component that needs to be managed very carefully. I expect to see the development of a *class librarian*, a senior systems analyst who has significant experience and is charged with management and handling of classes—particularly with an eye toward its reuse potential within the business.

OOP is a whole new way of looking and developing solutions for the business world. It's not that difficult to learn. Once you learn one OOP language, learning the others is easy. Once you start building these objects and the protocols to manipulate them, business will be able to develop solutions as fast as their users wish them to. In this environment, maintenance problems can be eradicated much more quickly.

application or family of applications. The framework often provides memory management, exception handling, and a library of tested ready-to-use code which can be inherited by programs developed within that framework. The framework concept was first introduced in the Macintosh, Smalltalk (ParcPlace), and Lisp communities, but is now found in many C++ and Pascal implementations (e.g., Turbo Vision), among others. As object-oriented programming matures, we see many more frameworks emerging.

example, the Apple Macintosh user interface is an outgrowth of Xerox PARC technology. Smalltalk was developed at Xerox PARC as an object-oriented language and, after years of research, was released as a commercial product in 1986. Xerox PARC has spun ParcPlace Systems off as a separate company, which continues to provide innovative research and state-of-the-art OOP tools (Smalltalk and ObjectWorks for C++) to the programming community.

## The Origins of OOP

Xerox Palo Alto Research Center (PARC) is universally credited as the source of OOP. Xerox PARC's technologies, past and present, have had a significant effect on the look and feel of most graphical user interfaces today. For

## Object-Oriented Programming: Glossary

Many of the terms used to describe object-oriented programming systems may be new to our readers. Here are definitions of some frequently used terms.

**abstract class**—a class with no instances, usually created to organize a class hierarchy. Also known as a *virtual class*.

**abstract data type**—one that is programmer defined, rather than built-in. An abstract data type defines a data space and hides procedures and other details unnecessary for data manipulation. Its type definition contains both an internal representation and a set of procedures that access and manipulate the data.

**abstraction**—a design methodology that moves shared class behaviors into a new superclass. When this happens, the classes become subclasses of the new superclass and they share inherited behavior from the superclass.

**active database**—a database system where the retrieval and update operations result in calling, then invocation of procedures (also known as *triggers*). These procedures are associated with certain fields and when the field is accessed, the trigger is activated.

**attribute**—a property of an entity, which in turn is described by its attributes. In an object-oriented database,

instance variables can be attributes of the objects stored in the database.

**base class**—a C++ specific term describing a class from which other classes are descended (similar to a Smalltalk *superclass*).

**behavior**—actions that can be called forth from an object. Behavior permits viewing of the object's information, changes to its behavior, and interaction with other objects (through messages).

**browser**—a tool to display an object's definition.

**class**—A group of similar objects that each contain a set of shared characteristics.

**class hierarchy**—a natural organization of classes (either tree or network). There is always a top node. In a hierarchy, the classes above a particular class are known as *superclasses*, those below are referred to as *subclasses*.

**class library**—related classes belonging to a particular domain.

**class method**—a method (procedure) that is invoked by sending it to a class rather than a class instance, usually to perform tasks that shouldn't be performed by the instance—like creating and destroying instances.

**class object**—a class definition. A class definition may be an instance of a generic class or of a metaclass.

**class variable**—The part of a class that sets it apart from other classes, but which also becomes the common part of the class' *instances*.

**code reuse**—the use of a piece of code for more than one purpose in a computer application. Code reuse reduces the amount of code needed for application implementation.

**complex object**—an object composed of other objects. These objects are actually collections of parts, each of which is an object. Also referred to as a *composite object*.

**composite object**—see *complex object*.

**concurrency control**—a mechanism to regulate user access to objects, preventing execution of inconsistent or destructive actions on the database.

**data abstraction**—a programming technique that leaves the user with a partial view of internals and operations of an object, displaying information relevant to a particular action on an as-needed basis. The computational methods remain hidden from external view.

**data model**—a specification of the database structure, operations, and integrity rules.

**database schema**—complete set of individual schema definitions describing the logical structure of a database. An ODB schema is a set of class definitions.

**dynamic binding**—Also known as "late binding," the process in which all linking of objects occurs at program execution; thus, all objects

are dynamic, and their functions depend on circumstances and conditions at runtime.

**encapsulation**—The joining of procedures and data so that the only part of the resulting object users see are the procedures; the data is hidden from view.

**extensible database management system**—a class of DBMS that includes additional data modeling capabilities combined with data management services for applications that cannot make easy use of conventional DBMS.

**extensibility**—ability to dynamically update the database schema. Generally this includes adding support for new data types (definition and manipulation) such as voice data, images, and knowledge data for AI.

**enterprise modeling**—creation of a working organizational model in order to understand it and translate that understanding into software function.

**entity**—a collection of information items that can be grouped together and distinguished from their surroundings. Entities are described by their attributes and can be linked to, or have relationships with, other entities.

**expert system**—software that considers a facet of human endeavor with regard to the rules governing its activities. This software can work out new problems by utilizing the rule set.

**factory object**—The original *objects* of each *class*, built at compile time.

**functional decomposition**—a method for analyzing requirements before designing a program.

## The Benefits of OOP

### Cost Reduction

First and foremost, OOP saves money. When properly implemented, it can reduce software production and imple-

mentation costs, providing easily maintained code and allowing programmers to decrease time spent prototyping and debugging code. It is important to note, however, that programmers new to object-oriented programming must accept a learning curve, and that in many cases the payoffs

A goal is established for the program, then the goal is broken down into a series of steps to meet the goal. Each step is broken down further into more steps, then each step becomes a separate program module.

**generalization**—relationship between superclass and subclass. A superclass is a generalization of its subclasses.

**handle**—pointer to, or address of, an object.

**information hiding**—hiding the internal details of a module from other modules to prevent unauthorized changes that could affect other dependent modules.

**inheritance**—The method by which *objects* pass attributes to dependent objects, called descendants, or receive attributes from objects, called ancestors, on which they depend.

**instance**—One of the components of a *class* that share the same code, but have differences in values. A single occurrence of an object.

**instance variables**—The parts of an *instance* that set it apart from other instances. The attribute of an object.

**message**—A predefined request to an *object* (from another object) that activates its *methods*.

**message passing**—this is the way objects communicate. A message may consist of the message name, the target object name, and any arguments. Upon receipt of a message, an object invokes a method, which performs an operation exhibiting the object's behavior.

**message selector**—The part of a *message* that tells the *object* what to do.

**method**—The procedure, or callable operation, contained in an *object*. The code that is executed in response to a message.

**modular programming**—breaking down of a program into separate parts, known as modules. Each module has its own data and procedures and has the ability to operate independently, with minimal intermodule interaction.

**multiple inheritance**—The capability of an *object* to inherit characteristics from more than one ancestor. Since descendant objects always dynamically refer to their ancestors for data, any changes made to an ancestor are always reflected in its descendent. Multiple provides for some overlap in the definition of objects, because an object can inherit characteristics from a number of superclasses. This helps provide flexibility.

**object**—code that contains data that defines both the characteristics of the code and its procedures or operations. An object performs predefined actions in response to *messages*.

**object-class library**—A hierarchy of possible *objects*.

**object database management system**—a DBMS built to store and manipulate object data types.

**object ID**—a permanent and unique identifier assigned to an object, independent of the value of the object's instance variables, and constant despite changes of object state. Often used instead of the term *handle*.

**object server**—software that supports transaction management and storage management functions for objects.

**overloading**—the practice of giving a number of meanings to the same method name, which in turn permits a message to perform different functions, depending on the receiving object and the accompanying parameters.

**overriding**—overloading with a twist—the same name is given to one variable at two or more levels of a branch of a class hierarchy. The lowest name in the hierarchy takes precedence.

**paradigm**—a learned way of thinking.

**paradigm shift**—transition to a new paradigm. Often met with resistance.

**parallel processing**—when a processing task is split among a number of hardware processing units.

**part hierarchy**—hierarchy of component objects that join to form a composite object. Composite objects can be constructed of component objects that also contain components. This is not the same as a class hierarchy where class relationships are established through *inheritance*.

**persistence**—data or an object that has a longer lifetime than the code that created it.

**persistent object store**—an object database.

**polymorphism**—Greek for "many forms"; sending the same *message* to more than one *object* in order to invoke *class-dependent methods*. Polymorphism allows generic code to be reused, since it performs the same task on different types of objects.

**protocol**—the interface or set of messages an object can respond to, generally defined in the class definition.

**referential integrity**—RDBMS term meaning that no record contains references to a primary key for a nonexistent record.

**runtime binding**—see dynamic binding.

**shadowing**—a subclass method to replace a method that would otherwise be inherited from a superclass.

**specialization**—relationship between super- and subclass. A subclass is a specialization of its superclass.

**state**—set of values for an object's instance variables. Changing the instance variables changes the object's state.

**static binding**—Also known as "early binding," the process in which all references by objects to items are resolved during compilation; therefore, nothing needs to be looked up during the program's actual execution.

**subclass**—In the hierarchy of *inheritance*, a group of objects that are descendants of another group, called a *superclass*, and which contain variations on the superclass.

**superclass**—In the hierarchy of *inheritance*, a group of objects that are ancestors of a *subclass*. A superclass may itself be a descendant of another superclass. A superclass is generally found in the Smalltalk language.

**virtual class**—See *abstract class*.

**virtual function**—a C++ function that only executes at runtime.

will not be evident until the second or third project. For more information, see the section titled "Planning for OOP Implementation: What a Manager Needs to Know about OOP".

Competition is forcing software developers to bring their products to market ever more quickly, but the applications the market demands are growing ever more complex. Key development companies, which have either implemented OOP techniques for in-house development or

have released object-oriented tools, unanimously cited productivity gains as a key strategic benefit of OOP.

Microsoft, for example, incorporated OOP techniques in developing Quick Pascal. So few bugs were found during the debugging process that the development team initially thought the results were wrong.

Borland International's Turbo Pascal 6.0 is the company's most recent version of Pascal with object-oriented extensions. At that time, we spoke with Borland's Gene Wang about the move to OOP techniques among microcomputer software vendors. Wang had cited productivity gains but cautioned that the design phase of OOP software development is critical to a successful implementation; revisions of the object hierarchies should be expected. His cautions and expectations still hold true.

### Helps Master Complexity

Systems and operating environments offer an increasing number of communications links, access to remote databases, graphical user interfaces (GUIs), and dynamic data exchange (DDE), yet conventional coding practices have done nothing to help programmers master these complexities.

The competitive pressure of a microcomputer market in transition has created a crisis for many software vendors. Because software vendors must meet their own impossible, but announced, release deadlines, some features are compromised in order to bring packages to market. Vendors must develop programming techniques and tools that permit programmers to incorporate all the facets of the complex environment into today's applications. Vendors overwhelmed by this complexity may view OOP as a lifesaver. The vendors we spoke with have cited productivity gains, reusable code, and ease of maintenance and upgrades as reasons for shifting to OOP. The upsurge of interest in OOP is a natural result of delays in software development. Although hardware capabilities have increased dramatically (about a thousandfold every 10 years), software has not kept up with hardware growth. OOP techniques may hold the promise of helping software earn its keep.

### Popular Misconceptions

OOP requires an approach that is different from the traditional structured methods prevalent today. This new approach has led to (and continues to carry with it) a number of popular misconceptions. Some of the myths surrounding OOP have been enumerated by Adele Goldberg of ParcPlace Systems:

- **You must throw everything away.**  
This is not true—good programming techniques, once learned, can be easily applied to this new paradigm.
- **Object-oriented programming makes everything easy.**  
It does not. In fact, according to Goldberg, the level of effort in creating readable code makes things harder. The implementation description is very important.
- **You can reuse everything.**  
There are those who believe that you just plug things together. This misconception leads to a poor implementation architecture.
- **Everything revolves around inheritance.**  
Actually, it is only one of several object relationships.
- **Object-oriented systems—and Smalltalk in particular—are interpreted and very slow.**

Compilers are now available, and the performance has been improved dramatically.

Some revisions to program design are given in OOP. In fact, Goldberg is quite emphatic that the first design is usually simply a first cut and should be discarded. Patching that first design is not a good idea because it generally leads to a hodgepodge implementation architecture. The recommended procedure is to budget for redesign.

### OOP versus Traditional Programming

In a traditional, procedural application, the user typically has a limited range of options. At any given point in the program, the programmer would have total control over what options the user would pick at that time. But in a GUI such as Windows 3.0, Presentation Manager, or the Apple Macintosh, the user can go to any window, icon, or menu choice at any given moment in an apparently random way. That is an event-driven system.

Object-oriented environments are message- and event-driven. That is, an object is a sort of black box that exists in a space, reacting to events (such as key presses and mouse clicks) and communicating with other objects through messages.

### OOP Methodology

Programmers working with object-oriented technologies have two learning curves to accept. They must learn to work with the tools provided to create object-oriented programs, but more important, they must learn to work with the programming style associated with OOP. In order for an OOP project to succeed, it is necessary to take a hard look at the expectations and concerns of the end user, the software consultant, and the systems manager, as well as the integrator and professional programmers—what each of them needs and wants and how they interact with each other. OOP projects tend to evolve as they are implemented, which differs from the definite specifications that generally accompany a traditional structured project.

### Varieties of OOP

Three distinct kinds of OOP exist today—high-level languages, environments, and other development tools.

### Enhanced High-Level Programming Languages.

Some high-level languages have been enhanced with object-oriented tools or extensions for creating object-oriented code. This combination of an existing programming language with object-oriented extensions is considered a *hybrid approach*. C++ from AT&T, Objective C from Stepstone, Turbo Pascal 6.0 from Borland, and Quick Pascal from Microsoft are examples of traditional languages enhanced with object-oriented functions. Programmers can use these languages to create structured programs, but the enhanced languages offer programmers the option of writing object-oriented code.

There are those within the OOP community who feel that corporate users would more readily accept OOP tools which are more an extension of existing programming tools.

### Completely Object-Oriented Environments

Languages (or environments) such as Smalltalk 80 from ParcPlace Systems, Smalltalk/V PM from Digitalk, MacApp from Apple, and Actor from the Whitewater Group provide a complete OOP environment. Because

they are usually written in an object-oriented language, a programmer working with these tools cannot *help* but create object-oriented code. These languages are considered a *pure* approach to OOP.

**Other Options**

Programs such as Hypercard and Easel provide some of the characteristics of an object-oriented environment, but

the user must work with an existing set of predefined objects and procedures (e.g., the cards with Hypercard). It is impossible to subclass using these products; the user cannot define new objects or procedures but must manipulate those provided with the program. ■



# An Introduction to Windows

## In this report:

What Windows Can Do .....	2
Features and Benefits of Windows .....	3
New Features for Windows 3 .....	5
Windows Operating Modes .....	7
Hardware Requirements for Windows 3 .....	7

## This report will help you to:

- Learn the benefits of using Windows.
- View the features and functions of Windows 3.
- Know the minimum hardware and software requirements for running Windows.

### Introduction

Microsoft Windows is a windowing and multitasking environment for personal computers that runs PC-DOS and MS-DOS. A *window* is a rectangular box that holds a running application. You can have several windows open at once. These windows can overlap one another, and the topmost window usually holds the currently running application. Since multiple applications can be open at once, you need not exit one to use another. Instead, you simply switch windows. *Multitasking* is an additional feature of Windows that gives it the ability to run more than one application at the same time. This means you can actively sort a database in one window while writing a letter in another. Multitasking

is only possible on 80386 or more advanced machines.

Windowing and multitasking are probably the most important characteristics of Windows, but the *data transfer* feature also enhances the value of the package. It gives you the ability to grab a block of text, a table of numbers, or a picture from a document in one window and “paste” it into another window. This means you can easily insert a graphics image in a letter, or insert spreadsheet information in a report. Transferring information between programs is possible because all applications written for Windows support the same data transfer capabilities. You can even cut and paste between some applications not specifically written for Windows.

Windows is similar in look and feel to its OS/2 cousin—the Presentation Manager. In fact, if you plan to use OS/2, and the Presentation Manager in the future, Windows is

---

This Datapro report is a reprint of Part 1, “Introducing Windows” pp. 5-25, from *Windows 3 Made Easy*, by Tom Sheldon. Copyright © 1990 by McGraw-Hill, Inc. Reprinted with permission.

an excellent environment in which to work in preparation for your transition.

Although Windows has been around since 1985, version 3 is a radical departure from previous versions in its power, flexibility, ease of use, and graphics interface. Many new features have been added, including the ability to handle memory beyond the 640K barrier normally associated with DOS. In addition, an enhanced appearance along with new ways of organizing your programs and working with files make Windows a program you will want to have running at all times, not just when you need to run multiple applications or transfer data between documents. Essentially, the DOS command prompt is a thing of the past.

### What Windows Can Do

At its most basic level, Windows offers a way to organize your programs on top of an *electronic desktop*, as shown in Figure 1. Each window holds its own set of *program icons*, and each icon represents a program or utility that starts when you double-click on the icon with the mouse.

Figure 1 shows three stackable, overlapping windows that can be opened or closed as you see fit. Note that the Main window is overlapping the Accessories window. These are both contained within the Program Manager, which is a window used to manage all of your icons in groups. The Program Manager window and most other windows can be expanded to cover the entire screen or reduced to an icon while working in other windows.

Figure 1.  
*The Windows Program Manager with its Main and Accessories Windows Open*

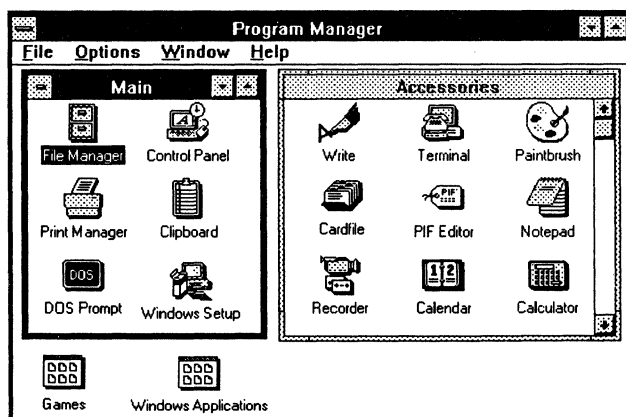
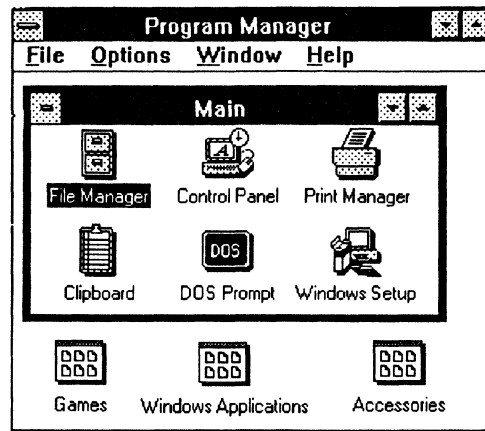


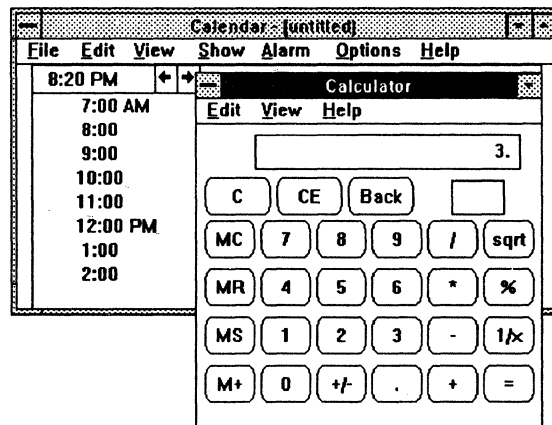
Figure 2.  
*The Program Manager Window Resized, with the Accessories Window Reduced to an Icon*



In Figure 2, the Program Manager window has been reduced to a smaller size and the Accessories window has been reduced to an icon at the bottom right of the screen. Windows reduced to icons are available for immediate use. Simply double-click on the icon to reopen the window.

Figure 3 shows two programs loaded in separate windows. The Windows Paintbrush drawing program is the active top window, and the Windows Cardfile accessory is the inactive window beneath it. Notice that the Program Manager has been reduced to an icon since it is not being used. The Paintbrush and Cardfile windows are both open at the same time so the art in the Paintbrush window can easily be copied and pasted to the Cardfile window. Cardfile can store both graphics images and text for later use. In this way it acts as a

Figure 3.  
*Pictures or Text in One Window Can Be Cut and Pasted into Another Window*





glossary, but as you will see later, it has many other uses.

As another example, Figure 4 shows the Windows Calculator open at the same time as the Calendar accessory. You can have the calculator resting on your desktop to perform calculations as you go through the previous days activities and expenses.

You can see there are many ways to use multiple windows and multiple applications at the same time. The more you use Windows, the more you'll find that you can't do without it.

## Features and Benefits of Windows

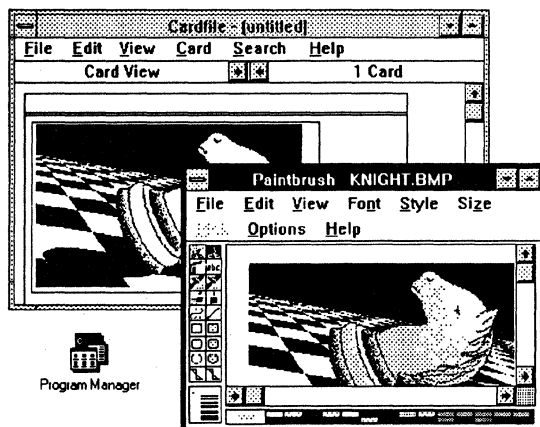
Windows offers many features that benefit both new and experienced PC users. Several of the operating features were discussed and illustrated in the previous section. This section describes other features that may not be so apparent.

### Foundation for Program Development

A wide range of software applications have been written to the Windows environment, including Aldus PageMaker, Microsoft Excel, and Micrografx Designer. The Windows user interface is already in place, so programmers simply design their programs to work with that interface. This lets them concentrate on the real guts of the program itself and thus create programs with advanced features that are relatively free of problems and easy to use.

Figure 4.

*The Windows Calculator Can Be Opened on the Desktop for Quick Calculations While Working in Other Applications*



Windows contains a complete programmer's toolbox that speeds the development of software. You can be assured that just about any type of application will be available in the Windows environment now that Windows (together with the Presentation Manager) has become a standardized graphics user interface.

### Consistent User Interface

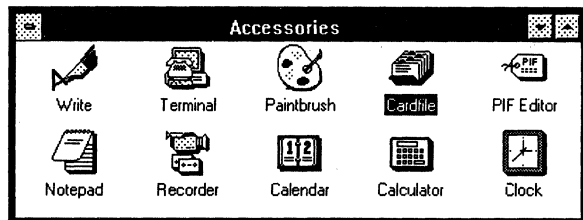
Since programmers design their applications to fit within the Windows user interface, you will already know how to start using most of the applications you buy for Windows. You will spend less time learning how to use your applications and more time getting productive work done.

### Multitasking Capabilities

*Multitasking* gives you the ability not only to load multiple applications in multiple windows, but also to actually run those applications simultaneously. This is not to be confused with *multiloading*, a term used to describe multiple loaded applications that do not run simultaneously. A loaded application resides in memory but sits idle while another application runs. When you switch to another application, it begins to run and the other becomes idle. True Windows multitasking occurs when two or more applications in different windows actually run or process information simultaneously. Technically, even this is an illusion since each application is simply given a small piece of the microprocessor's time, a process known as *time slicing*. But with advanced processors like the 80386 and 80486, this happens so quickly that you can take advantage of multitasking to do several things at once, like write a letter while your mailing list is being sorted or printed.

Multiloading simply saves you the trouble of exiting one application and loading another as you go from one to the other. Each sits idle in its own window until you make the window active. Obviously, true multitasking would seem to be more desirable, but your system must have at least an 80386 microprocessor to take advantage of it. As you run additional applications, you will see some slowdown as the microprocessor time is sliced up among each running application. But consider the advantages. For example, a chess playing computer may take some time to calculate its next move against you, the opponent. This time often in-

Figure 5.  
The Accessories Window



creases with the skill level you have chosen. Multi-tasking lets you switch to another window and write a letter while the computer “thinks.”

### Data Transfer Capabilities

The Windows Clipboard utility is used to exchange graphics, scanned images, spreadsheet data, or text between applications. The Clipboard makes all of your Windows applications (and some non-Windows applications) act as if they are part of a single integrated software package. You can think of it as a translator, because in most cases it will convert images in one format to another during the transfer between applications. Imagine the Clipboard as you would a real clipboard—a place where you clip pieces of art or text while moving to another location. In Windows, the Clipboard holds this information as you switch between windows, allowing you to mimic the common routines of cutting and pasting.

### Desktop Accessories

Windows offers several accessories and utilities you will want to use every day. These are located in the Accessories Group window, shown in Figure 5.

A description of each item follows. During your normal Windows sessions, you may want to keep one or more of these accessories loaded on the desktop for easy access.

**Write**—Used to write, edit, format, and print documents. Although not a full-featured word processor like Microsoft Word, it is useful for memos, letters, reports, and other every day documents.

**Paintbrush**—A convenient, easy-to-use drawing program. Used to create simple to complex drawings for your Write documents or any document created with any Windows application.

**Terminal**—A communications program used with a modem to connect with other computer systems or on-line data services over the phone lines.

**Notepad**—Allows you to create, store, and quickly retrieve notes, and memos your Windows sessions.

**Recorder**—Allows you to save keystrokes and mouse movements so you can repeat them at any time. Useful for establishing procedures for novice users or for replaying keystrokes and mouse sequences you perform on a regular basis.

**Cardfile**—An electronic card filing system with many of the sorting and searching features of advanced database systems.

**Calendar**—An appointment scheduling utility with day and month views, appointment scheduling, alarms, and other features.

**Calculator**—A desktop calculator with two modes, standard and scientific.

**Clock**—Produces an analog or digital time display you can place on the screen during idle computer time.

**PIF Editor**—Program Information File (PIF) editor creates and alters files used by Windows to run non-Windows applications.

### Windows Control Panel

The Windows Control Panel offers a complete set of utilities for controlling the operation of your computer and Windows. The Control Panel window with the utility icons it contains is shown in Figure 6.

Here is a description of each utility in the Control Panel:

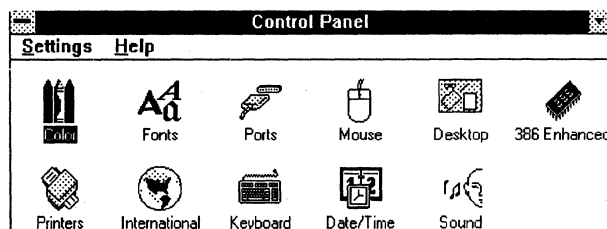
**Color**—Used to alter the foreground, background, border, and other color schemes of Windows.

**Fonts**—Used to add, alter, and remove screen fonts.

**Ports**—Controls parallel and serial ports for printers, modems, and other devices that connect to your computer.

**Mouse**—Controls the operating characteristics of the mouse.

Figure 6.  
The Control Panel



**Desktop**—Used to alter the features of the desktop.

**386 Enhanced**—Only appears on 80386 and 80486 systems and is used to control multitasking features.

**Printers**—Used to add and remove printers or alter their settings.

**International**—Sets keyboard, date, time, currency, and number features and formats to US or international standards.

**Keyboard**—Sets the keyboard repeat rate.

**Date/Time**—Sets the date and time.

**Sound**—Sets the warning beep on or off.

### Dynamic Data Exchange

Another feature used to exchange data is Dynamic Data Exchange (DDE), which is more automatic than the Clipboard and relies on features programmed into the application being used. For example, a data analysis package in one window could supply data points to a graphics program running in another window. As the analysis package processes in the background, you can watch the bar charts change in real time. The plotting of election results comes to mind. Taking this a step further, election results could be entered in a database by an operator or input device as they arrive. The new data is fed to the analysis package as required, which then updates the graphics display.

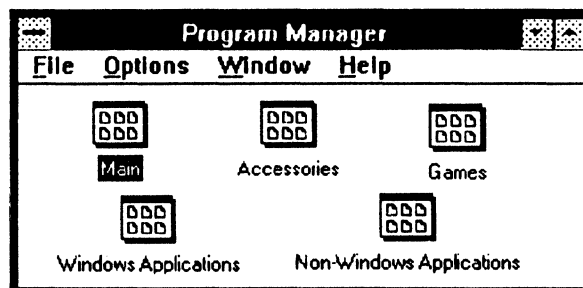
### DOS Compatibility

Since Windows is a shell over the DOS environment, you can still run most DOS applications by simply opening a window to the DOS command prompt. In addition, many DOS commands can be executed from the File Manager using the familiar Windows interface. These include the DOS commands for copying, deleting, renaming, and changing file attributes, as well as commands for creating and removing directories. You can also format and label floppy disks.

### Increased Productivity

While increasing productivity may sound like computer jargon from the '80s, the concept is valid when it comes to Windows 3. In the past, some people probably wondered whether a computer really made them more productive, since the process of learning and using the computer took more time than it was worth. But, as hardware and software become more sophisticated, we can begin to

Figure 7.  
Program Manager Screen



see what increased productivity is all about, and Windows is no doubt one of its realizations. It brings you a working computer environment that offers software integration, speed, and ease of use. At the same time, your learning time decreases because Windows applications have the same consistent user interface that you need not relearn with every new application you buy.

### OS/2 Look and Feel

If you are interested in making a transition to the OS/2 Presentation Manager environment at some time in the future, or if you often move to other systems that run OS/2, you will find Windows an excellent environment to use and grow with.

### New Features for Windows 3

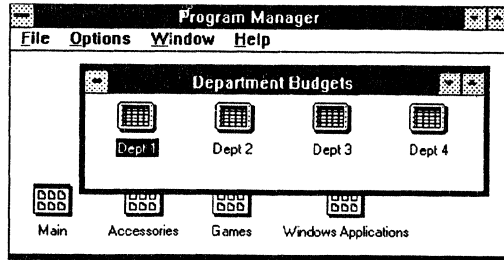
If you have been using previous versions of Windows, you will be interested in learning about the new features in Windows version 3. The most significant additions are the Program Manager, File Manager, and Task List.

#### Program Manager: The Group Organizer

The Program Manager is the first window displayed when you start Windows 3 and is used to organize your applications, utilities, and files into meaningful groups, making access to them easier. Figure 7 shows how a typical Program Manager screen appears with all windows reduced to icons.

The icons are created during the installation process by Windows to hold the various programs and utilities that come with Windows, or that may already exist on the system. A powerful feature of Windows 3 is its ability to completely search a system for existing applications and create start-up icons for those it recognizes.

Figure 8.  
Document Icons



Not all group windows necessarily hold programs and utilities. In Figure 8, a group window called Department Budgets has been opened to reveal *document icons*, in this case, Excel spreadsheet documents.

By double-clicking on any of the document icons, Excel is started and the spreadsheet for the associated department is loaded. This example reveals how the Program Manager can be used to organize your system according to the tasks you perform on a regular basis. It also points out Windows' ability to start an application and load a document by clicking on the document icon.

### File Manager

The File Manager is one of the most practical additions to Windows. It allows you to use the mouse to work with files and directories in a whole new way. The File Manager is illustrated in Figure 9. Inside its window is the Directory Tree window on the left, which shows a graphic representation of your hard drive filing system. On the right is a *directory window*, which shows the files in a directory

Figure 9.  
The File Manager

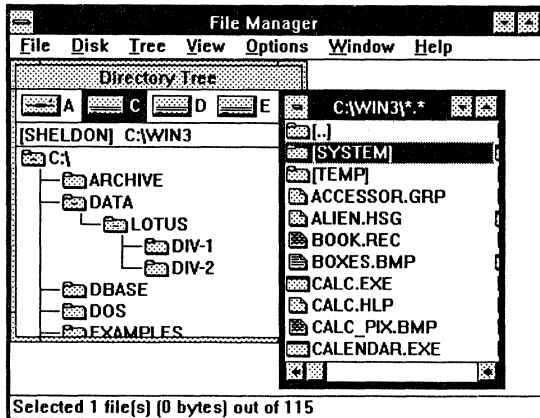
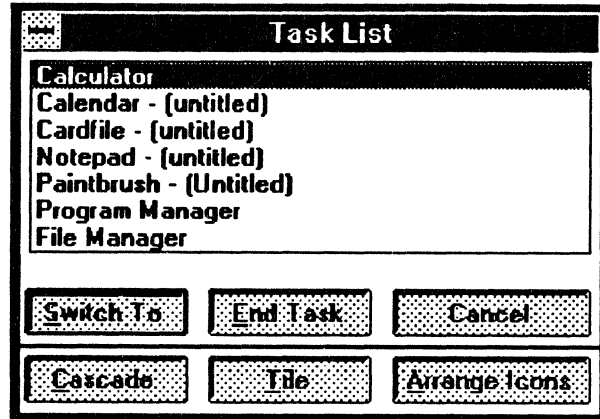


Figure 10.  
Task List



called WIN3. Notice that files have icons associated with them to help you differentiate between programs, directories, and normal files. Copying a file is easy: simply drag a file icon from a directory window to one of the directory icons on the Directory Tree.

Floppy disk and hard-disk drives are represented as icons to which you can switch by clicking with the mouse. Directories are also easy to get to by clicking the mouse. Subdirectories branch from their parent directories, making it easy to move to any directory quickly and easily. Files can be viewed and managed by double-clicking on any directory or subdirectory icon.

### Task List

The Task List is a useful tool for moving from one window to another, especially when a lot of windows are on the desktop. The Task List can be opened at any time by double-clicking on the desktop. In Figure 10, a number of applications are listed in the Task List, each representing an application in an open window or icon.

To quickly switch to any application, simply click on its name. The Task List is also used to rearrange the windows and icons on the desktop.

### Network Support

Windows 3 recognizes network connections and allows you to display and work with files on network drives and to use network printers. Popular networks such as Novell NetWare, Microsoft LAN Manager, and Banyan VINES are supported.

### Additional New Features

Windows 3 presents a whole new appearance that is the result of enhanced colors, proportionally spaced fonts, and interesting graphics icons. Along with this new interface are these features:

- Easy installation. Windows determines how it should be installed and searches your hard drive for applications that can be run under Windows. It then creates start-up icons for the applications.
- Full, on-line help facilities.
- Enhanced Print Manager now supports network printing.
- The Recorder accessory allows you to save keystrokes and mouse movements so they can be repeated at any time.
- Enhanced terminal communications program.
- Expanded printer support.
- New, expanded symbol font.
- An enhanced 386 mode provides even more compatibility with non-Windows applications.
- Solitaire, a new game.

---

### Windows Operating Modes

Windows operates in one of three modes, depending on the type of hardware you have. When Windows is first started, it automatically determines which mode to use, but you can choose to run Windows in a particular mode if the software you are attempting to run will not operate in the mode Windows selects. This may be the case with older Windows applications or non-Windows applications.

The three operating modes are discussed here.

#### Real Mode

Windows automatically runs in Real mode if your system has less than 1 megabyte of memory or if it uses an Intel 8086 or 8088 microprocessor. Real mode is the slowest and least desirable running mode for Windows. In addition, it does not take advantage of extended memory, which means that Windows has less memory available to run multiple applications and may not be able to do so at all. In some cases, you may want to run Windows in

Real mode, even if you have a more advanced system, because applications written for previous versions of Windows may only be able to run in Real mode. If you have such applications, Real mode can be activated by starting Windows by typing WIN/R on the DOS command line.

#### Standard Mode

Windows automatically runs in Standard mode if your system has 1MB or more of memory and if it has an Intel 80286 microprocessor. In Standard mode, Windows can access extended memory, and you can switch among non-Windows applications.

#### 386 Enhanced Mode

Windows loads in 386 Enhanced mode if your system has an Intel 80386/80486 microprocessor and 2MB or more of memory. True multitasking of Windows and non-Windows applications is possible in this mode. When Windows runs in 386 Enhanced mode, it uses a special operating mode of the 80386/80486 microprocessor known as the virtual 86 mode. In this mode, the 80386/80486 acts like separate 8086 microprocessors for each open window with a running application. Each virtual 8086 machine runs in its own protected environment—if a program crashes in another virtual 8086 window, the entire computer system is not brought down.

---

### Hardware Requirements for Windows 3

Windows operates on most DOS-based computers, including personal computers with Intel 8088 and 8086 processors, although the performance on these systems is not spectacular. Windows runs best on AT-type systems with Intel 80286 microprocessors or advanced systems with Intel 80386 or 80486 microprocessors.

In general, the hardware you use should be 100 percent compatible with the tested hardware list you received with Windows. If you have problems running Windows, you may need to obtain special software from the hardware manufacturer to make your system compatible with Windows.

The minimum software and hardware requirements are outlined here:

- PC-DOS or MS-DOS 3.1 or higher.

- A personal computer based on the Intel family of 8088/8086 microprocessors. This family also includes the Intel 80286, 80386, and 80486.
- Memory requirements as listed in the next section.
- A hard-disk drive and at least 8MB of free space.
- Graphics monitor.
- Although a mouse is not required, operating Windows without a mouse is difficult.
- If you intend to use the Windows Enhanced Terminal communications software, you need a Hayes-compatible modem.

### Other Considerations

There are a few things you must consider before you install Windows. These have to do mainly with the type of system you intend to use and the mode it will be operating in.

### Memory

Windows may use expanded or extended memory, depending on the run mode. Expanded memory is only used in the Real mode, so if you have an application that requires expanded memory, you may need to run your system in this mode. Standard and 386 Enhanced modes use extended memory because it is much faster and efficient than expanded memory. If you have an 80386 system and you need to run an application that requires expanded memory, you can install a special expanded memory simulator.

The expanded memory used in Real mode must conform to version 4.0 of the Lotus-Intel-Microsoft Expanded Memory Specification (LIM EMS 4.0). Because extended memory is preferable to expanded memory, you should use whatever method possible to take advantage of it. While this is not possible on systems based on the 8088 and 8086 microprocessor, 80286 systems with installed memory boards like the AST RAMpage! or Intel Above Board/AT should be installed with those boards set for extended memory. You should then run in the Standard mode whenever possible.

The following is an outline of the memory requirements for the three different types of systems:

**8088/8086**—Systems in this class can only run in Real mode and can only use expanded memory. Set your memory expansion boards for expanded memory. You will need 640K of conventional memory and as much expanded memory as possible to conform the LIM EMS 4.0 specification.

**80286**—Systems in this class can run in the Standard mode. One megabyte of memory is required but 2 megabytes or more is recommended.

**80386/80486**—Systems in this class will run 386 Enhanced mode. Two megabytes of memory is required but more is recommended.

### Older Windows Applications

If you have applications that were written for older versions of Windows, you may not be able to run them in the Standard or 386 Enhanced mode. You will need to start Windows in the Real mode to use these programs until you can get an update from the software manufacturer.

### Memory-Resident Software

Memory-resident software remains in memory after you start it, even if you start another application. This type of software may pose problems for Windows. Most applications that are simply loaded and have no further interaction with the user can be loaded in the normal way before starting Windows. So called *pop-up programs* that interact with users at any time may need to be loaded after Windows is started.

### Disk-Caching Programs

Windows uses its own disk-caching program called SMARTDrive to improve access to your hard drive. In most cases, you should replace your existing disk caching program with SMARTDrive. SMARTDrive increases the efficiency of your system by keeping previously read hard-disk information in memory just in case you need it again. Most computer systems typically read the same blocks of information from a disk system on a regular basis, so programs like SMARTDrive offer a dramatic improvement in overall speed. This is especially true if you are running several applications at once. SMARTDrive requires at least 512K of extended memory or 256K of expanded memory. ■

---

# An Overview of Environments and GUIs

---

## Datapro Summary

Initially developed to provide multitasking of DOS applications and a user interface that was friendlier than the DOS command line, environments for microcomputers have undergone significant change over the past few years. While Apple Computer provided the Macintosh with the graphical user interface that came to epitomize "user friendly," DOS and UNIX environments (or graphical user interfaces) have changed many users' perception of what a computer can do. When Hewlett-Packard added object management features to the DOS/Windows environment, the functionality available from a DOS-based system became the equivalent of the Macintosh.

---

## Technology Overview

Studies prove that a well-designed environment or graphical user interface (GUI) increases user productivity, decreases training costs, and provides access to multiple programs after a user learns just one package. While Apple Computer pioneered the commercial GUI, other developers in the DOS, OS/2, and UNIX markets have recognized the merits of GUIs and are working to provide graphical front ends for most operating systems and applications software.

As for environments, when an application-dependent user considers moving from the single-tasking MS-DOS to a multitasking operating system, only to find few available applications, a DOS environment such as Windows 3.0 or DESQview 386 will allow both multitasking and the use of the existing DOS applications.

## What Is, or Isn't, a GUI?

Graphical user interface (GUI) presents a graphical representation of program and system features through icons, pull-down

menus, and scroll bars. Technically speaking, to qualify as a GUI, an interface must run in the computer's graphics mode; a number of packages commonly referred to as GUIs run in character mode only. Clearly, no strict GUI definition exists.

## Benefits of Using a GUI

### GUIs Save Money

First and foremost, GUIs save money. Several Datapro subscribers with whom we spoke have mixed IBM and Macintosh environments to determine whether their users find any advantage in a graphical environment. Many of them found the graphical environment faster to learn, and as a result, they get work done quicker.

### Ease of Training

Among the proven advantages of a graphical (as opposed to command-line) interface is ease of training. Why? In a graphical environment, most functions are the same from program to program. A user need only learn once how to print, how to copy files, or how to transfer data or graphics to another program.

### Ease of Use

Computer users range from programmers to novices, and commands are not always

---

—By Karen J. Offermann  
Associate Editor

intuitive or easy to learn. Some users find commands represented by icons (pictorial representations) easier to understand and remember. What could make more sense than pointing a mouse and clicking on a file's icon to open it up? The command line alternative could require the user to change directories, understand what file opens an application, and know the operating system's syntax (all without prompting) to do the same task.

### Tendency to Use More Applications

Because functions are identical across all applications, GUI users tend to use more applications than users who work at the command line. A GUI user who has learned one application then knows the basic functions of other applications that use the same interface.

### The Peat Marwick Study

A study conducted by Peat Marwick Main and Company in July 1987 focused on the benefits provided by the Macintosh environment to corporate end users. (It is important to note that the case studies used to reach these conclusions were taken from a sampling of major Apple accounts and do not claim statistical validity.)

Peat Marwick reached the following broad conclusions:

- The ease of use of the Macintosh appears to promote greater use of the computer;
- Gains in productivity, quality, efficiency, and effectiveness are reported by all levels of white collar workers;
- The quality of white collar products (specifically reports, correspondence, and budgets) is improved; and
- In most cases management has leveraged productivity gains from Macintosh into strategic and competitive advantages.

For more information, contact Peat Marwick at (415) 951-0100.

### GUI Market Trends

At this point, four GUI standards influence microcomputer interfaces—Microsoft Windows 3.0, the Apple Macintosh interface, IBM's SAA and Presentation Manager, and the UNIX-based X Windows standard.

#### The Windows 3.0 Revolution

While the initial versions of Microsoft Windows were met with disinterest, the release of Windows 3.0 in May of 1990 has changed the face of personal computing. Corporate users, although slow to commit to the OS/2 operating system, are displaying a favorable response to Windows 3.0. The reason for the easy acceptance of Windows 3.0 seems straightforward: users can run their existing DOS programs in a graphical environment with up to 16M bytes of memory per application (on an 80386 PC).

#### Macintosh: First and Purest

The Macintosh user interface was the first commercially available GUI. Its unique look and feel has never been successfully replicated, partly because it was designed from the ground up to be a GUI-based system but primarily because Apple has demonstrated a litigious method of guarding the technology. While the company has succeeded in protecting its interface, Apple may have lost an opportunity to set a GUI *de facto* standard.

### Even IBM Is Doing It

In 1987, IBM announced the OS/2 Presentation Manager and the Systems Applications Architecture (SAA). These developments bring a GUI environment to the IBM world and standardize the user interface across the entire IBM product line. Although IBM was late in entering the GUI market, it has been instrumental in defining the standards. The SAA standard gained huge success among graphical user interfaces; nearly a third of all current GUIs provide some level of SAA compliance. IBM and Microsoft have worked together to develop Presentation Manager and OS/2, through which they hope to offer the definitive user interface and operating system for the nineties.

Although Microsoft, Interactive Systems, and The Santa Cruz Operation are working together to provide Presentation Manager for UNIX, IBM has licensed the rights to use NeXTStep, the GUI that provides the front end for Steve Jobs' NeXT computer. As long as multiple UNIX standards exist, IBM may just be covering all its bases with the NeXTStep license, and we believe it is too early to rule out the possibility that IBM will offer PM for UNIX as well.

### . . . and Then There Is UNIX

Multiple and incompatible standards characterize the UNIX market, and UNIX GUIs are affected by such fragmentation. Today, both vendors and the standards bodies market competing UNIX GUIs, and many products are still in development.

Several UNIX GUIs are based on the X Windows windowing environment developed at MIT in 1984. Although X Windows is not a GUI and does not actually define a windowing style, it does provide a library of tested routines that allow a programmer to rely on certain standard methods while developing a GUI.

### Sun and Open Look

Sun Microsystems' NeWS (Network Extensible Windowing System) is a display manager for Sun's engineering workstations. NeWS allows use of the PostScript video display language in a network environment. Licensed to over 90 software developers, NeWS software will be available for UNIX and OS/2.

Sun and AT&T have combined forces to create Open Look, supporting both Berkeley and AT&T UNIX versions. The same network features in NeWS are found in Open Look, and functional routines are similar across all applications. AT&T's System/V Version 4.0 will include Open Look as a standard operating environment. To promote Open Look, Sun has provided a developers' toolkit for converting existing applications.

### OSF Motif

The Open Software Foundation has pledged to standardize the UNIX operating system, using IBM's AIX as a development platform. To ensure that all OSF applications operate similarly, the OSF has chosen a user interface standard—*OSF/Motif*. Motif combines the look and feel of Presentation Manager with an X Windows application program interface.

### Not X-11

Microsoft, Hewlett-Packard, and The Santa Cruz Operation have combined to offer Presentation Manager for UNIX (PM/X). PM/X uses the same applications program interface and graphics programming interface as OS/2. Any OS/2 C program which makes only display calls to



OS/2 will be portable to the PM/X environment. PM/X is targeted for portability to a number of environments and will not depend solely on the Intel family of microprocessors.

NeXTStep from NeXT includes Display PostScript, and its icons give a high resolution, three-dimensional appearance. NextStep has been licensed to IBM and is running on IBM's RISC AIX RT PCs.

### Beyond GUIs

Some vendors' products offer a graphical interface and other features that surpass the ordinary requirements of a GUI. Hewlett-Packard's NewWave and Wang's Freestyle are two examples.

NewWave's uniqueness stems from its reliance on Microsoft Windows to provide a windowing environment and extends to provide object management services. The object management services allow users to integrate data from multiple programs and to bind applications and data into objects. NewWave 3.0 provides a WYSIWYG word processor, computer-based training (CBT), and a task automation mechanism utility called NewWave Agent.

Wang Freestyle combines software and dedicated hardware to provide an easy-to-use, elegant interface. Priced at \$1,995, Freestyle includes the software, interface card, and a tablet and pen (which take the place of a mouse). Freestyle permits integration of different data types in the same document. The basic Freestyle icon is a page, actually an object containing both data and the code to manipulate them. Freestyle permits the integration of voice and data via a voice interface and handset which sells for \$1,495.

### DOS GUIs

The DOS operating system, with its command line interface and huge installed base, cries out for a friendly front end. Microsoft developed Windows to provide a graphical user interface for the DOS environment. Other vendors realized that the limitations and slow acceptance of the first two releases of Windows provided an opportunity to compete. As a result, users can choose from among several DOS GUIs. But, the release of Windows 3.0 in May 1990 has altered market perception of Microsoft's offering, and all indications are that the industry has accepted Windows 3.0 as a standard.

### Hardware Required

A DOS-based graphical user interface has certain hardware requirements. Even character-based interfaces are best served on a platform designed to accommodate the bit-mapped graphical interfaces. The minimal GUI configuration should include a mouse or other pointing device, a video adapter, and a video display.

Graphics hardware has evolved to support the graphical interface, and emerging display adapter standards promise ever-increasing resolution. The CGA standard, with its 320-by-200 pixel resolution, was considered revolutionary when introduced in 1983, but now it seems primitive. The VGA's 640-by-480 resolution now rates as the minimum hardware for adequate graphics support. The "Super VGA" (SVGA) provides 800-by-600 resolution, and the IBM 8514A standard's minimum specification provides 1,024-by-768 pixel resolution.

Graphics co-processors off-load the management of graphics facilities, such as screen redrawing, from the CPU and prevent the system from loading down. They promote faster performance of graphics programs, permitting acceptable GUI performance on computers with a comparatively slow clock rate. Graphics co-processors are available from Intel, Hitachi, and Texas Instruments.

### Decision Points

The environment will determine the face a PC shows to the user and will affect every aspect of the user's interaction with the computer. For this reason, this choice should be made with the following considerations in mind.

The first issue to consider when choosing an environment is whether there is a need to process DOS programs and data. Most GUIs provide processing for existing DOS programs without requiring any modification to the DOS code. But in most cases, a DOS program must be written to take advantage of the unique properties of the environment. For example, dynamic data exchange, a highly desirable feature of Windows 3.0, is available only to programs whose developers have implemented DDE. ■



---

# An Overview of Utility Software

---

## Datapro Summary

Although no one thinks much about utility software, nearly everyone with a personal computer needs and uses it. Because these small and inexpensive programs can significantly increase both data integrity and user productivity, they pay for themselves several times over. Utility software is an important part of any personal computer system and should be acquired with an eye toward both current needs and future expansion. All too often, a utilities package is purchased only after a costly disaster. For example: the hard disk crashes during an electrical storm, and during the weeks it takes to re-create the data, users realize that a disk backup or format recovery package should be in place; or an unauthorized entrance into confidential files indicates to the system administrator the urgent need for passwords to protect sensitive data.

---

## Technology Basics

By recognizing the benefits of utility software, users can plan a computer installation that prevents costly damage and increases productivity. Everyone should be acquainted with utility software.

- When an entire group of employees uses a computer, yet sensitive information needs to be protected from indiscriminate access, a *data security* package will prevent unauthorized users from viewing confidential data.
- When a user is accustomed to working only with Lotus 1-2-3 and is familiar with the keyboard requirements of that program, a *keyboard enhancement* program will allow the office manager to create a common user keyboard across all applications.
- When a hard disk is corrupted during a power outage, a *disaster recovery* program will allow the user to restore the working environment.

---

—By Karen J. Offermann  
Associate Editor

- When a user downloads software from a bulletin board, a *vaccine* program may detect the presence of potentially destructive virus code and alert the user.
- An *operating system menu* will allow the user to view a graphical representation of the system files and applications.

Utilities programs meet all of these needs, and more. At a minimum, when installing a system, a user should consider file management, disk backup, and keyboard enhancement software. If users access electronic bulletin boards, a vaccine package may be in order. In other words, look at utilities packages when you are looking at applications packages. Your choice and prompt implementation of a utility package may prove to be an important service to system users.

## Disadvantages

As always, convenience comes at a price. Since most utilities are “memory resident,” that is, they run in the background at all times, the trade-off is that they use valuable RAM memory. While the RAM capacity of most systems is now great enough that utilities use is not critical, it must, nevertheless, be weighed against the cost of today’s

RAM chips. Notorious for their consumption of RAM memory are keyboard enhancers, desktop organizers, and DOS menus. For the most efficient performance, a user should know how much RAM the operating system, applications software, and utilities consume. To help you calculate these requirements, our survey results show the amount of RAM consumed by each loaded program.

## Types of Utilities

In general, we find eight utilities categories that are important to the average user—file management and file maintenance; operating system menus; vaccines; disk backup/disaster recovery; data encryption; keyboard enhancement; desktop organizers; and, our catchall, miscellaneous. Each category is defined by the most important features it supports, although some features are identical. File management packages tend to provide some features found in data encryption packages, and operating system menus may offer features common to file management packages. Our comparison columns point out this overlap when it occurs.

A detailed description of each category follows.

### File Management and File Maintenance

These programs manipulate files, unerase files, provide hard disk format recovery, compress data files, secure files through password protection, relocate files, rename files and directories, and encrypt data files.

Hard disk format recovery can restore an accidentally reformatted hard disk to its previous condition. Unerasing files is one of the handiest features provided by a file maintenance utility. The re-creation of an erroneously deleted data file can be time consuming, so utilities of this type are quite valuable. Some packages unerase a sector at a time, allowing the recovery of major portions of the damaged file.

### Operating System Menus

Operating system menus are primarily a DOS utility that has the sole purpose of replacing the command line and providing a visual front end to the operating system and applications. Changing from a command line to a menu often implements a *point and shoot* method to choose the proper command or application. Most menuing packages are mouse compatible.

In addition to their basic features, many menus provide logs of system usage, rudimentary file management, and data exchange. Others permit the user to create a custom menu, which only allows access to permitted applications. Menus that group applications of the same type can be created. Most packages allow the addition of new menu items (usually programs) and the deletion of those no longer required.

### Vaccines

Vaccines, an emerging category of utility programs, screen for a type of renegade software known as a virus. A virus is a program designed to modify or "infect" other programs, possibly inserting an evolved copy of itself into the other program's code. The function of a virus can range from a benign message on the screen that appears once and is never seen again to catastrophic system failure.

The term "virus" has been associated with these dangerous programs because they reproduce and spread like their biological counterparts. Viruses tend to modify system and program files.

The most widely publicized effect of viruses is the erasure of a hard disk; however, manipulation of data or symptoms that resemble intermittent system malfunctions can be more serious. Viruses can be transmitted either by using an infected disk on a computer system or through a telecommunications link, where infected programs are downloaded to a "clean" computer system.

In response to the threat viruses pose, many small software companies (and a few large ones) have developed programs designed to detect the presence of viruses, keep them from replicating, and notify the user so the offending program can be removed.

Most vaccines are still rudimentary. Software developers must chase a moving target, basically playing a catch-up game. While most products screen for known viruses, more invaders are likely. As viruses gain in sophistication, developers hope their vaccines will also become more adept at screening for invading code. While some vaccines require special hardware, such as an add-on board, most do not.

Antiviral programs use many methods to detect a virus' presence. Some of these are:

- Scanning of executable files. By scanning all of the disk's executable files, antiviral programs can detect known viruses by comparing filenames to those of known viruses.
- Text string searches. Known viruses can be detected by scanning all executable files for text strings found in previously detected viruses.
- Checksum comparison. Infection of executable files can be detected by calculating checksums of programs known to be "clean" and recalculating the checksums on a regular basis. Some antiviral programs recalculate checksums for selected programs each time the computer starts, while others perform the task on demand.
- Detecting writes to executable and system files. By write-protecting system and executable files, some antiviral programs can protect these files from erasure or modification; but just as important, most programs offering this feature will signal the user when writes to these files are attempted, indicating possible viral presence.
- Terminate-and-stay-resident (TSR) program check. Programs employing this feature allow the user to create a list of permitted applications. Unlisted programs attempting to stay resident warn the user that a virus may be present.

Since most antiviral programs cannot with certainty identify the presence of a virus, most programs present a warning on the computer's display, then allow the user to decide whether the threat is real.

### Disk Backup/Disaster Recovery

Both disk backup and disaster recovery programs provide protection for data and applications stored on a hard disk. Both program types share some common features, and their differences are not always clearly defined. Primarily, a backup utility provides security by backing up data and program files, and disaster recovery software usually takes a snapshot of the system, writing to two hard disks at once or saving to videotape in realtime.

Disk backup tends to be incremental and uses a diskette or streaming tape as the storage medium for backed up data. Some programs will back up to another hard disk.

Disk backup programs often contain subroutines that estimate the number of diskettes required, format the diskettes, and compress files. Disaster recovery programs often back up data to a second hard disk or to videotape.

### Data Encryption

Data encryption programs take data files and, using an encryption algorithm, encode the data so that it cannot be read without reversing the encoding process. Other features include password protection, creation of hidden virtual drives (RAM disks), and the creation of hidden virtual drive subdirectories and executable files.

There are several encryption algorithms in common use, including DES, XOR, and RSA. Many programs use a proprietary algorithm.

### Keyboard Enhancement

Keyboard enhancers increase the computer keyboard's functional capability. These programs are RAM-resident and consume some RAM for macro processing, but they trade off the use of RAM storage for ease of use. Features range from keystroke redefinition to multiple keystroke commands. Also known as keyboard macro processors, these programs translate single keystrokes into multiple command strings. (*Macro* means that one instruction represents many instructions.) Any frequently used character string can be redefined as a macro, reducing keystroke repetition.

The macros created with these programs can contain complex logic statements and allow a character string to be redefined as a number of keystrokes. Each application can have its own macros to allow identical keyboard use across a broad base of applications, providing rudimentary applications program interface (API). An API allows consistent keyboard control across a variety of unlike applications.

The timesaving features provided in keyboard enhancers, combined with additional features such as file encryption, make them attractive to end users.

### Desktop Organizers

These RAM-resident programs are time-savers that help users organize their activities. They frequently contain an editor bundled with *emulators* for calculators, notepads, phone directories, calendars, and appointment books. Most desktop organizers can be accessed from within other programs. For example, the user can call up an appointment book from within a word processing program, check the date and time of a meeting, and return to the program. Some desktop organizers allow "cut and paste" between applications, so the results can be placed directly into the running program.

Desktop organizers can help schedule appointments, dial the phone, and keep track of phone numbers and addresses. They also can help manipulate data, sort mailing lists, and even provide rudimentary word processing, but their chief benefit is saving the user time.

### Miscellaneous

These programs tend to cross over the defined boundaries of other utilities packages. Most provide multiple features, while others represent categories too small to warrant individual attention in this report. Among the miscellaneous programs are thesauri, spelling checkers, and printer utilities.

In the past, single-purpose utility packages were common. The trend now is to provide multiple, complementary functions in the same package. Many utility packages

on the market offer features that go beyond their primary purpose.

## Things to Consider

Utilities, because they supplement the computer's operating system capabilities, are vulnerable to periodic updates made to the operating system software. Many DOS environment and utility packages, for example, use undocumented MS-DOS system calls or make assumptions about the file allocation table (FAT) found on the computer's diskettes and disks. The introduction of MS-DOS 4.0 eliminated many of these undocumented system calls and changed the file allocation tables, forcing many utility vendors to modify their code in order to run successfully under the new version of MS-DOS. Utilities that use undocumented system calls or go directly through interim vectors may crash under MS-DOS 4.0.

## Selection Guidelines

Utility software should be an integral part of any PC system. Ideally, planning for utility needs should be done when planning the initial installation. In fact, users are advised to consider their systems in light of any "worst case" possibilities.

- If you have a hard disk, then a file recovery program is advised. Any loss of power can result in a corresponding data loss. A file recovery program will attempt to reconstruct lost or damaged files, after the fact, using the fragments that result from some disaster.
- For the same reason, hard disk users must seriously consider disk backup and disaster recovery programs. A disk backup/disaster recovery utility will prepare for the possibility of future damage, and replicate the data contained on a hard disk by copying it to tape, diskette, or another hard disk in realtime. This approach is recommended over use of a file recovery program alone.
- In this day of electronic bulletin boards, a vaccine program for the purpose of detecting software viruses is a basic requirement. Any user who downloads freeware, program fragments, bug fixes, and other code risks downloading destructive code at the same time. By routinely screening any incoming code, a vaccine program can protect both program and data files.
- In an ideal environment, no user would deliberately view or manipulate programs or files that contain confidential information. In the real world, protecting the confidentiality of sensitive information is a legitimate concern. Proprietary code, payroll records, letters, product suggestions, and other documents may require protection from indiscriminate access. A data encryption utility will encode the software, preventing programs from being read or run, and prevent users from accessing documents containing sensitive information.
- *Make mine multipurpose.* Many utility programs take the "Swiss Army Knife" approach, providing a number of features in one program. Most often, these programs will provide file management features, a disk optimizer, file recovery, a menu interface, disk backup, and unerase features. The advantage of this approach is that users are afforded a number of features at one price. Among these multipurpose programs are: the Norton Utilities (now also available for the Macintosh), PC Tools Deluxe, and PC Tools. ■



# An Overview of Language Compilers and Interpreters

## In this report:

Compilers versus Interpreters.....	2
Cautions and Considerations .....	4

## Datapro Summary

The language compiler market has changed rapidly in the last two years. With the introduction of OS/2 and the renewed interest in UNIX, compilers reflect the power of these new system platforms. Object-oriented languages are becoming commonplace. In the past, language compilers were largely the province of experienced programmers; end users relied on off-the-shelf applications. As the general PC community grows more sophisticated, however, many end users are experimenting with coding their own applications, tailoring them to fit personal needs. As a result, more and more compilers offer features—such as interactive debuggers—that appeal specifically to end users. The mystique of the hacker has faded; programming is becoming a common skill.

## Technology Background

The products covered in this report fall into several categories. *Quick* compilers generally are geared for the novice or casual programmer and feature a menu-driven user interface, automated compilation, and assembly and link and assume little programming experience. The professional or development systems, on the other hand, have a level of sophistication that is rare in low-end compilers.

Today's significant issues in mass-market compilers are object-oriented extensions, interactive compiling, on-screen help, and fast compile times. Object-oriented extensions to conventional languages provide the programmer with features such as the ability to subclass. They also allow for inheritance, multiple inheritance, data abstraction, data hiding, and a choice of linking methods (dynamic or static).

The movement to low-cost software began with Borland's Turbo Pascal about six years ago. The rage for low-cost, easy compilers has since swept the market for Basic, C, and even Prolog. Fortran and Cobol, long the exclusive territory of professionals and hard-core hackers, are now making

concessions to the ease of use offered by more interactive languages.

Many vendors have found that low-cost compilers are an excellent vehicle for marketing high-end compilers. More and more low-end products are offering compatibility with high-end compilers, as seen in the relationship between Microsoft Quick C and Microsoft C 6.0. Microsoft, long known for its professional languages, has moved into the low-cost compiler market with its Quick Basic (QB) compilers. Microsoft has pledged to expand its quick-language family and will, no doubt, include the QB productivity aids in the scheduled upgrade of Quick C. There is also an increasing emphasis on optimizing the code for fast execution.

Borland International, first in the market with quick compilers, has added object-oriented and professional Pascal compilers and a professional C compiler along with an interactive assembler and debugger positioned to compete with Microsoft's CodeView.

## Programming: Not Just for Hackers Anymore

The first commercially available programming languages were geared to the mainframe and minicomputer environments.

For years after the introduction of microcomputers, the industry focused on applications, with languages remaining the developers' province. Early microcomputer-based language compilers and interpreters offered limited features and were difficult to use, necessitating a formal education in programming skills. Although rarely stated explicitly, there was a certain snob appeal in being a member of the programming community—the more difficult a language was to use, the more exclusive the fraternity. For years, the mystique lingered.

The first microcomputers were intimidating to the uninitiated. Even now, computer phobia is a reality for many novices. As a result, vendors' early attempts to sell microcomputers concentrated on packaged applications that required little or no user programming experience.

Now, compilers are a hot topic in the general user community. More and more users want specialized routines or functions that are not available in off-the-shelf software. Some tasks are so user-specific that vendors cannot profit by developing a program to handle them. The clear alternative is for end users to write customized routines. The software community response is programming tools that range from inexpensive low-end compilers and interpreters to full-blown professional development packages. Interactive, menu-driven language compilers, unheard of only a few years ago, are now commonplace.

Borland International led the software developers in ease of use, interactive debugging, and reasonable prices with the introduction of Turbo Pascal in 1984, followed by the release of Turbo C and Turbo Basic. Other additions to the Borland line include Prolog (an artificial intelligence language). The interactive features of the new wave of compilers demystified the coding process and made programming languages available to a wider group of users.

Portions of many vendors' high-end compilers, geared to professional software developers, are now in a version that offers a subset of the features at a much lower price. These low-end products are generally easy to use—offering tutorials and menu-driven interfaces—but they are often upwardly compatible with the high-end product, allowing full use of both source and executable code. Microsoft is competing with Borland by offering "Quick" versions of its conventional compilers. Microsoft's Quick C, for example, is upwardly compatible with Microsoft's Professional C compiler.

Finally, as the introduction of new operating systems continues, users can choose from a larger variety of languages for even the more arcane operating environments. Recognition that an operating system platform cannot become an industry standard without a broad variety of available tools has sent many operating systems software developers (e.g., IBM and Microsoft) scurrying to develop, acquire, or cooperate in the development of programming languages and tools for operating systems such as OS/2, Xenix, UNIX, and AIX.

## Programming Languages: A Definition

A programming language, generally speaking, is any system of logic and notation, implemented on a computer, for the processing of algorithms and data. A computer's central processing unit (CPU) understands binary code, and high-level languages provide a bridge from natural-language constructs to the binary code required by the CPU.

Binary (or machine) code consists of a series of zeros and ones. Programming in binary code is difficult and time consuming. Debugging binary code is incredibly complex,

as the generated code is huge. A simple program to print the words "Hello, world" to the screen would take about 100 lines of binary code. The same program, in the C programming language, takes four lines of easily understood code.

The first computer programs were written in machine code, composed purely of binary values, or in assembly language, a mnemonic code only one step above binary machine code. Although machine and assembly languages are still used (mostly for speeding program execution), the process is tedious, the source code is unwieldy, and debugging is time-consuming.

Basic, C, Cobol, Fortran, and Pascal are among the most widely used high-level programming languages. The survey also covers a variety of miscellaneous languages, including Forth and Modula2. Today's programming languages have their roots in the early days of electronic data processing. The languages evolved in power and sophistication along with the computers.

## Formal Language Standards

Throughout this report, mention is made of "formal language standards." Language standards are those conventions of syntax and semantics that combine to form programs in the subject language. There are a variety of standards for programming languages, and not all standards apply to all languages.

### BNF

One of the oldest formal language standards, which is still commonly used, is the *Backus-Naur Form (BNF)*. This notational method specifies a "generative grammar" and defines the set of all strings of symbols, and their syntax, which is used to create programs in a subject language. The BNF grammar offers a set of rules, presented with a left side and a right side split by the *metasymbol* "::="". The left side contains a *nonterminal* symbol that contains information about the construct or syntactic type of the subject language. This *nonterminal* symbol is a string of one or more characters enclosed by "<>". The "::=" symbol reads *consists of* or *is defined as*. The example below illustrates this principle.

```
<series> ::= <statement> | <statement>
           ; <series>
```

Other language standards widely recognized and accepted throughout the industry are the American National Standards Institute (ANSI) standard; the ISO standard; the GSA standard, taken from the General Services Administration of the United States Federal Government; the Kernighan-Ritchie (K&R) standard (C); the Jensen-Wirth standard (Pascal); and the University of California-San Diego (UCSD) standard (also Pascal).

## Compilers versus Interpreters

Each programming language uses different constructs and procedures for identifying data and different formulas for manipulating that data. "Source code" is, basically, a program written following the rules of structure and syntax necessary for the language. Since the computer's CPU cannot directly execute the source code produced by a high-level language, some form of translation into machine language is necessary. There are two forms of translation available for this purpose—compilers and interpreters.



## Compilers

A compiler takes the high-level language source code as input and manipulates it into machine language. This code is appended with a suffix indicating the language it represents. In the case of Basic, the suffix is *.bas*, for C—*.c*, for Fortran—*.for*, for Cobol—*.cbl*, and for Pascal—*.pas*.

The process of conversion takes several steps. Generally, the code is manipulated in this fashion. Assume a source file called *examples.pas*. Note that not all compilers or interpreters follow all these steps and that some of these steps may be invisible to the user. Some compilers and interpreters provide an automated MAKE or other command sequence that executes all steps in this sequence, masking the procedure with a batch file.

- The compiler is invoked with the command *pascal examples (options)*. Calling the compiler takes the high-level language code and converts it into an intermediate language. This is the first step in the conversion to machine code and results in a file with the *.il* suffix.
- Next, the *.il* file is optimized with the command *opt examples*. This results in a file with an *.opt.il* suffix. Optimized code is more compact and, therefore, more efficiently stored and executed.
- A translation step may follow, using the command *transil averages*. This step results in the *.asm* file.
- The code is then assembled with the command *asm example*. The result of the assembly process is a relocatable file with the suffix *.rel*. Generally, relocatable code may be moved from one system to another, as it is processor independent and does not include absolute addresses.
- Finally, the relocatable code is linked with the libraries and routines used by the program (for graphics, disk I/O, etc.). The command to link is *link examples*. The linker is a utility that pulls in all related routines and library calls necessary for the successful compilation and execution of the source code. The link file must include all external routines and libraries required by the program, or the attempt to link will result in *unresolved externals*. The resulting code is now in executable form and results in a processor-specific file (*.68000*, etc.) and the executable file (which is reflected by the *.exe* suffix).

**Executable** code is identified with the *.exe* suffix and, as the name implies, is the code that actually causes the computer to run the program. Once the code is in executable form, it is no longer necessary to invoke the compiler in order to invoke the code. With all steps necessary to manipulate the source code in our sample session completed, the program can be started by typing the word *examples*.

## Interpreters

An interpreter takes the high-level language source code and data as input, generally a line at a time, and translates it directly into machine code, which is then executed immediately. This translated code can be executed more easily than the source code but cannot use the hardware interpreter for program execution. This means of language processing is less efficient than a compiled version, since the program is translating and executing a line at a time. Optimization is impossible, and each call to a statement or routine is translated as many times as it appears in the program. This causes interpreters to execute more slowly than a compiled version of the program in the same subject

language. In contrast to the **executable** code file produced by a compiler, the end result of an interpreter's processing is **executed** code.

The Basic programming language was traditionally offered as an interpreted language.

The sequence of events in the interpreter's operation differs from that of the compiler. The interpreter fetches the instruction found in the source code, decodes it into machine (or binary) code, fetches the called arguments, and calls the designated primitive operation. Once the primitive operation is executed, the interpreter repeats the sequence until the program is finished. In a loop, where the same instruction is repeated over and over, the interpreter reads in the same instruction as many times as necessary for execution, repeating all steps of the interpretation sequence for each occurrence of an instruction.

## Performance Trade-Off

The performance trade-off between compilers and interpreters is a factor in choosing the proper language processing program. The compiler translates each instruction only once and is, therefore, generally faster. Because the compiler stores the object code until execution halts, it requires more disk storage space. An interpreter is considerably slower because it analyzes a statement each time it is called, but it creates no object code and requires less storage space.

## Choice of a Programming Language

Naturally, a new programmer wants to know which language is best suited for the job at hand. The following items discuss some of the issues to consider when choosing a high-level programming language.

### Simplicity of Language Conventions

It is essential to factor in the simplicity of the actual language conventions—does the language assist the programmer in conceptualizing the problem before the actual coding?

### Simplicity of Syntax

Language syntax greatly affects the initial success of the coding process and later affects any modification of the programs. In addition to avoiding misleading or confusing syntax, a programmer should select a language that allows anyone reading the program to follow the thread of logical execution through the source code. For example, Fortran uses *GOTO* statements, making the initial algorithm often difficult to decipher from the resulting code. Tracking the *GOTO* statement can be confusing when it moves the thread of execution to a completely different area of the program. On the other hand, Pascal code denies any use of *GOTO* statements, allowing nested routines which reflect more accurately the algorithm beneath the code, since the thread of execution is more clearly reflected in the code and lends itself to more concise program planning.

### Abstract Support

Support for user-defined and abstract data types should be as wide as possible in order to offer the greatest performance from a high-level language.

### Tools Provided

In a situation where a certain compiler or interpreter includes an assembler, editor, debugger, menu interface, or

other feature providing ease of use or an enhanced operating environment, the user may opt for the more self-contained product. The facilities such as assembler, editor, and debugger, which may be provided with the programming language, may be enough of a consideration to make a lesser performer the better choice for a job.

#### **Portability of Source and Object Code**

Portability of both source and object code should be considered when determining which programming language product to choose. A language specifically designed for one computer system may not be usable with others. A good rule of thumb is to look for a definition or standard that is hardware independent to the greatest degree possible.

#### **Purchase and Development Costs**

The cost factor is measured in two ways—the actual cost of the software and the more subjective cost of developing and maintaining applications code.

---

### **Cautions and Considerations**

Perhaps the greatest mistake in the microcomputer industry has been the tendency to think small. Although the high-performance products may seem daunting, plan before making a selection. Clearly understand the types of records to be maintained and the processing to be performed before deciding on a programming language or purchasing a compiler or interpreter. Look at the size of the source and data files a compiler can successfully process. If processing needs are growing, do not commit to a package with upper limits that will soon be exceeded.

Speed is a concern in many mathematical operations. Some graphics and computer-aided design (CAD) applications require complex, lengthy calculations that may take excessive CPU time. Timed results for processing industry-standard benchmarks—such as the sieve of Eratosthenes (a classic benchmark to test a language compiler's ability to create fast, compact executable code)—are a good indicator of processing performance in number-crunching programs and provide a reasonable basis for comparison from compiler to compiler.

Consider vendor support and experience. A novice would be hard pressed to deal with a sophisticated product from an obscure vendor. Many of the new interactive programming languages provide tutorials, context-sensitive help, and telephone hot line support at no additional cost. Generally, help or additional information is easily acquired if a popular product has been available long enough to inspire book and article publications for the general marketplace.

---

### **Future**

At the risk of redundancy, the most significant trend in programming compilers today is the introduction of object-oriented languages. The combination of the object-oriented tools and paradigm provides developers with a means to better control the complexity of today's programming environments while reducing development time and costs. ■

# An Overview of Electronic Mail

## In this report:

Platform Options .....	2
Selection Criteria .....	8

## Datapro Summary

Electronic mail, once a peripheral communications technology with limited uses, has moved to the center of many office automation strategies. Spurred by the growth of desktop computing, data communications, and application software standardization, E-Mail has shifted from a standalone resource used by small groups to an important technology for information processing, resource sharing, and group coordination.

## Technology Overview

E-Mail is transforming office communications. The growth of standard interchange methods such as the X.400 standard promises to eliminate the barriers to message interchange erected by proprietary vendor strategies. New applications are making E-Mail a crucial technology for information resource sharing.

## The New Face of E-Mail

No longer considered simply an alternative to conventional mail services, E-Mail has become an entirely new way of sharing documents, graphics, and applications. Using E-Mail and applications integrated with it, office workers can adopt a new style of direct communications with one another. Managers gain extraordinary tools for referencing and recalling information and monitoring subordinates' communications. These developments point toward increased personal and departmental productivity.

## What Is Electronic Mail?

The term electronic mail encompasses various technologies that support electronic transmission of text and graphics. E-Mail technology, in its broadest sense, encompasses all types of electronic messaging, including facsimile, telex, mailgrams, and voice. In recent years, however, E-Mail has increasingly been used more narrowly to

describe messages transmitted between users of networked computers.

According to the Electronic Mail Association, electronic mail is the generic name for "non-interactive communication of text, data, image, or voice messages between a sender and designated recipients by systems using telecommunications links." Information deliverable via E-Mail can generally also be sent by such conventional means as:

- The public postal service,
- A special courier service (e.g., Federal Express, Purolator),
- Inter- or intracompany mail, and
- Telephone voice messages.

The decision whether to use E-Mail or one of these conventional means turns primarily on the *time factor*. A message's priority determines how much time there is to deliver it. Priority, in turn, restricts the possible methods for getting the material to its intended location on time when using conventional services. It is in this area—delivering messages rapidly while preserving the option of hard copy—that E-Mail offers its greatest advantages.

Increasingly, E-Mail is hailed as an important "enabling technology" for office communications. By combining electronic messaging systems with other applications, information workers obtain tools to disseminate information for comment

quickly, coordinate meetings and scheduling, hold meetings on-line electronically, and design applications that use messaging features to automate such office processes as requisitioning supplies.

As E-Mail becomes increasingly integrated with a host of other office automation tools, it becomes easier and more cost effective to add value to information quickly and route it in any form desired. In organizations with a strategic vision for managing information resources, E-Mail can become a platform for swift, value-added information processing involving many individuals.

## Electronic Mail Technology

E-Mail is not a technology in and of itself. It combines software and hardware invoked at a given place and time for communication. E-Mail software, which provides the ability to transmit and receive messages, is useless without the necessary hardware to create, receive, and transmit messages. Thus, it is difficult to isolate software and hardware from the overall perspective of message creation, transmission, and reception.

### Basic Technology

Regardless of operating environment, E-Mail systems share some basic technologies that support message creation, transmission, and reception. The individual E-Mail applications provide these technologies. The network that supports this application carries out the communication functions. E-Mail systems do not necessarily use all of the following technologies, but most full-featured mail systems include provisions for each of them.

**Mailboxes.** Most currently available E-Mail systems provide user mailboxes for message reception. A mailbox is a work space provided by the E-Mail application where users manage correspondence. The functions for managing the mailbox range from a simple read function through menu-driven cutting and pasting of stored messages to formulate responses.

**Editors.** Most E-Mail systems today provide at least a simple text editor for message creation. Not all systems are integrated with an editor, however. Some systems still require users to create a text message separately before using E-Mail to transmit it.

**Store and Forward.** E-Mail systems based on this technology send messages to a central storage point or repository. The addressee is then notified of the pending message and reads the message from the central repository into his or her work space. Store-and-forward messaging eliminates the need for receiving parties to be connected to the system when mail is routed to them. The system stores the message and notifies the recipient when he or she logs on.

Most mail systems use store-and-forward technology. Only rudimentary messaging systems require direct connection between the sender and receiver, as in telephone communications.

**Directories.** Most E-Mail systems employ a user directory to simplify message addressing. The directory lists all accessible usernames or numbers and their logical address on the network. The user can call up the directory and choose the intended receiving party(ies) from the list. As mail systems become more widespread, a universal E-Mail address, similar to a telephone number—accessible from almost all points—has emerged as a goal. The CCITT's X.500 series of recommendations specify standards for such a universal directory.

**Distribution Lists.** These lists go hand-in-hand with directories. Distribution lists allow users to group recipients into lists and forward messages automatically to all users on a chosen list. Some systems include carbon copy and blind carbon copy facilities to automatically forward message copies to other involved parties.

**Bulletin Boards.** Another distribution technology, bulletin boards provide access to common communicating areas that allow each user to read the messages stored there. Bulletin boards provide a more interactive means for communicating among members of a group than do traditional, individually routed messaging methods.

**Message Headers.** Each E-Mail system uses its own method to identify outgoing and incoming messages. The header is the initial information that a recipient views about a given message. Some E-Mail systems give users control over a variety of available header contents. Others simply provide a time and date stamp. Some message headers indicate appropriate follow-up actions such as response requested; other headers generate a receipt record for receipt-requested messages.

**Message Storage.** Once a message is received, most E-Mail systems provide some facility for storing it for further reference. Storage may be automatic or require executing a command. Some systems also provide tools for managing stored messages. These range from simple acknowledgment functions to sophisticated text retrieval functions involving Boolean operators.

**Security.** The features that E-Mail systems provide for security vary greatly. Some systems make few security provisions; anyone could walk up to a terminal and read a user's mail. Other systems provide password access to the mail application. Still others support separate definitions of read and write access levels. The E-Mail system security features must be placed in the context of overall communications security, which may involve encrypting or scrambling message transmissions to thwart unauthorized access.

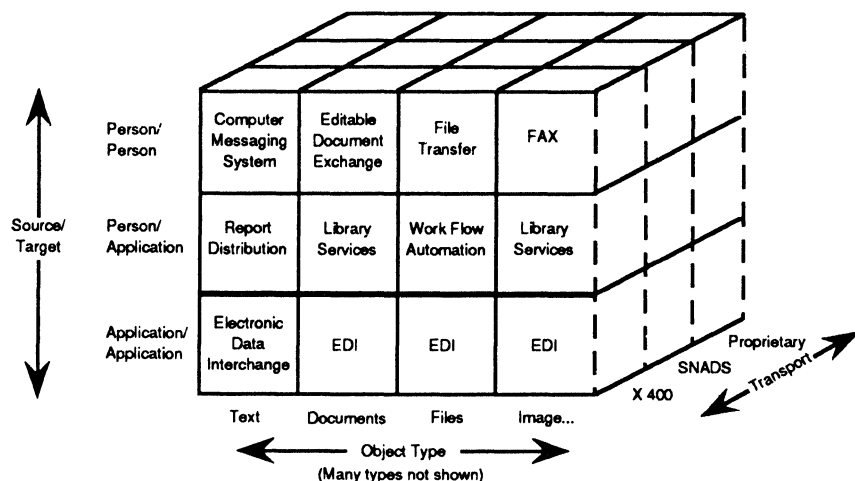
**Access to Outside Systems.** Optimally, an electronic mail system allows users to access all messages addressed to them and send mail to whomever they want. For users in large organizations, this implies that the local mail system must be connected to other systems to supply complete electronic access. Many packages now include gateways to various on-line services, facsimile, and popular Integrated Office Systems such as IBM's OfficeVision, DI-SOSS and PROFS and to Digital's ALL-IN-1.

## Platform Options

Most organizations with centralized or networked computing resources are implementing some form of E-Mail. There are two basic approaches to this task. The first is to upgrade existing computing equipment to handle text communications tasks. This usually entails implementing an integrated office suite that includes E-Mail capabilities. The second is to implement a complete turnkey system specifically designed to provide E-Mail functions. Many companies already using some form of computer messaging in-house have upgraded their systems to handle electronic mail.

### Integrated Office Systems

A special type of in-house system, the integrated office system, uses centralized and/or networked computing resources to provide a set of office applications. All these systems feature E-Mail, often as a cornerstone application



*Figure 1.  
E-Mail Architecture  
Schematic*

*Adequately specifying a complete E-Mail network requires multidimensional thinking. The questions of to-where, in what form, and by what resources all must be solved both independently and interdependently. The implications of solving this matrix constitute the foundation of an eventual, all-encompassing, worldwide E-Mail network.*

Source: Soft-Switch, Inc.

used to tie other applications together. The emergence of new, network-based office tools, such as IBM's recently announced OfficeVision, signals a new round of interest in these integrated tools.

Integrated office systems combine departmental computing technology with office software applications. In addition to E-Mail, these systems combine other support functions such as word and data processing, graphics/image processing, and voice processing. They provide a consistent user interface to all these facilities. This consistency implies an ease of use that makes them attractive for linking office workers.

The E-Mail component of these systems usually incorporates features that simulate business communications. Options include forms simulating memos such as "While You Were Out" slips, distribution lists, and carbon copies. Message receipts are easily logged. In addition, third-party and user-developed add-ons extend E-Mail functionality by combining the messaging functions with graphics or financial analysis tools to automate certain processes. Other systems support on-line conferencing that allows meetings to be held with the participants never leaving their offices. Many integrated office systems also combine electronic messaging with an interface to voice-based messaging.

**LAN-Based Systems**

The potential for running E-Mail applications on microcomputer local area networks (LANs) is becoming ever greater because of increasingly sophisticated hardware and software. Microcomputer LAN applications now include some of the best features of larger implementations. LAN packages increasingly used network architecture elements such as client/server computing (discussed below) to allow E-Mail networks to be implemented largely on microprocessor-based networks. This presents obvious advantages when compared to running resource intensive E-mail applications on host resources.

LAN-based systems run primarily as an application on local workstation. Increasingly these packages grow more sophisticated and run in conjunction with other resources such as an application run atop the network operating system as well as at the workstation. This structure allows LAN-based E-Mail more capability by providing resources for interconnecting with larger networks seamlessly. Again, taking these resource-intensive operations away

from host resources is proving increasingly desirable. Running E-Mail as a server process is an important innovation—one that will speed the growth of E-Mail networks.

Simultaneously, E-Mail is growing more closely integrated with microcomputer-based software applications. Rather than operate as a separate application, E-mail can be directly accessed by applications such as Word Processors and spreadsheets. By integrating E-Mail more closely with a user's operations, it becomes the prime interface for underlying communications resources—resources that grow more powerful by implementing the aforementioned server-based processes.

**System Components**

E-Mail technology depends on two fundamental hardware components: a terminal or user workstation and a communications network. Workstations can be either *smart*—containing a processor and capable of executing software (such as a microcomputer), or *dumb*—processorless and requiring connection to a smart resource such as a mainframe, minicomputer, or a supermicrocomputer where the mail software is executed. The results are then passed back to the dumb terminal for interactive display. The E-Mail software provides an application-level interface between the user and the workstation.

**Workstations**

Any device connected to a host resource or network can be considered an adequate workstation for running E-Mail. Terminals have been most common to date, but microcomputers are emerging as the most prevalent communications device available for electronic mail input and output. Microcomputers are used as terminals for connection to remote computing services and to in-house systems, and electronic mail is a vital application in the burgeoning market for microcomputer LAN applications. The concepts underlying the development of LAN mail software are on the cutting edge of the evolution of the electronic office. Mail software enables people using a departmental LAN to share documents, graphs, and even applications. The design of these applications draws on new understanding about the nature of communications in the information-age workplace.

Professionals with personal computers are most likely to take advantage of the benefits of electronic mail. Microcomputers are used throughout the workplace and run a variety of applications besides electronic mail. Today, microcomputers are networked, connected to mainframes, minis, and to each other in all conceivable arrangements, and they can be equipped to emulate most standard communicating terminals. Virtually all the major office automation vendors provide microcomputers that act both as standalone application processors and workstations for their mini- or mainframe-based integrated office software which includes the electronic mail application.

Dedicated terminals connect users to centralized, dedicated computing resources. These terminals usually depend entirely on the host computer for input and output. E-Mail is usually a menu item or a command entered locally at the workstation, processed and managed as a session on the host. Dedicated terminals often have special function keys that allow customized responses to applications; a key for "Read Messages" may be defined on the keyboard, for example.

Printers for creating hard copy are another vital part of the workstation picture. Other peripheral devices, such as modems to connect workstations to on-line services, image scanners, fax machines, and host gateways or LAN interfaces, are among the components that define the E-Mail workstation.

### Communications Network

The communications network that transmits messages is the largest variable in electronic mail. The physical medium and communications controller can be the public telephone system, a private or semiprivate data communications or telecommunications network, a hard-wired connection to a host mainframe, a collection of small networks in a wide area network (WAN) connected by a dedicated backbone, or a single LAN with its own wiring and connections. A system-level provision for communicating over the network is necessary. E-Mail software interfaces with this system-level communications control resource to transmit messages.

Hierarchical networks, which were the initial environment for electronic mail implementation, are now being supplemented with departmental systems based on local network technology. The key component in a hierarchical network is often the communications processor, a computer dedicated to the control of electronic data and text traffic in a company's data communications network. This communication processor generally handles the operations of the electronic mail application. In a distributed network, some of this controlling power may be decentralized, but the function of the communications processors is the same.

A communications processor comprises the processor itself, an interface to the host computer, a communications multiplexer for the control of incoming and outgoing information, interfaces to terminals and peripherals, storage facilities, and control software. The primary functions of a communications processor are line control, character and message assembly, data and protocol conversion, error control, message switching, and application-oriented functions programmed by the user. Communications processors can be limited to data transmission in a computer network or can be more general devices, such as digital PBXs, that support both voice and data traffic.

Communications processors or switches often employ store-and-forward message switching. In order to add electronic mail, the following considerations are important:

- Can the code conversion capabilities of the communications processor be upgraded to handle a wider variety of workstation types, as well as telex terminals and facsimile?
- Can the on-line storage capability of the host computer or intelligent workstations be upgraded and linked to the communications processor to provide additional electronic filing space for mailboxes and records?
- Perhaps most important is the issue of connectivity. Is the network's architecture capable of easily lend itself to integrating with a variety of other networks?

Network architecture is little appreciated beyond the level of system developers, but now everyone should understand *something* about it. Terms such as *open system* and *open network* are based on sound principles, and these principles deserve to become universal—the OSI formalism is a start. Be aware that *open* is a relative term to most vendors, however. In many cases this is partly by necessity, given the technical demands of designing and then implementing networks. Vendors do have an investment to protect—in their design work and programming and in assembling systems capable of carrying the network. Vendors in turn should understand that investing in standards *does* pay off.

A local area network serving the electronic mail application obviates the need for a communications processor. Several LANs of the same type can be connected via bridges or combined with other types of networks (foreign LAN, SNA, X.25, X.400) via gateways to create a larger area of service. Electronic mail software can be maintained and distributed from one system designated for this purpose. This system can range from a PC on a departmental LAN to a minicomputer on several bridged LANs to mainframe host systems on an SNA network. Of course, the larger the network, the more complicated it is to implement and maintain. One of the advantages of LAN processing is the inherent advantage of distributed systems—eliminating the possibility of total network failure.

### Messaging Architectures

The variety of network architectures employed in E-Mail networks is implemented on a variety of platforms. The need for a messaging architecture and the tools to implemented have evolved for over 20 years. Now the effort to forge an architectural standard by which all existing systems can interconnect is the delivery phase. All of the architectures mentioned are topics for a great deal of in-depth coverage. It would be wise to study the texts available. One that has been influential to our understanding is a vendor publication, but its inclusiveness as a managerial text for message systems architecture is worth noting. *Electronic Mail: Technology, Applications, and Infrastructure* from Soft-Switch provides an executive summary for E-Mail implementation, and a practical look at messaging network architecture and its implementation.

### X.400

The Consultative Committee on International Telephony and Telegraphy (CCITT) formally approved its Recommendation X.400 for Message Handling Systems (MHS) in 1984. The goal of the recommendation was to specify a set of standards that users and vendors alike could adopt

and thereby ensure global compatibility for electronic mail and other message-oriented information exchanges. Among its desired effects is a uniformity in communications protocols that would break down barriers imposed by varying standards used by different software vendors.

Most major computer system vendors have already announced or demonstrated basic high-level interconnectivity based on X.400 protocols. Even IBM has embraced X.400 to enable its large installed base to achieve a degree of connectivity with non-IBM systems. In addition, a growing number of domestic network facilities vendors have added X.400 interfaces to their set of basic product offerings in anticipation of new user demand. Together, these developments have provided network managers and integration specialists with enough functionality to construct private, multivendor networks for electronic mail exchange.

For devices to communicate, they must be compatible on various levels. These levels are outlined in the International Organization for Standardization (ISO) Open Systems Interconnection (OSI) reference model for data communications, which consists of a seven-layer hierarchy that defines the electrical characteristics, communications standards, and software applications for computer systems. This model does not define a single system and should not be regarded as a specification for any data communications system. Rather, it is a reference point for the establishment of a data communications system.

Beyond X.400 is the OSI recommendation for directory services (DS), commonly known as CCITT X.500. X.500 specifies an online directory for message communications, ultimately allowing network providers to map a common, interconnected directory of worldwide users. X.500 dictates naming conventions, how users access directory information, and what services are available. The 1988 standard faces several obstacles to formal ratification, however, including network security concerns. A new version is being developed and may be adopted in 1992; the vision of a worldwide messaging directory probably will not be realized until the late 1990s.

### SNADS

As IBM's proprietary messaging protocol, the SNA Distribution Services occupies a key role in a discussion of messaging protocols. SNADS is a set of Distribution Transaction Programs that run in an APPC LU6.2 environment. It provides asynchronous (store-and-forward) communications. Although compatible with IBM's Document Interchange Architecture (DIA), SNADS is becoming the preferred method for distributing documents. The primary application for SNADS is electronic mail, making it IBM's proprietary equivalent to the CCITT's X.400 protocol. Yet as an equivalent to X.400, SNADS betrays IBM's traditionally proprietary stance, and is generally a target protocol for E-Mail software only as an acknowledgment of IBM's installed base. In fact IBM's gradual evolution toward ISO standards for its network strategy spells either eventual obsolescence for SNADS, or a likely overhaul to bring it into some degree of ISO compliance.

### TCP/IP

Transmission Control Protocol/Internet Protocol (TCP/IP) is an abbreviated way to refer to the Department of Defense standard Internet architecture. These protocols implement the transport and network protocol layers, respectively, of the OSI model, and are the central part of several networks used in the military and in many research

institutions, universities and manufacturers. TCP/IP is the method used by the UNIX operating systems for message handling and processing. Although it is a standard component of the otherwise proprietary UNIX system, TCP/IP's acceptance as part of the U.S. Federal government's overall communication and computing strategy make it an important protocol for the foreseeable future.

TCP/IP is well-established in a variety of computing environments. It is tested, reliable, and relatively low-cost. This is in contrast to X.400 which is a relatively new technology, one where observance of the *de facto* standard is often overshadowed by the relative immaturity of vendor offerings. Competition from TCP/IP is frequently cited as one of the major roadblocks for widespread OSI acceptance. The tension in this market has yet to be satisfactorily resolved. In the interim, committing to TCP/IP allows users access to a reliable and low-cost method to develop intra-enterprise networks. Ultimately, TCP/IP's inability to serve as a true *interenterprise* protocol limits it, and it is in this case where X.400 is clearly necessary. Now, however, the impetus for widespread X.400 implementation is insufficient to remove focus on TCP/IP.

### MAILbus

In September 1987, Digital Equipment Corporation announced a number of advances for DECnet and its overall communications strategy. Since then, Digital has maintained its pioneering stance in the integration of standards-based protocols into its communication architecture. Among these were Phase V of Digital Networking Architecture (DNA) and the MAILbus suite of communications software products for linking Digital systems in multivendor application networks.

The MAILbus products, which include the Message Router X.400 Gateway (MRX), comprise a logical grouping of new and existing communications software. The MAILbus concept offers a modular approach for adding gateways to major environments, including IBM, OSI, VMSmail, and UNIX (VAX ULTRIX Mail Plus).

### AT MHS

Action Technology's Message Handling System (MHS) has achieved near-standard status in the LAN-based messaging market as a result of Novell's adoption of it as the messaging protocol for its market-leading NetWare LAN operating system. MHS is defined as a platform for the development and operation of groupware products, integrated Material Requirements Planning (MRP) in manufacturing, and other communications applications. MHS provides standard X.400 gateways, a centralized store-and-forward architecture, application program interfaces (APIs) for building on the MHS services, standardized addressing conventions, and communicating bridges (e.g., to SNADS and MCI Mail).

Designed around a hub architecture, MHS was designed primarily for remote access to the network and will find its use in a variety of applications. MHS is a solution for system integrators connecting applications among remote sites, for example. It spares the expense of an X.400 implementation when interconnection among different types of systems is unnecessary. MHS gateways are widely available, particularly in the swelling ranks of LAN-based systems. Novell has announced plans to develop its own messaging gateway, one that maintains backward compatibility with MHS. For its part Action Technology is committed to maintaining MHS: the product fits its market niche fits well.

## Advanced Features and Applications

In addition to the basic set of technologies that facilitate electronic messaging, the emergence of advanced features has paralleled the growing sophistication of all application software. The twin trends toward increased ease of use and increasing numbers of added features have transformed E-Mail packages into sophisticated applications. Also, the demands of the emerging office of tomorrow have produced new classes of software to organize and manage office communications. This has further E-Mail's applicability as a modern office technology.

**Application Programming Interfaces (APIs).** APIs provide tools for extending E-Mail software by accessing the underlying communications and filing technology employed by the package. "Open" APIs imply that a package can be extended either by sophisticated end users capable of programming, or by third-party developers, who can use the E-Mail package as a platform for further development. APIs have especially important implications for building bridges among heterogeneous networks, providing the same application functionality for a package across multiple platforms.

**Attachments.** Another popular E-Mail feature allows users to attach files to messages or otherwise include information available at the user's workstation in a message. Advanced systems allow inclusion of binary (program) files, and others allow attachment of graphics to messages which can then be viewed at the recipient's workstation.

**Bulletin Boards.** Another distribution technology, BBs allow users to access common communicating areas and provide a more interactive means for communicating.

**Client/Server Architecture.** Emerging as the key to 90s computer architecture, the client/server scheme is being increasingly used in E-Mail implementation. E-Mail is a resource-intensive application, particularly from the perspective of operating in a mainframe environment. Using microprocessor-based systems to handle these "back end" or server mail processes, which will likely include implementing X.400 transfer sessions and X.500 directories, and other routing and bridging processes will become a growing area in the next few years. On the "client" end, the end user's interface with the underlying network will provide seamless access to the entirety of the available network through connection to the server.

**Document Interchange.** E-Mail can be used for document interchange, and some E-Mail systems conform to the emerging document interchange specifications. E-Mail is much more widely established in the communications infrastructure than are some data and document interchange standards.

**Electronic Data Interchange (EDI).** EDI generally refers to the set of standards established to allow businesses to exchange data in common formats. Examples include such common business processing items such as purchase orders and invoices. As organizations become increasingly interconnected through the demand to communicate through E-Mail networks, EDI will ride on the communications infrastructure being established. Thus E-Mail growth will spur the widespread acceptance of EDI, an acceptance which has seemed imminent for the past decade.

**Facsimile.** Another significant electronic messaging technology, facsimile, has become almost universal in the business environment. Fax interfaces to E-Mail systems are natural extensions of both sets of installed bases. E-Mail packages with fax interfaces allow an electronic

message to be sent to any fax machine, combining timeliness with immediate hard copy.

**Forms.** An emerging E-Mail feature is the ability to create and use custom forms for messaging. Office work often revolves around managing forms, and using an in-place E-Mail network for forms transmission is a natural application. Forms-based applications are widely touted as an application of the future, and integrating form creation with the ability to transmit them electronically is an idea in its infancy, but spreading fast.

**Gateways.** The growth of internetworking technology to extend links between heterogeneous systems has opened a market for both standard and dedicated gateways among these systems. Providing the means to simply exchange messages is a basic application and one that has generated an entire submarket for internetwork message passing products. As the X.400 standard spreads, more gateway products will appear.

**Graphical User Interfaces (GUIs).** Apple's Macintosh began the GUI revolution in earnest, changing the face of microcomputer applications, and ultimately all applications. Naturally, E-Mail packages running on the Macintosh use a GUI, as do others that run under Microsoft Windows. Eventually, most applications will evolve away from text orientation and use graphical elements in their interface with users.

**Image Processing.** One of the hot technologies for the 90s, document imaging systems are using underlying E-Mail capabilities to transmit images through networks. As imaging systems grow in popularity and as images become more commonplace, the ability to share large image files will become more crucial. Although the X.400 standard does not provide specific support for image objects, its built-in support for Office Document Architecture/Office Document Interchange Format (ODA/ODIF)—which provides a model for defining the logical structure, content elements, and layout of a compound document (one which incorporates text and graphics) and for regenerating the transmitted document—supplies users with an adequate framework for sharing document data in accordance with the OSI model.

**Interprocess Communications.** This refers to the ability of operating systems to allow two or more applications running either at a single workstation or at more than one workstation to share data. An emerging concept, multiuser interprocess communications will allow users to use the same file simultaneously, with the communications handled transparently to the user. This concept is part of the OS/2 operating system and of Apple's forthcoming System 7.0 for the Macintosh.

**Mail Enabled Applications.** One of the major trends in applications software is to enable standard desktop applications such as word processors to communicate using underlying E-Mail capabilities. This takes the form of E-Mail being made available to a user as a menu item within an application—a user editing a document can then transmit the document (spreadsheet, etc.) using the communications capabilities of the E-Mail systems without exiting the application. In the future most applications will grow more tightly coupled with underlying communications facilities that exist either as part of the operating system itself or which run as separate applications.

**Memory Residence.** Memory residence, primarily in reference to microcomputers, implies that the E-Mail application runs as a background task, while the user runs another application in the foreground. Users can then call



the mail application and send messages while the foreground application continues to run, and the mail application is available to receive mail and then notify the user when a message is received.

**Multimedia.** Related to the imaging issue, multimedia capabilities are becoming increasingly common, and E-Mail packages are being built with the capability to share large documents that contain graphics, sound, and elements other than text. Once again E-Mail serves as a convenient enabling technology for communicating these

**Objects.** The software environment is being slowly and inexorably altered by the emergence of object-oriented programming and principles. This trend extends to E-Mail where packages that use object technology treat messages as objects which can then perform predefined actions. Objects enable creation and distribution of "smart" messages which include not only message data but information about how the message data can be used and perhaps even the tools to manipulate the message itself.

**Voice Integration.** Another emerging capability in office systems is voice integration and annotation. Capabilities such as software that "reads" computer-based messages to a user through standard telephone lines demonstrate the extent to which E-Mail is building a parallel communications infrastructure to the telephone network by forging bonds with voice-based telephone technology. Other applications include voice/E-mail integration found in a package such as Wang Office which notifies users of voice messages, and allows these messages to be distributed using the E-Mail network.

### Computer Supported Cooperative Work

Computer supported cooperative work (CSCW) is an emerging concept based, like E-Mail, on communications resources. The software tools that implement this concept, generally referred to as "groupware," consist of products that provide the means to create and share information as well as some framework for managing and documenting its interchange.

Groupware tools, at minimum, automate the logging of a document's movement from one worker to another. Some groupware tools provide on-line group calendars that allow meetings to be scheduled when the entire group is available. Most groupware applications include store-and-forward E-Mail modules.

Some groupware products are specifically intended for use in design environments. Designs in process are circulated among group members. The package manages the movement of documents from one step to another and automatically creates an audit trail that indicates what was done at each step in the design.

Other groupware products are designed for editorial environments and include tools for critiquing and editing text documents. Still others manage office communications, providing a framework for exchanging information and allowing managers to monitor this exchange.

Implementing groupware entails some risks. Privacy issues must be resolved, since the flow of information in groupware applications is public. If used too strongly for group control, groupware may inhibit group creativity and teamwork. The public nature of groupware also entails security risks. Breaking into another user's calendar and changing a schedule would undermine the cooperation that groupware can provide. A manager must ensure that all group members are trustworthy before granting them access to groupware tools.

## Electronic Mail's Advantages and Restrictions

When properly implemented, E-Mail has measurable advantages over alternative messaging methods such as:

**Telephones**—The telephone is a poor substitute for E-Mail. It is very unreliable as a means for delivering high-priority messages. A telephone conversation is not concrete enough to consummate many business transactions because it lacks hard copy and thus cannot document verbal agreements. Voice communications content is much less compact than text communications. Reading a 10-page document over the phone for someone to copy by hand would take much longer than sending it through some form of E-Mail or by courier service.

**The Postal Service**—The initial motivation for developing E-Mail sprang from dissatisfaction with the postal service. The postal service is designed to handle mail of low priority by today's standards. The age of electronic telecommunications has vastly increased our expectations concerning the speed of information delivery. E-Mail is popular because it satisfies this emerging need for instant delivery of information.

**Courier Services**—Local, regional, and national couriers often transport packages of information that must be delivered overnight. A combination of airplanes and trucks is usually employed for the swift movement of packages from city to city. While these services have proven reliable and swift by conventional standards, they are costly and cannot usually provide anything better than overnight delivery in most city-to-city situations. For companies engaged in the regular, heavy use of couriers, the cost of E-Mail can usually be justified by the elimination of the courier costs. Even so, when E-Mail is not possible between two points, services such as Federal Express, Purolator, or the U.S.P.S. Express Mail service can be effective delivery tools.

**Intracompany Mail**—The delivery of messages and documents within an organization can sometimes be the most aggravating bottleneck of all. Stories about interoffice mail taking days to circulate in some large companies abound and, once it gets there, no guarantee exists that the recipient will immediately look at it. Internal E-Mail systems seek to eliminate this problem by sending messages instantly and by notifying the recipient that a message has arrived.

A growing body of experience and a booming market testify to E-Mail's potential benefits. In a business environment increasingly fixated on speed, E-Mail builds rapid message turnaround into an organization's information infrastructure.

### Restrictions

Despite the allure of E-Mail as a strategic information management and communications tool, it is not a panacea. E-Mail's restrictions involve the following issues:

- **The Network's Scope.** The network should serve mail to a sufficient number of people so that the timeliness advantage is not overshadowed by the number of inaccessible people. Often, where some sort of mail network is in place, not everyone may be connected, and not all who are connected may use the network for mail. Electronic mail is not likely to be as universally accepted as the telephone until the next century.
- **Ease of Use.** A great majority of people can use a telephone and mail a letter. The same cannot be asserted for

computers and related networks. Because extensive user training is usually required, the goal of universal electronic mail remains far in the future.

- **Cost.** Certainly the cost of a microcomputer connected to a network is greater than that of a telephone or the postage necessary for one person's correspondence. Can efficiency gains justify the large start-up costs implied by microcomputer-based messaging systems? Because the microcomputer has uses other than for mail, the benefits of the mail component must be separated and stated explicitly.
- **Security.** Threats to individual privacy are great with the widespread introduction of computer technology. Once a computer is connected to a network, the information stored by an individual becomes accessible by others, and control over this access is generally out of that individual's hands. All electronic mail applications must make some provision for security.

One of the downsides to E-Mail, one growing more widely publicized, is workers getting swamped with an unmanageable volume of mail. Once organizations become dependent on E-Mail they discover that it is a resource-intensive application. In these cases organizations must recognize that indeed the systems they have in place may be inadequate to handle this new conception of a widely integrated organizations dependent on communications. Further, the E-Mail "front end" must provide users with capabilities to manage E-Mail effectively—a factor missing from mainframe-based systems. Host processing resources were clearly not designed for the demands that growing E-Mail volume entails.

### Cost Justification

Traditional cost justification methods miss the mark in the case of E-Mail. Defining the up-front costs on a per-user basis does not account for the possible traffic volume, and focusing on message volume does not factor in the intangible benefits of enhanced communications. An accounting of both per-user and message volume costs still neglects *value added* to knowledge work. Thus justifying E-Mail is difficult in terms of traditional methods. Not having it, however, is now practically impossible to justify. Why is this so?

The time savings offered by E-Mail are of central importance to its justification in an organization. E-Mail facilitates a new kind of work, one that allows users to create and receive messages while in the midst of other tasks. Ideas shared electronically in a workgroup can accelerate quickly, and all of the intermediate steps in the creation process are documented automatically. Actual time spent in the E-Mail process is usually negligible. Thus messaging becomes an extension of the *thinking* process as opposed to the automation of some process otherwise handled in some other medium (such as face-to-face or telephone communications). This is where users realize the real benefits of E-Mail.

It must be recognized that any benefits are in direct proportion to their planning. Thus, in a situation where implementing E-Mail is considered a cost rather than as an investment, the benefits are likely to be marginal. Alternatively, where the system's expected benefits are carefully planned from the outset, then the system is likely to take on a life of its own and to benefit the organization in positive, unforeseen ways.

The shift in demand to microprocessor-based solutions such as E-Mail server products that handle the required throughput with lower up-front investment grow more attractive to firms. Otherwise organizations end up in the position of one well-publicized firm—it was forced to simply disconnect the E-Mail system in order to run more strategically important processes. Over 20,000 employees lost their E-Mail resources in the process. This points up a failure to recognize that today's modern organizations indeed depend on E-Mail as a critical resource, and that system planning models are unlikely to account for the incipient system demands for E-mail service. This calls for completely new thinking about the value of interconnection both internally and increasingly externally as well.

Those responsible for information systems justification are developing their own special set of tools. Surveying user needs and opinions provides a starting point for determining actual benefits. Once perceived benefits are stated explicitly, users can establish the means to test and measure the perceptions. By treating and managing E-Mail as an investment and an asset, strategic payoffs—in terms of productivity gained and value added—can be far greater than explicit benefits such as decreased courier costs. This is the perspective required to make E-Mail pay off many times the amount invested in it.

### Selection Criteria

This report covers technology applicable across platforms—from microcomputer to mainframes. The networks using E-Mail vary from isolated LANs to international enterprise-wide networks. Moreover, large users may be locked into a small number of options, if any exist at all, due to the specific requirements of their installed hardware base. Thus specific criteria for judging and evaluating a given package are impractical to specify. Nonetheless, given the emerging strategic importance of interconnectedness in the workplace, it is vital for those charged with designing and specifying an E-Mail network to recognize its role in the organization.

As stated, common criteria are difficult to specify for a topic as far-ranging as E-Mail. The most important choices focus on:

- **Ease of Use.** Is the front-end system—the part that users encounter and use—easily navigated? Is it menu driven, suggesting likely actions at any given point? A good package guides users almost intuitively into the basic functions of message creation and sending, and message reading. All options available from a given point should be readily available.
- **Directory Services.** The package should allow easy access to anyone the user would want or need to communicate with. A listing of available "mailboxes" that is accessed in a single step is a reasonable expectation.
- **Security.** Basic security such as password access should be given. Beyond that, messages should be encrypted upon transmission, and stored in an encrypted form to prevent unauthorized access. Security levels available among E-Mail vary widely; if security is a primary concern, the vendor should be closely scrutinized as to its ability to develop secure systems.
- **Connectivity.** The package should be able to access as wide an audience as possible. With few exceptions, the package should be expected to eventually communicate with a variety of other systems. The software should at least allow some gateway through which the ability to

translate messages from foreign messages can be accomplished. An X.400 gateway is an expensive option for most right now, but it should at least be an available option should the user need it.

- **Distributed Processing.** As the computer environment grows ever more sophisticated, placing increasing amounts of power on desktops, the ability to take advantage of a variety of systems design elements that revolve around distributing computing power should be part of the E-Mail system. This implies use of a client/server architecture to separate message editing and local message handling from "back end" processes such as implementing X.400 gateways.
- **Interoperability.** This has several implications. At the desktop level, the E-Mail package must not interfere

with running any other package. In fact developers are increasingly including hooks into E-Mail functions in their applications software to provide transparent integration of communications into other applications. At another level it means that the same package from a single vendor provides the same capabilities and the same interface irrespective of the hardware platform on which it is running. This consistency is vital in saving on training costs.

- **Multimedia Support.** Increasingly, office applications entail use of graphics, voice, and image data. Certainly anyone planning for the future uses of E-Mail must factor in a package's capability to manage various data types. ■



# An Overview of Mass Storage

## In this report:

Selection Guidelines.....	3
Emerging Technologies.....	5

## Datapro Summary

Vendors in the dynamic microcomputer hard disk market continue to offer a wider variety of drives. The mass storage market includes drive component companies, OEM drive manufacturers, controller makers, storage subsystem integrators, and retail market specialists. Competition among the vendors has been beneficial to end users—prices have continued to decline, while capacity levels have risen. This report examines what this trend means to vendors. Products from leaders in the microcomputer mass storage market are reviewed and the results of Datapro's 1991 survey are analyzed.

## Technology Overview

Magnetic storage products differ widely in form factor, capacity, speed, and compatibility. All market segments are experiencing somewhat steady growth, despite tough economic times. On the whole, price per megabyte of memory is decreasing in the face of severe competition and increasing capacities.

Technology highlights can be summarized as follows:

- Internal drives are more prevalent than external (standalone) units.
- Drive capacities are increasing rapidly; some 5.25-inch drives now provide over a gigabyte of storage.
- Drive performance (speed and reliability) is improving.
- Removable drives provide portability and security for low-end systems.
- 2.5-inch drives are becoming the new standard for portable computers.

- The SCSI interface is by far the most popular for hard drives—74% of the drives in our survey use SCSI.

## Technology Basics

Disk drives continue to offer better interfaces, larger storage capacities, faster access times, reduced cost per megabyte, and smaller form factors. Magnetic hard disks have evolved into vital components for all computer types and sizes, from mainframes and supercomputers to today's smallest palmtop personal computers. The huge storage needs of imaging and management applications, graphical user interfaces (GUIs), and large databases have spawned many higher-capacity products. Once expensive options, microcomputer hard drives are now affordable even for beginning computer buyers, and more powerful and memory-intensive software has made 20M bytes of disk storage a minimum requirement for microcomputers.

## The Storage Subsystem

A computer's storage subsystem has two main components: disk drive and controller. The controller translates requests from the CPU for data to be read from or written to the drive. Today, controllers are included in the packaging of many larger hard

—By Todd R. Denton  
Assistant Editor

drives. The controller contains a microprocessor responsible for executing instructions independently of the system's CPU. Together, the drive and controller define system performance as experienced by the user. The integration of different components in a disk subsystem can greatly affect overall subsystem performance.

### Winchester Disk Technology

Winchester technology refers to the process of encasing data disks and the accompanying read/write heads in an airtight enclosure. Fixed disk drives offer some significant advantages over traditional removable disk drives. The sealed HDA virtually eliminates the problem of head crashes caused by contamination. Thus, no preventive maintenance, such as changing air filters or cleaning and aligning heads, is required, and mean-time-between-failure (MTBF) is significantly increased.

Greater bit packing densities and greater track densities have been achieved as a result of the low head flying height, which is approximately 14 microinches from the disk. Because magnetic flux spreads with distance, the greater the separation between the read/write head and the disk surface, the greater the area occupied by a bit of information. The low head flying height is made possible by the light head load pressure and a low head mass. Because of the thinner coating on the disk surface, the magnetic field of the head varies less through the medium, so that the magnetized regions acquire greater definition.

### Disk Coatings

Much of the current development in magnetic storage technology revolves around the evolution of new materials for coating the disks. Current disks are coated with iron oxide or composites based on iron, and developers have reached the physical saturation points of these materials. To expand beyond this physical barrier, vendors have begun producing disks with special aluminum-based composites. Where the current limits of existing materials lie at about 50M bytes per square inch, the new aluminum materials promise densities of 160M bytes per square inch by 1993. This implies capacities of 2.5G bytes on a 5.25-inch disk and 1G byte on a 3.5-inch disk. Beyond aluminum disks, vendors are developing magnetic disks composed of glass, again promising increased capacities and smaller form factors. These technical breakthroughs suggest that magnetic drives will remain a component of microcomputers into the next century.

Thin-film coating for disks involves a process called plating; platters are either coated through a chemical batch process or electroplated. The thin-film coating process gives the magnetic layer of the disk a more uniform composition than that afforded by the older and more common oxide coating process, and the thinner surface layer allows faster head flight.

Low-end Winchester disks are coated with either iron oxide or a composite metallic oxide substance suitable for storing data written by magnetic heads. The heads, mounted at the end of an arm assembly that allows them to access the entire disk surface, hover about 20 millionths of an inch above the disks. The disks reside on a spindle attached to a small high-torque motor and spin at a standard 3,600 revolutions per minute.

### Actuators

One notable and increasingly implemented small-drive innovation is the incorporation of closed loop actuator systems with linear or rotary voice coils—a technology once

almost exclusive to higher capacity, higher performance drives—into lower capacity and lower performance drives. (The actuator system positions the read/write heads over specific tracks on the surface of the disk.) Normally, such small, low-capacity drives employ an open loop actuator that uses a stepper motor. While stepper-based systems are inexpensive and simple, they are also somewhat imprecise and slow; the stepper receives no feedback from the read/write heads as they move over the surface of the disk and is limited in the precision of its movements by the size of its incremental steps. A stepper-based actuator cannot make fine adjustments for head positioning, leading to less efficient packing of tracks and slower access times, which are frequently twice those provided by closed loop systems.

Closed loop systems with voice coil actuators receive feedback from the read/write heads and gain greater precision of movement across the disk surface by accounting for the position of the heads and the velocity and acceleration of the disk. By compensating for environmental conditions, closed loop systems provide faster and more accurate read/write operations.

To further enhance storage capacity and transfer speed, almost all DASDs now employ thin-film technology in read/write heads, platters, or both. Thin-film heads feature a spiral film of electrical conductor deposited on a silicon substrate (instead of a coil of wire wrapped around a ferrite core). These heads are smaller than conventional heads, and respond more quickly to changes in the magnetic fields on the disk to allow data to be read and written faster. Thin-film heads can be more accurately aligned and give good frequency response up to 100MHz. The more homogenous thin-film surface provides a disk with a greater recording density than can be achieved with an oxide disk.

### Coding Schemes

A vendor can use one of two techniques to encode data on disk. One technique, Modified Frequency Modulation (MFM), uses sensitive heads to squeeze bits, represented by magnetic stripes known as flux changes, onto disk surfaces. The density of these flux changes can be increased by plating the disk surfaces with a thin film of alloy in place of traditional iron oxide. Putting more bits of data onto the surfaces, however, strains the capability of the heads to recognize where one bit ends and the next begins, so data integrity is decreased.

The other technique, the Run Length Limited (RLL) scheme, increases the amount of data on a disk by one half. RLL more accurately encodes and decodes information on the disk in accordance with a complex set of rules. Again, the encoding has to do with the flux changes on the disk surface. Most RLL controllers use 2,7 RLL encoding, meaning that in the flux pattern written to the disk, no fewer than 2, and no more than 7, *off* or *0* readings on the disk separate each *on* or *1* reading on the disk. MFM controllers use a 1,3 encoding scheme, meaning that smaller amounts of information are written to a given unit area of the disk. In principle, using an RLL scheme increases the density of installed disks that use the MFM scheme. Drives must be certified as RLL-compatible. Low-end, uncertified hard disks cannot support an RLL controller, as their electronics are inadequate to work at the much closer tolerances that RLL requires.

## The Controller

The controller translates requests from the CPU into instructions understood by the disk drive and transfers the data between the CPU and hard disk. To obtain optimal performance from the storage subsystem, the drive controller must fit exactly with the disk drive component. Otherwise the user will incur a performance deficit. Various controllers support different hard disk capacities and data transfer rates. Others provide higher level services such as error checking and correction.

- **ST-506**—Shugart Technology (now Seagate Technology) pioneered this standard in the early 1980s, using it in small-form-factor disk drives beginning with the 5M-byte Model ST506. The ST506 drive used a 34-pin daisy-chain cable for control signals and individual 20-pin radial cables to carry data between the controller and the drive; this same physical arrangement is used in the ST-506 controller. Used by the IBM PC/XT/AT and compatibles to combine diskette drives and hard drives into one system, the ST-506 controller supports up to 127.5M bytes of data and is used by most hard disks under 40M bytes. Two hard disks can be daisy-chained to an ST-506 controller. An ST-506 controller transfers data in a serial bit stream at a rate of 625K bytes a second; however, since the bit stream contains information that the controller needs to identify and delimit the data, the actual data throughput is 510K bytes per second.
- **ST-506 RLL**—This is an RLL version of the ST-506 controller. It packs 50% more data onto a disk than the ST-506 controller by more accurately encoding and decoding information on the disk using complex rules. As a result, RLL encoding requires more accuracy from the hard disk's electronics, heads, and media, and not all hard disks have the necessary electronics to support the ST-506 RLL controller. An ST-506 RLL controller supports capacities up to 200M bytes and transfers data at 750K bytes per second. As mentioned, RLL disks must be certified as such. *Consult the vendor or your dealer before attempting to combine a hard disk with an ST-506 RLL controller. Incompatible configurations will lose data.*

## Selection Guidelines

The best plan for purchasing a hard disk is to consider applications and find a drive that satisfies all that application's requirements. The extremely wide variety of drives on the market makes it possible to find a suitable drive for any application. The following are some key issues:

### Capacity

Disk drive capacities increase continually in all segments of the market. Maxtor, Seagate, and others now market 5.25-inch disks with capacities over 1 gigabyte (1,000 megabytes). Average new desktop system configurations have moved from no fixed storage to 10M-byte drives, from 10M to 20M bytes, through 40M bytes, to 70M bytes and beyond in the latest systems. Today, even machines used primarily for word processing often have 60M-byte drives. Higher capacities result from:

- evolving material technology that produces disks with much finer packing of metal oxide on the surface, allowing more data to be written per unit of surface area;
- improvements in all mechanical components, allowing much finer tolerances in their production and operation; and
- increasingly sophisticated electronics, implying both miniaturization and integration of components allowing more precise operation of drives and creating more space for disk surface inside the drive housing.

The optimal capacity for either a replacement disk or for the disk in a new system is largely determined by the application. Applications are changing, as are operating systems. These conditions make accurately forecasting future requirements difficult. Some certainties exist: OS/2 and UNIX require more storage than MS-DOS. Applications using graphical user interfaces consume far more storage than do character-based applications. Multimedia tools are becoming increasingly commonplace; storing full-motion video and digitized audio, even when compressed, requires copious storage. Those planning resource requirements are well advised to overestimate the future requirements for all but the most predictable tasks, such as character-based word processing. Users should purchase abundant storage resources now or plan on adding it later. Demand for storage always rises, and the rate at which it rises is unlikely to bear a linear relationship with past requirements.

## Controller Standards

Growing computer applications such as wide area networks, distributed processing, imaging, and CAD have increased user need to optimize storage resources. Critical to the process is increased processing power, and flexibility in the management of memory. In addition, the move to full 32-bit systems implies a necessity for much higher performance levels than previously attainable and demands a higher level of integration with the system.

## SCSI

The Small Computer System Interface (SCSI) was developed in the late 1970s as a connecting channel for computers and intelligent peripherals. Today, it is the most popular hard drive interface. SCSI (pronounced *scuzzy*) uses a parallel bus, with data flowing through the bus in a parallel (instead of serial) bit stream. A SCSI controller performs parity checking on the data that traverses the bus to further increase the accuracy of the data transfers between disk and RAM. SCSI is a channel, not a device controller; therefore messages are passed along in a logical, rather than physical manner. This allows SCSI devices to be daisy-chained along the bus, with logical control signals and data from one or more peripherals flowing through each device to the CPU.

SCSI is a very strong candidate for the controller standard of the future. Its major advantage is its ability to connect up to seven compatible peripherals to a single controller. Thus, a single controller card could manage an array of peripherals such as a Winchester drive, a tape backup unit, a WORM drive, or even a printer. With the current trend toward smaller desktop units, and hence fewer expansion slots available for peripherals, SCSI provides system and peripherals vendors with the expandability that users will require from the small footprint systems of the future. To its detriment, SCSI does entail a performance deficit, as the instructions to control the different signals take time for the devices to interpret, execute, and then transmit down the line. Later versions of the standard, such as SCSI-2, have been designed to reduce this performance deficit and will serve to broaden SCSI's appeal.

SCSI, though still not fully integrated into the desktop PC marketplace, is ideal for memory over 20 megabytes.

As such, it holds the most promise as a long-term industry standard. The SCSI interface is most popular in our 1991 survey—74% of the drives use SCSI. In 1990, we found only 69% that used the SCSI interface. SCSI is a system-oriented standard which minimizes CPU (system) control over peripherals and allows *plug and play* subsystem compatibility for systems using the interface. Thus a SCSI-compatible peripheral works with any system that has a SCSI port, allowing cross-system compatibility. SCSI allows daisy chaining of many drives as applications and memory requirements increase.

### ESDI

The Enhanced Small Device Interface (ESDI) is an extension of the physical arrangement of the ST-506 controller, introduced to support higher capacities and faster data transfer rates. ESDI supports a maximum capacity of 765M bytes and transfers data at up to 15M bytes per second. Through the addition of a data buffer to hold data during transfers, ESDI controllers significantly enhance data integrity, enabling recovery of data that would normally be lost with an ST-506 controller. Additionally, ESDI implements features such as on-disk error-checking, and optimizes throughput through control of the location of data and placement of the read/write heads. Both these factors provide ESDI with its distinct performance advantage in comparison with SCSI.

ESDI is essentially a device-oriented standard that allows drive or system vendors to modify the drive subsystem in order to better integrate it into the overall system. ESDI features high data transfer rates (up to 15M bytes per second) and promises much greater data integrity than previously attainable. Ten percent of the drives in our 1991 survey use the ESDI interface.

### Other Issues

The final issue regarding controller standards involves the ongoing debate in the microcomputer industry pitting IBM's proprietary Micro Channel Architecture against the Industry Standard Architecture forces. As the IBM-compatible, MS-DOS environment comprises the industry's largest market, the technology choices its vendors make determine the future direction of all corresponding subsystem vendors. High-performance, 32-bit systems used either as standalone systems or as servers on LANs require the high performance provided by the ESDI standard. Multiple storage devices connected to a single controller, as SCSI allows, are an increasingly common solution, especially for LANs. The advent of multimastering system buses expands the potential for even more intelligent drive controllers that operate independently of the CPU. Now in their infancy, these issues will emerge to define the next generation of storage subsystems.

### Removability

Removability remains an application-dependent factor for hard drive users. Security is the most common reason to select a removable drive. Users working with confidential information, such as financial data, often need to remove the storage medium and secure it. Other removable drive applications include application distribution (especially for large, vertical market applications) and media transfer (especially in large sites exchanging incompatible removable media). Removable drives have benefited from technical advances in miniaturization, allowing increasingly sophisticated drive mechanisms to fit into small, removable cartridges. Bernoulli technology has been available for

some time, and its chief proponent, Iomega Corporation, has rebounded to extend the line's capacity and establish its presence in the Macintosh market. Though the Bernoulli Box has found its successful niche, it remains a technology whose performance precludes it from serving as a primary storage device for most users.

### Form Factor

Most drives included with microcomputer systems are internal drives. Originally, systems were physically configured to hold a single add-in drive, often forcing users to remove an existing diskette drive. Newer systems include more internal space (drive bays) for additional drives. Tower configurations usually provide more internal space than desktop units and are a good choice if more storage may be added in the future. In general, it is preferable to have Winchester drives mounted internally, where they are much less likely to be exposed to impact. External drives occupy more space than internally mounted units, and must be housed sturdily.

Hard disk drives are constantly shrinking in size. Though 5.25-inch drives are still prevalent and increasing in performance, 3.5-inch drives are also becoming increasingly popular. A trend toward smaller desktop units and the explosion of the market for portables have significantly boosted demand for smaller drives. The future holds an emerging 2.5-inch form factor. These tiny drives will no doubt serve as the standard for future portables and even desktops.

### Networks

Network capability and compatibility is still a developing area for hard drives. The rapid growth of LAN installations has taxed the available resources for establishing and supporting adequate LAN hardware configurations. Many factors (controllers, capacity, performance) determine a drive's applicability as a network drive. Network drives need large capacity, require a high-speed controller, and must be reliable. In many cases they also must be tested for compatibility by the network software publisher. Most network-targeted storage units are geared toward Novell NetWare, the most popular network in use today. However, the task of matching hard drives to networks is still a tough proposition for many users establishing a network.

### Optical or Magnetic?

This issue becomes hotter each month, as various writable/erasable (magneto-optical) and multifunction products reach the market. Optical media are rapidly moving into position as primary storage technologies for microcomputers. Several distinctions need to be drawn to understand this technology debate. Optical disks no longer offer only three technologies (CD-ROM, WORM, and erasable); multifunction drives combine the technologies for greater user flexibility. CD-ROM is a read-only medium; it cannot be reused for storage. WORM (Write Once/Read Many) drives are primarily an archival medium—users can write to the disk only once, after which the data is available only for reading.

Erasable optical media pose a more significant threat to magnetic storage, as their operation duplicates that of a Winchester drive. Rather than using a magnetic head, writable/erasable drives use a laser to heat the area being written while exposing the same area to a magnetic field. Depending on the polarity of the field, the surface of the disk (which is magnetic) becomes more or less reflective.



The pattern of reflective and nonreflective patches represents the binary code of 1s and 0s. When reading, the drive shines a weaker laser beam on the patch, and light-sensing circuitry detects the pattern. Erasable drive capacities depend on the capacity of each individual disk, rather than having fixed capacities like magnetic drives. These drives allow a user to store over a gigabyte of data on one disk. Therefore, a single disk could satisfy the storage requirements of almost any individual user; a cluster of six disks could satisfy the storage needs of a workgroup. Thus the threat posed to magnetic storage is clear. The results, however, have been far less so.

Erasable storage is still moderately expensive (vendors must recapture high development costs until unit volume allows prices to fall). Magnetic drive makers do not yet view erasable optical storage technology as a major threat, as magnetic storage is widely installed and well-trusted. Magnetic storage technology, still in its infancy, promises to deliver higher capacity, faster access, and less expensive drives.

### Performance

A combination of factors determines disk subsystem performance. Despite manufacturer's emphasis on a disk's average access time, this measures but one of nearly a dozen separate procedures required to transfer a byte of data between a disk and a CPU. The average access time measures the average time between a request to read or store data and the beginning of data transfer. The lower the specified average access time, the faster the drive performs the operation. Access time is important, but bottlenecks elsewhere in the subsystem could significantly reduce performance expected from a drive with a low average access time. Other factors, such as the data transfer rate from the read/write head to the disk's internal electronics and the transfer rate from the electronics to the controller, must be considered. Any weak link in this sequence will impede performance. Users are thus at the mercy of system or drive vendors that match components, as the decisions revolving around the final configuration are the major determinants of storage subsystem performance.

Although it may be possible for an individual to match controllers with drive units to boost performance, it is not a recommended practice for corporate installations. The money saved in mixing and matching may not offset the risk that parts may be mismatched. Vendors put their reputations on the line with their component choices, and most do a solid job in assembling the best performing subsystems possible.

### Price

As the disk drive market matures, two trends have emerged to shape end-user purchasing patterns. First, storage prices have continually decreased. In 1981, a 10M-byte disk cost over \$1,000; today a 20M-byte drive is the smallest new drive available and costs under \$500. It is doubtful that some analysts' rosy predictions of \$1 per megabyte by 1992 will come true.

### Reliability

Disk drive vendors dutifully claim laboratory-determined mean-time-between-failure (MTBF) ratings, but these are small consolation for a user with a failed disk. As a general statement, hard disks are reliable by nature—they are a closed system kept free of environmental trauma and, short of physical abuse, they tend to last as long as vendors expect. They are not, however, immortal. Depending upon the quantity and quality of use, practically all disks are bound to fail. Some failures will yield unrecoverable data, generally from a power surge or a chronic power problem.

Hard disk buyers should pursue every possible avenue to determine product reliability. MTBF specifications notwithstanding, there is no substitute for information regarding a vendor's field-determined reliability experience.

---

## Emerging Technologies

Storage technology is very dynamic, and fundamentally new concepts constantly arrive on the scene. One such technology is the *floptical disk* drive developed by Insite Peripherals. Using a laser-servo to write control tracks on magnetic media, floptical drives combine laser and magnetic technology to produce 20M-byte capacities on a 3.5-inch removable disk. Iomega Corp. is using the technology in its Bernoulli Box line.

Hard disk drives are very reliable, yet not infallible. Magnetic drives contain moving parts, and mechanical devices are always subject to eventual obsolescence. If we look far into the future for an ultimate replacement for magnetic storage, the answer is both simple and remote. Mass storage will eventually be solidstate. Flash EEPROM (Electrically Erasable Programmable Read-Only Memory) technology points the direction for development. A type of nonvolatile memory, it does not require electricity to hold its contents the way RAM does. Once prices fall sufficiently, microcomputers will contain no moving parts, primary storage will be solidstate, and archiving and removable storage will be laser based. ■



# An Overview of Optical Storage

## In this report:

Products .....	3
Future Directions.....	5

## Datapro Summary

Optical storage is an international business, with many of today's technological developments coming from Japan. Most drive and media makers are based in Japan, but Kodak, LMS, Philips, and Dupont are notable exceptions. The available selection of optical disk drives, especially of rewritable drives and jukeboxes, continues to grow. Multifunction drives, an emerging technology, double as both write-once and rewritable drives. The optical storage market will continue to grow, with the fastest growth and the steepest price drops in rewritable and multifunction drives. But despite a growing market, optical storage will be used with, rather than instead of, magnetic storage.

## Technology Overview

This section of the report explains the three main optical disk technologies—CD-ROM, write-once, and rewritable—but focuses on the latter two, which are more directly applicable to document imaging applications.

- CD-ROM—a read-only disk used to publish and distribute large databases and reference works.
- Write-once—also known as WORM, read/write or writable disks, write-once disks archive information. Since they cannot be erased or overwritten, write-once disks are ideal for archives.
- Rewritable—like magnetic media, this product (also called erasable) can be written to many times, but rewritable optical disks have storage capacities far beyond that of magnetic media. A new type of rewritable drive, called the multifunction drive, accepts both write-once and rewritable disks.

—By *Kenny Weilerstein*  
Associate Editor/Analyst

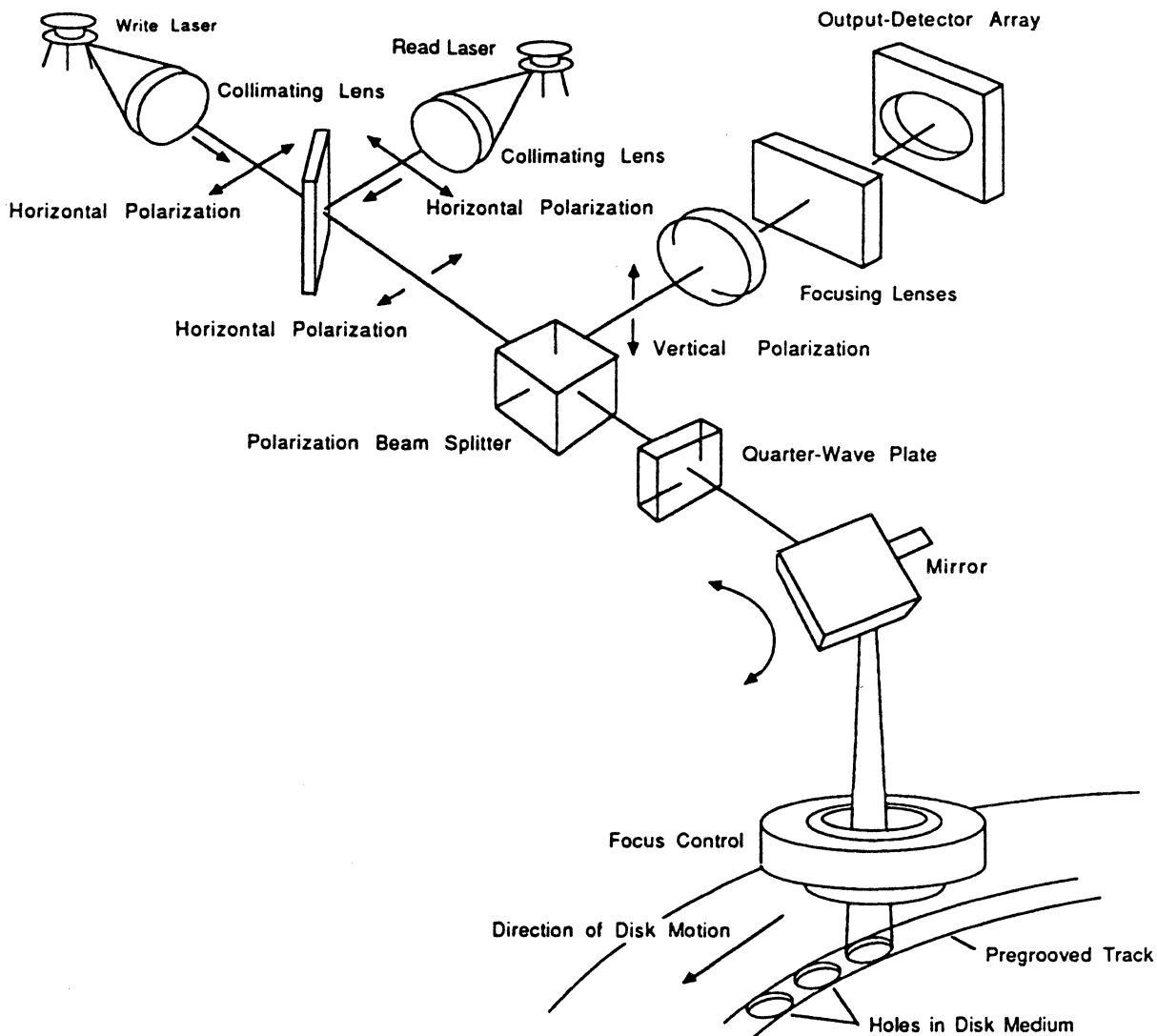
## Technology Basics

### CD-ROM

Though CD-ROM is used by only a few companies in the imaging market, it is a good starting point for understanding how other kinds of optical disks work. Compact disks have become one of the most popular ways of recording music. Readers are familiar with the 4.7-inch plastic disks with the mirror-like surface. The kindred CD-ROM disk, and a similar player (drive), are used to distribute information for microcomputers. Digital technology, the recording of information as zeros or ones, makes this possible.

Under a powerful magnifier, a CD-ROM would show spiraling reflective tracks interrupted by pits. The pits carry the information, and the manufacturer presses them into the surface during duplication. When the user puts the CD-ROM in the drive, the drive reads the information with a laser beam and a light detector. Just as the familiar phonograph needle runs along a groove on the platter, the drive shines a laser beam on the narrow track. Because the CD-ROM has a reflective surface, it reflects the laser beam back to the light detector. The drive circuitry interprets an unwritten spot as a 0. When the beam crosses a pit on the track, it is scattered and

Figure 1.  
A Write-Once Drive



The components of a simple optical disk system. Information is written with a high-powered laser beam and read with another, low-powered laser beam. The main components of the system are the lasers, the polarizing beam splitter, the focusing lenses, the output-detector array, and the recording material.

does not trigger the light detector. The circuitry interprets the interruption as a 1. Another change back to the reflective surface signifies another 1.

A powerful magnifier is needed to see the tracks and pits on the disk because 16,000 tracks fill each inch of disk surface; about 2K bytes of data are stored in each 0.75-inch length of track—a track density tenfold that of magnetic disks. One CD-ROM holds about 635M bytes, yielding a total of about 553 usable megabytes. The remaining 82M bytes are used for a variety of functions.

Some are used to separate the surface of the disk into segments so the read/write head can move from one place on the disk to another. Some are used for error correction. Although a clear, hard plastic layer protects the surface of the disk, it can be obscured by dirt and fingerprints or

damaged in manufacturing. About 77M bytes per disk allow the drive to recover from small interruptions in the datastream. Typical error rates range from 1 in 10 billion to 1 in 10 trillion, or, at most, about 1 error in 18 fully loaded disks.

CD-ROM manufacturing comprises three major steps: premastering, mastering, and duplication. Most CD-ROM publishers distribute databases, indexes, or other reference works. If the information is already stored electronically, the publisher must convert it to the suitable format. If printed on paper, it must be scanned or keyed in. Software must be prepared to access the information on the disk; such programs are available commercially.

The mastering system, in the past usually a Digital Equipment Corporation VAX but today often a specially equipped microcomputer, formats the information into

2,048-byte sectors. These sectors are laid out on a nine-track magnetic tape in the same sequence they will appear on the disk, and an error correction code is added. Then a glass master disk is created, from which the manufacturer eventually makes the "son" disks used to stamp data onto raw disks. Some mastering systems, such as Meridian Data's CD Professional, can produce a sample disk in minutes on a special writable CD. At a media cost of over \$100 per disk, the samples are suitable only for checking the mastering or for very small runs of disks in a hurry.

### Write-Once

Unlike CD-ROM drives, write-once drives can both write and read. As shown in Figure 2, the hardware components of a write-once disk drive are the lasers (one write laser and one read laser), the polarizing beam splitter, the focusing lenses, the recording material, and the output-detector array.

The optical disk itself consists of a three-layered "sandwich" and an airspace. Two of the three layers are either glass or plastic. These two layers are designed to reduce the oxidation of the third layer, which is usually tellurium, rhodium, or an alloy of either. The third layer provides the reflective surface and stores the data. The airspace is created as either a controlled atmosphere or a vacuum, reducing oxidation of the reflective surface. The inner and outer edges of the disk seal the vacuum and complete the optical disk sandwich.

Data can be written to the disk surface using one of several approaches—ablative, bubble forming, bimetallic alloy, dye polymer, or phase change. Using one of these techniques, a write laser applies binary digital information to the disk. A mark or the absence of a mark is read as a binary 1 or 0. Using the ablative approach, a write laser melts a small pit less than one micron in diameter (less than the diameter of a human hair) on the sensitive metal alloy surface. Laser Magnetic Storage International uses the ablative approach in the LaserDrive 1200E.

To read the data, a lower intensity laser is beamed through the glass or plastic to the metal coating. When the beam is focused on an area with no hole, sensors detect a high-intensity reflection that translates into a binary 1. A low-intensity reflection caused by refraction of the light by the hole signifies a binary 0. The ablative media last about 10 years but are expensive to manufacture.

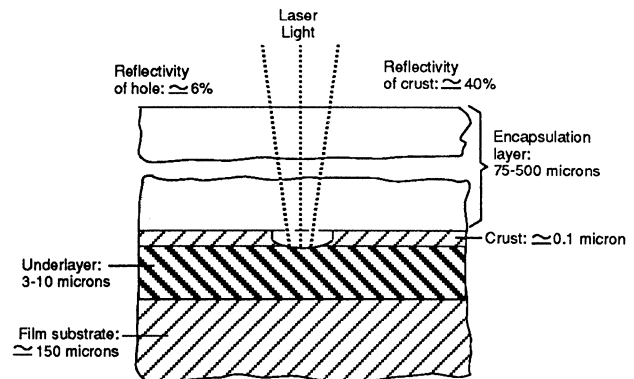
With the bimetallic alloy method, a variation on the ablative approach, the laser heats two metallic layers, causing them to fuse into an alloy that has a different reflectivity. Sony, which developed this approach, estimates that the media will last 100 years.

Using the bubble-forming approach, a laser beam heats a sensitive layer causing thermal decomposition. Each bubble so formed represents a binary digit. When a lower intensity laser reads the recorded data, the bubble scatters light, reducing reflectivity. Again, the variations correspond to a binary 1 or 0. The formation of bubbles on the disk surface creates a more defined circular edge for reading by the low-power laser beam than the rough edge produced by the ablation method.

In the dye polymer approach, a layer containing a dye polymer changes color when heated by the laser beam. Such disks are inexpensive but last only about five years.

The phase-change methods use the laser to alter the condition of the disk coating from crystalline (more reflective) to amorphous (less reflective).

Figure 2.  
A Write-Once Disk



*Reflectivity differs between the original surface of an optical disk and the hole that has been burned into that surface.*

### Rewritable

After years of expectation and speculation, rewritable optical disks are a commercial reality. The drives store between 128M and 2,008M bytes per side on a 5.25-inch, removable disk. The models currently available are mostly magneto-optical.

Magneto-optical disks resemble write-once disks with one major difference: The same track on a disk can be written over 1 million times. A phenomenon called the Curie point makes this possible. When heated to the Curie point, magnetic materials realign their polarity to that of an adjacent magnet. In optical drives, the read/write head shines an intense laser beam to heat the area being written while exposing the same area to a magnetic field. Depending on the polarity of the field, the surface of the disk (which is magnetic) becomes more reflective or less reflective. As with other optical disks, the pattern of reflective and non-reflective patches represents the binary code of 1s and 0s. When reading, the drive shines a weaker laser beam on the patch, and the light-sensing circuitry detects the pattern.

Multifunction drives, which were announced in the spring of 1990 and should be shipped in production quantities soon, take both write-once and rewritable disks. Pioneer's multifunction drive uses the same magneto-optical media as most rewritable disks. For the write-once function, though, it uses ablative media, which cannot be overwritten. Matsushita's multifunction drive combines two distinct innovations. The rewritable disk, which uses the same phase-change technology as the write-once disk, is the first innovation. One-pass writing is the second innovation. Instead of erasing each portion of the disk before it writes to it, Matsushita's drive writes directly to the un-erased surface, significantly shortening the write time.

### Products

The optical storage technologies discussed previously are used in different ways. CD-ROM is typically used for distribution rather than storage. Write-once disks are used mainly for archiving. Rewritable disks are usually used for on-line or near-line storage of large, data-intensive images and for backup of magnetic disks. Multifunction drives take either write-once or rewritable disks and can be used

in either application. The following paragraphs discuss standards, advantages, and disadvantages of each technology.

## CD-ROM

### Standards

CD-ROM adheres to more standards than any other optical disk type. Sony and Philips, the codevelopers of CD-ROM, specified strict mastering and production guidelines. In 1987, the key vendors in the CD-ROM market announced standards for the arrangement of information on the disk. The High Sierra format (named for the hotel in Lake Tahoe, NV, where the vendors met) allows different retrieval programs to read data from the disk in the same way, regardless of the host computer or operating system. The ISO codified most of the High Sierra format in its ISO 9660 standard.

### Advantages

A similarity to CD audio constitutes CD-ROM's biggest advantage. The huge audio market for blank compact disk media and duplication services means that publishers can duplicate and distribute the disks for a few dollars each. If there are enough clients, the additional cost of mastering the disk (about \$3,000 to \$10,000) can be absorbed. CD-ROM can disseminate voluminous information for a much lower cost than print, and it can be searched by computer. Thanks to its resistance to electromagnetic or mechanical damage and its error correction codes, CD-ROM is more secure than any other medium. Finally, as the word compact denotes, the disk fits in a small, thin package, making distribution easy.

### Disadvantages

As a read-only medium that is expensive to master and duplicate, CD-ROM is suitable only for publishing. The disks are inexpensive but cannot be updated. Master disks are expensive to manufacture, making them unsuitable for small runs. Mastering and duplication also take days. By contrast, on-line information can be updated in about one day. The drives are also costly, with OEM prices from \$450 and up. Even after retail discounts, the cost of the drive and host microcomputer makes the initial cost of owning CD-ROM higher than that of most reference books, even lavishly printed encyclopedias. Finally, CD-ROM drives are slow—most take 330 milliseconds or more to retrieve data. This is slightly faster than last year, but more than 20 times longer than high-capacity magnetic drives and much longer than write-once or rewritable drives.

## Write-Once

### Standards

Vendors cannot agree on a single write-once standard, so the result is a split standard. The ANSI XB311 standard and its ISO counterpart permit both the continuous-composite formatting that most vendors support and the sampled servo formatting found in Hitachi, LMSI, and Pioneer drives. Buyers must also weigh the trade-offs between different disk sizes, file structures, and interfaces.

### Advantages

Like other types of optical drives, write-once disks hold huge amounts of data, the equivalent of many filing cabinets of paper or rolls of microfilm. They are removable and

can be swapped by hand. Write-once disks are rugged enough and light enough to be swapped quickly by jukeboxes, yet are almost invulnerable, short of fire or vandalism. Unlike fixed magnetic disks, they fit compactly in a cabinet or safe. Write-once disks also have a long archival life—up to 100 years.

### Disadvantages

Some disadvantages are inherent in the write-once design, while others come from its novelty. Despite access times of about 60 ms on the fastest drives, write-once disks are still much slower than high-capacity magnetic hard disk drives. Because the disks are not rewritable, space is lost when files are updated. The drives are expensive (\$1,200 and up), as are the disks (\$65 and up for single-sided disks).

The legal aspects of documents stored on write-once disks are not entirely clear. In the opinion of George S. Kondos, a computer specialist and attorney with the U.S. Department of Justice, the courts have not fully addressed the admissibility of information stored on optical disks, but they probably will find them at least as admissible as magnetic disks and possibly more trustworthy. The legal acceptability of signatures stored on optical disks is also in question.

## Rewritable and Multifunction

### Standards

Standards are emerging for rewritable disks. First, all the current disks have a diameter of 5.25 inches. Second, the ANSI XB311 committee and the ISO are expected to issue a standard for rewritable drives using continuous-composite formatting. Like the write-once drives, most rewritable drives have a SCSI interface, but a few have ESDI instead.

According to Bob Abraham of Freeman Associates, there has been much activity in the ANSI X3B9 technical committee and in the corresponding ECMA committee on developing a standard for multifunction drives. The developers are divided into three camps:

- Hewlett-Packard leads an industry group that favors the magneto-optical technology used in most rewritable drives, but with differently coded media for write-once and rewritable disks. This proposal maintains compatibility with existing media, and it is supported by Hitachi, Maxoptics, Ricoh, Sony, Toshiba, and by most or all of the media makers. Since the coding on the disk is the only thing that keeps the drive from overwriting a "write-once" disk, it is conceivable that data could be altered by accident or by design.
- Pioneer, LMSI, and Optimem (an OEM rather than a multifunction drive maker) favor the usual magneto-optical media for the rewritable function, but ablative media for the write-once function. This proposal ensures that archival disks and those containing legal documents cannot be overwritten. Several media makers back this proposal as well as the first one.
- Matsushita favors its own proprietary phase-change technology, which should have a very fast write access time.

### Advantages

Rewritable drives have the same high capacity, removability, and random access that make other optical disks attractive. Unlike CD-ROM disks, they can be written as

well as read; unlike write-once drives, they can be rewritten. For the amount of data they store, the cost per megabyte is lower than the cost of magnetic media. Multifunction drives also accept write-once media, so they are good for users who need both temporary storage, such as for recent data backups, as well as archival storage, such as for computer files that are seldom needed but too important to discard. Another use for multifunction drives is in document imaging systems that discard some information after a few weeks but keep other information for years. Eventually, the multifunction drives may cost less than write-once or rewritable drives to manufacture, since the manufacturer can reach a larger market with a single product.

#### **Disadvantages**

The main problem with rewritable drives is the slow access speed—the fastest have an average access time of 48 ms., compared to under 10 ms. for high-capacity magnetic disk

drives. The access speed is a barrier to using the drives on networks and multiuser computers. On the other hand, the high price—a minimum of over \$3,000 for retail models—is hard to justify for a single-user computer. Speed is also the main problem with multifunction drives, combined with the incompatibility of the different models.

---

#### **Future Directions**

The coming years will bring improvements to the current optical storage technologies, along with growth in sales and lower prices. We expect to see some improvement in the slow access times, possibly using the single-pass write technology found in Matsushita's multifunction drive. Today's drives use infrared laser diodes, but IBM has announced a laboratory breakthrough that involves light in the blue spectrum. The technology has promise for optical storage, where the shorter wavelength could double the density of data on the disk. ■





# An Overview of Displays

## In this report:

Products .....	2
Selection Guidelines.....	5
Trends and Issues .....	7

## Datapro Summary

With an annual growth of 18% to 20%, the microcomputer display market is expected to reach \$2 billion by 1993. Large system vendors such as Apple, IBM, and NEC continue to be market leaders since microcomputer system buyers usually purchase complete systems. Meanwhile, most technological advances in displays have occurred in the area of resolution.

## Technology Overview

- Large-screen displays continue to provide the working area needed for desktop publishing and engineering applications.
- Resolution capabilities increase as new video standards emerge. Most new displays are capable of the 1024- $\times$  768-pixel resolution provided by the 8514/A and XGA graphics standards.
- Flat screens, growing more popular, provide wider image spaces than curved screens.
- Many displays support more than one graphics standard, allowing them downward compatibility with a variety of standard video cards and software.
- Low emissions and antistatic coatings are features of many new models and upgrades.
- The most exciting new display feature is Moniterm's "Virtual Display" technology, which allows simultaneous, hardware-driven scrolling of the screen within a larger virtual screen, providing four times the normal viewing area. This is especially useful for large documents such as maps and engineering drawings.

## Technology Basics

Most technological advances in microcomputer displays have occurred in the area of resolution. Until recently, video adapter cards have only gradually increased resolution, but this trend has sped up. We predict that graphics standards will surpass 1024 x 768 resolution (provided by IBM XGA) in 1992.

New graphics standards now emerge from year to year. IBM's Color Graphics Adaptor (CGA) was the first color graphics standard, eventually surpassed by the IBM Enhanced Graphics Adaptor (EGA), which offered better resolution with the same 16 colors. Many displays still use both of these adaptors, but today's most widely accepted standard is IBM's Video Graphics Array (VGA), which uses 640 x 480 resolution and 256 colors. Super VGA, a three-year-old, non-IBM standard, provides 800 x 600 resolution with 256 colors, improving upon the VGA format. Despite the fact that many displays support Super VGA, many of the more sophisticated (more expensive) displays are now capable of much higher resolution than the current Super VGA and 8514/A standards. We expect IBM's new XGA format (1024 x 768, 65,536 colors) to be better received than Super VGA, as many monitors already support this soon-to-be standard.

—By *Todd R. Denton*  
Assistant Editor

## Vendors Respond to VDT Problems

In response to the problems from prolonged viewing of Video Display Terminals (VDTs), display, peripherals, and software manufacturers have developed products to reduce their effects on users. These products perform a variety of tasks, from eye relaxation exercises to the shielding of electromagnetic radiation (EMR), often divided into very low frequency (VLF) and extremely low frequency (ELF) radiation. Among these products are low-EMR displays, safety screens which reflect harmful radiation, and software packages designed to strengthen

eye muscles and reduce eye-strain. The following companies make products to help VDT users.

**Cornerstone**, a prominent display manufacturer, conducts extensive research and is very outspoken on the subject of VDT safety. Cornerstone's 19" DualPage display significantly surpasses the Swedish SEMKO standard for electromagnetic emissions, the world's strictest at present. The DualPage features a protective coil which reduces radiation by counteracting magnetic fields. The display is available as a \$150 option or a \$200 upgrade for current users.

**NoRad Corporation** is a producer of thin-film shield screens that fit many popular curved-bezel monitors. The NoRad SuperShield contains a high-resolution mesh, which is metalized to drain static and block radiation. It is grounded to discharge the VDT's static electric field, which, according to NoRad, attracts dust and ambient particles to the VDT operator's skin and eyes. The screen also markedly improves image contrast. It retails for \$349.

**Optical Coating Laboratory, Inc. (OCLI)** manufactures Glare/Guard antireflection filters, which use OCLI's patented High Efficiency Antireflection (HEA) coating, originally developed to eliminate glare on space shuttle windows and instruments. The products reduce glare by 99% and enhance screen contrast to reduce eyestrain. They also eliminate the perception of screen flicker,

maintain resolution, eliminate static and dust buildup, and reduce VLF/ELF radiation by 98%. The products range in price from \$59 to \$249.

**Sigma Designs Inc.** offers monitors which also meet the Swedish government's VLF emissions standards. The company is also developing a monitor technology that will meet new standards for ELF emissions expected to be set by Sweden next year. Current products include Low-VLF PageView, L-View, and SilverView models, which range in price from \$1,499 to \$3,695.

**Vision Aerobics Inc.** offers a self-titled software package designed to exercise eye muscles, relax and reduce eye tension, and reduce user fatigue. This package consists of a series of arcade-like eye muscle exercise games. The \$129 package includes a pair of 3-D glasses which are necessary for some of the exercises.

## Products

### CRT Displays

The cathode-ray tube (CRT) is by far the most popular display technology. Used also in most television sets, the cathode ray (electron beam) tube display consists of a vacuum tube enclosed in glass. One end of the tube contains an electron gun assembly, which includes a heating element, a cathode, a focus control, and a deflection yoke. The other end contains a screen with a phosphorous coating.

When heated, the cathode emits a stream of high-speed electrons. A strong positive voltage attracts the negatively charged electron beam, accelerating it towards the other end of the tube. On the way, the focus control and deflection coil steer the beam to a specific point on the phosphorous screen. When struck by the beam, the phosphor fluoresces (glows). This is the light that users perceive when viewing a display.

The beam sweeps the screen from left to right in lines from top to bottom, in a pattern called a raster. During its sweep, the beam strikes the phosphor wherever an image should appear on the screen. It also varies its intensity to produce different levels of brightness. A high-intensity beam produces a bright glow, and vice versa.

Since the glow fades almost immediately, the electron beam must continue to sweep the screen to maintain an image, a practice called redrawing or refreshing the screen. Most displays have a refresh rate (also called a vertical scan rate) of about 60 hertz, meaning that the screen is refreshed 60 times a second. Low refresh rates cause the

screen to flicker and contribute to eye strain; therefore, high refresh rates are preferred.

### Monochrome Displays

Monochrome CRT displays produce images of one color. The most popular color is green, followed by amber and white. The color of the display is determined by the color of the phosphors on the CRT screen. For example, a green phosphor emits a green light when struck by the electron beam. Some monochrome displays with white phosphors can support many shades of gray; for example, the Multi-Sync GS from NEC Home Electronics supports 64 gray shades.

Many of the popular 8-bit and 16-bit grayscale displays have been introduced in the past year. These sophisticated, high-resolution displays provide an ideal working environment for desktop publishing applications which do not involve color. High-end, 8-bit grayscale monitors provide 256 shades of gray, in comparison to 16.7 million shades in 24-bit color displays. Yet, most users agree that 256 shades is more than enough for their applications. Gray shades are shown as pixel rows on screen.

Monochrome displays usually cost less than color models, and users who perform mainly character-based applications—word processing, spreadsheet analysis, database management, and computer programming—will find them perfectly suitable. Monochrome displays designed for specialized applications such as desktop publishing and CAD/CAM, however, cost hundreds of dollars more than even color displays.

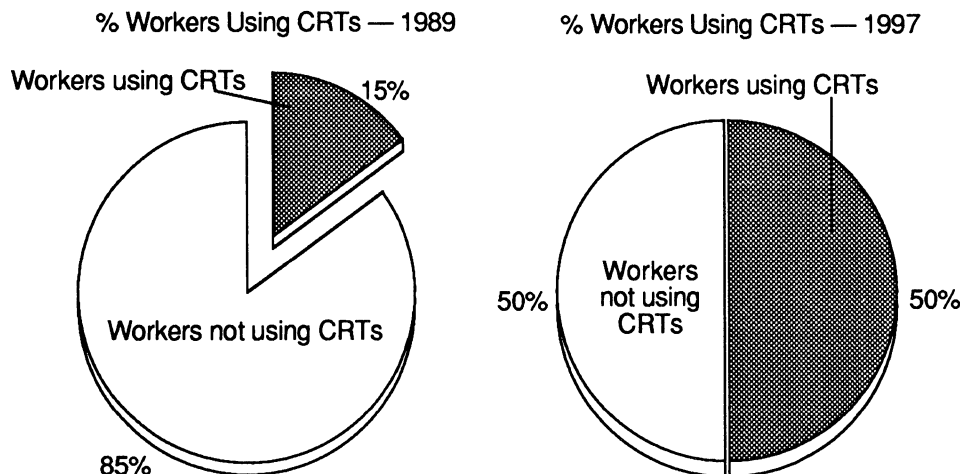


Figure 1.  
Increase of Workers Using  
CRT Displays

By 1997, 50% of the American workforce is expected to use CRTs.

Source: United States Department of Commerce.

### Color Displays

Color CRTs employ more sophisticated technology than monochrome displays, which accounts for their higher prices. While a monochrome picture tube contains one electron gun, a color tube contains three guns arranged in a triangular shape referred to as a delta configuration. Instead of green, amber, or white phosphors, the display screen contains phosphor triads, which consist of one red, one green, and one blue phosphor arranged in the same pattern as the electron guns. Red, green, and blue are the three additive primary colors of optics. They can be mixed to produce all other colors.

The focusing and deflection coils train the electron beam upon a phosphor triad, illuminating different phosphors to produce the desired images and varying the intensity of the beams to produce different colors. For example, illuminating the red and blue phosphor dots produces various shades of color between red and blue, such as purple, depending on the intensity of the beam.

A metal grille called a shadow mask is mounted directly in front of the screen. The shadow mask contains round holes that are precisely aligned with the phosphor triads in order to prevent the electron beam from illuminating the wrong triads. Shadow masks have one flaw: They tend to heat up as the electron beams pass through them, causing the holes to expand and the mask to vibrate. This distortion makes the electron beam illuminate the wrong phosphors and produces a blurred picture. Some vendors now use thinner, flatter shadow masks that neither expand under heat nor vibrate. The results are remarkable, although the technology is expensive.

Instead of a delta configuration, some displays use an in-line configuration in which the electron guns are arranged in a horizontal line. The display screen consists of alternating stripes of red, green, and blue phosphors, arranged horizontally to correspond with the electron guns. In place of a shadow mask sits an aperture grille containing long, vertical openings. This system allows more electrons to hit the screen, increasing the picture's brilliance. Sony and others use Trinitron picture tubes with in-line gun systems.

Although color displays cost more than monochrome displays, they are more versatile, allowing users to take advantage of color applications such as presentation graphics. The development of inexpensive color output devices,

including printers, plotters, and slide makers, should increase the sales of color displays and bring prices down.

Our survey showed the upper limit of resolution in available products to be 2560 x 2048 pixels, provided by Barco Chromatics. This same display topped our resolution charts last year as well. Screen sizes are increasing: Mitsubishi offers a 37" display. Super-large displays open a variety of new applications: Desktop publishers can create and manipulate actual-size posters and advertisements, and managers can easily make presentations to small groups.

### Alternative Technologies

Though virtually all currently available displays for desktop microcomputers use cathode-ray tubes (CRTs), flat-screen displays are necessary for portable microcomputers such as laptops and notebooks to conserve space and power. Developments in flat-screen display technology have been directly responsible for the astonishing popularity of portable microcomputers. Analysts predict that sales of laptops will increase from 832,000 in 1991 to 1.8 million in 1992 and 2.6 million by 1993.

CRT displays shoot an electron beam against a phosphor-coated screen to turn each pixel on and off separately, whereas flat panels select a pixel through an electric grid without turning on the adjoining pixels. Flat panels were once expensive and limited by low-resolution, almost unreadable, characters. The technology has improved considerably, providing a reasonable alternative to CRTs in both readability and price.

There are three kinds of flat-panel displays:

- Liquid crystal displays (LCDs)
- Electroluminescent displays (ELDs)
- Gas plasma displays (GPDs)

The LCD is a reflective display; ELDs and GPDs are light-emitting displays.

LCDs, the early choice for laptops, are still by far the most commonly used type of display in laptops. They are inexpensive to manufacture and consume less power than ELDs or GPDs. Because of their extremely low power consumption (about 80 milliwatts), the majority of today's laptops use LCD display technology, which produces dark pixels on a gray background. In an LCD display, each pixel

## Display Emissions Still Problematic

The hazards of prolonged display use continue to be a hot topic in the computer community. Studies show that users who spend extensive time in front of a video display terminal (VDT) are likely to experience fatigue, reduced productivity and possible health problems. The subject of VDT emissions has been thoroughly studied, and ongoing disputes and regulations are quickly reshaping the way we use computer workstations.

In the early days of VDT products, X-ray emissions were the chief concern of users. As medical problems became attributed to prolonged VDT use, the electromagnetic waves and radiation that VDTs emit were proven harmful to viewers, and regulatory agencies set out to eliminate emissions. However, according to NoRad Corp., a manufacturer of VDT shields, "Post-1971 video display terminals [no longer] present health risks from emissions of ionizing

(high-frequency) radiation, such as X-rays, because under federal law they are shielded by their manufacturers to reduce such radiation to essentially undetectable levels." Other manufacturers share NoRad's view that federal regulations have reduced VDT emissions to almost undetectable (unharmful) levels, attributing more tangible external problems such as glare, bad resolution, and dust buildup to decreased productivity and user health problems. Despite their claims, questions remain over whether long-term exposure to low-level electromagnetic radiation (EMR) emissions are still a threat to user health. The most questioned emissions problem today is EMR, which is broken down to very low frequency (VLF) and extremely low frequency (ELF) radiation.

Manufacturers of screen shield products say recent studies link VDT usage to serious health problems.

"Recent studies have established a highly probable link between VDT use and miscarriage, and studies in increasing numbers over the last decade have shown that similar low-level, non-ionizing EMR can cause chromosomal defects, alter the rate and quality of bone growth, increase fetal malformations in mice, and disrupt the basic genetic apparatus of cells, among other biological effects," states NoRad Corp.

Researchers suggest that CRT displays have by far the worst electromagnetic emissions. Alternate display types such as electroluminescent, gas plasma and liquid crystal displays (LCDs) are said to be much safer. The vast majority of today's desktop displays, however, are CRTs.

Though emissions regulations vary from country to country, Sweden currently has the most restrictive policies for VDT emissions. VDT manufacturers increasingly consider the Swedish SEMKO standard when designing new products. According to Sigma Designs, Inc., a leading U.S. VDT manufacturer, "Sweden has received international attention as the most progressive country in regulating magnetic field emissions." This

standard mandates induction (the rate of fluctuation within the magnetic field) to 25 milliTeslas per second (mT/s) and magnetic field emissions to 50 nano Teslas (nT)<sup>1</sup>. A news release on the standard states that Cathode Ray Tube (CRT) monitors produce X-rays when electrons are rapidly decelerated. Electron beams are directed at many points on the display screen by means of a magnetic field. However, measurements have shown that the radiation does not penetrate the monitor's glass tube.

Heavy-duty users will be pleased to know that emissions reductions were featured in more new monitors in 1991 than in any previous year. Manufacturers who wish to sell their displays in Europe are forced to conform. In the U.S., the state of California has progressive new legislation on display use, regulating VDT emissions and even time spent in front of the display. Many new displays also feature anti-glare coatings, and a variety of add-on glare and radiation-reducing screen shields are currently available.

<sup>1</sup> A tesla is the unit of measurement for the strength of a magnetic field

is a twisted nematic (loosely structured liquid crystals) cell. The cells consist of liquid crystals on glass plates that are oriented at 90-degree angles and sandwiched between two polarizing filters, thus forming a matrix. An electric current causes the pixels to change from a polarization that transmits light to one that blocks light. However, an LCD's external light source determines character intensity, which creates a problem in inadequately lit environments such as trains or planes.

To improve readability, LCD manufacturers have turned to a "supertwisted" LCD. Supertwisted crystals change the polarization to 270 degrees instead of the usual 90 degrees. The 270-degree orientation provides better contrast than a typical LCD. To further improve screen legibility, laptop manufacturers have put an electroluminescent light source in the back of the display to further sharpen characters in dim lighting. As a result, LCD laptops are more popular than before.

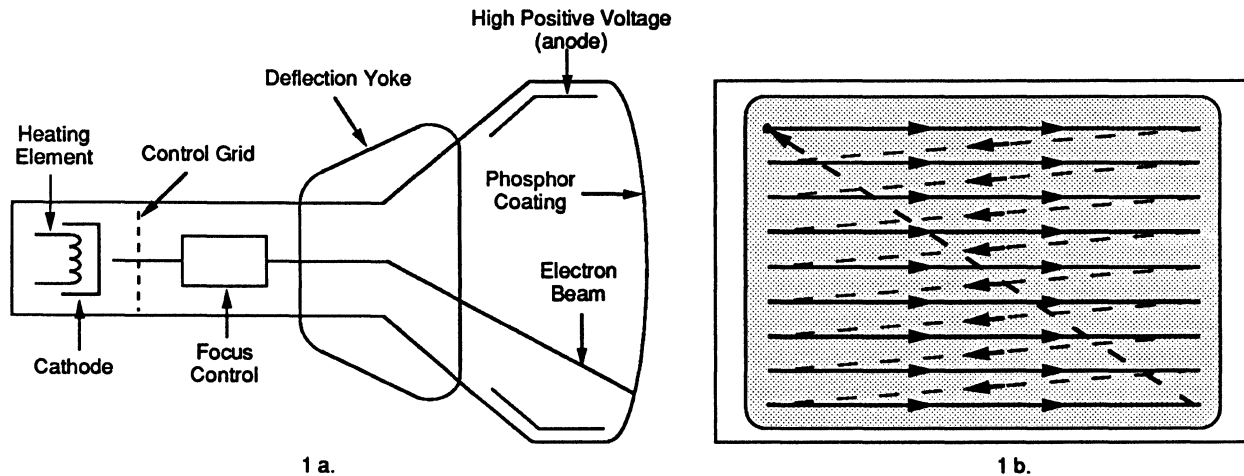
An ELD consists of a layer of manganese-doped zinc sulfide sandwiched between two electrodes. When voltage is applied, the manganese atoms are excited and emit a

green-yellow light. As a result, ELDs are difficult to manufacture, consume more power than LCDs, and require higher voltage drive circuits to power the display. ELD display readability is superior to the LCD's, however.

The GPD uses an inert gas (usually a combination of neon and argon) sandwiched between an x-axis panel and a y-axis panel to form a matrix. A pixel is produced by charging wires that intersect at the specific x-y coordinates. When the wires are charged, the gas (called the Penning mixture) around the intersection glows bright orange. GPDs require AC power to maintain the glow at a constant level. GPDs draw less energy than an ELD, and images disappear from the screen in significantly less time than with the LCD. This produces a quicker, sharper image that does not depend on any external light source.

Increasing numbers of displays manufacturers are working to bring flat-screen display technology to desktop systems. Recent developments include flatter screens that produce less glare, and flat-panel displays that hang against a wall like a picture frame. Desktop displays using these technologies will become widely available in the next few years.

Figure 2.  
CRTs



CRTs produce images by generating a beam of electrons and focusing it onto a phosphorous-coated screen, which fluoresces or emits light at the point of contact. Figure 2b. The beam sweeps the screen from top to bottom and left to right in a pattern called a raster scan.

## Selection Guidelines

### Screen Resolution

Resolution is the amount of detail on screen that the display is capable of rendering. This quantity is expressed in the number of horizontal and vertical picture elements (called pixels for short) contained in the screen. The greater the number of pixels, the more detailed the images produced. The resolution required depends on the application. Character-based applications require little resolution, while graphics-intensive applications such as desktop publishing require a great deal.

### Dot Pitch

The term pixel or picture element refers not to a specific device or screen phosphor but to the smallest unit that the CRT uses to produce an image. In a monochrome display, the picture element is a screen phosphor, but in a color display, the picture element is a phosphor triad. This difference raises another consideration called dot pitch, which applies only to color displays. Dot pitch is the distance, measured in millimeters, between phosphor triads. Screens with a smaller dot pitch contain less distance between the phosphor triads; as a result, the picture elements lie closer together, producing a sharper picture. Conversely, screens with a larger dot pitch tend to produce blurry images. Small dot pitches are a technological challenge and are ultimately more expensive for the buyer. Most displays currently have a dot pitch between 0.31 and 0.28 millimeter, though some new monitors feature an 0.26-millimeter dot pitch. The smallest currently available dot pitch is 0.25 millimeter.

### Interlacing

Some displays support interlaced resolution. In noninterlaced (conventional) mode, the electron beam sweeps the screen in lines from top to bottom, one line after the other, completing the screen in one pass. In interlaced mode, the electron beam also sweeps the screen from top to bottom, but in two passes, sweeping the odd lines first and the even lines second. Each pass takes half the time of a full pass in

noninterlaced mode; therefore, both modes refresh the entire screen in the same amount of time. This technique allows faster screen redrawing and more stable images. Broadcast television transmissions in the United States are interlaced.

The drawback is that interlacing depends on the ability of the eye to average two nearly identical lines separated by a gap into one solid line. While television pictures have nearly identical lines, microcomputer graphics do not because of the difference between television and computer-generated images. Since unmatched lines produce flickering and shimmering images, manufacturers of graphics equipment reserve interlacing for generating extremely high-resolution images (over 1024 x 768 pixels), such as those produced by the highest mode of the IBM 8514/A and XGA graphics cards.

### Graphics Standards

Like all output devices, displays require a source of input. Input signals for displays originate from graphics hardware that resides inside the system unit of the microcomputer. A few systems, notably the IBM Personal System/2 product line, contain graphics circuitry on the motherboard. The majority of systems, however, use a separate circuit board that fits into an expansion slot within the system unit. Expansion boards that produce video signals are called video cards or graphics cards.

Most graphics formats gain standardization through the Video Electronics Standard Association (VESA); however, some become widely used without this certification. The majority of video cards adhere to the following industry standards:

### XGA

IBM recently joined forces with VESA's XGA technical committee to promote the Extended Graphics Array (XGA) as a new video standard. This proposed standard, providing 1,024- x 768-pixel resolution, has quickly become a popular capability for graphics monitors. XGA offers much better picture quality and higher speed than

VGA, as well as several features (including faster device drivers) not found in 8514/A. (See sidebar for further information.)

### 8514/A

IBM's answer to the non-IBM Super VGA, 8514/A provides a maximum resolution of 1,024 x 768 resolution and 256 simultaneous colors from a palette of 256,000. This standard is not nearly as popular as IBM's previous standard (VGA).

### Super VGA

This non-IBM standard improved upon IBM's VGA standard with higher resolution (up to 800 x 600 pixels) and the same colors as IBM's VGA (256 colors from a palette of 256,000).

### VGA (Video Graphics Array)

Today's most widely used standard, VGA was introduced by IBM in 1987 as the base video for all PS/2 systems. VGA provides a maximum resolution of 640 x 480 pixels, 256 simultaneous colors from a palette of 256,000.

### EGA

IBM's Enhanced Graphics Adapter (EGA) surpassed CGA, providing a maximum resolution of 640 x 350 pixels and 16 simultaneous colors from a palette of 64.

### CGA

IBM's Color Graphics Adaptor (CGA) was the first color graphics standard, offering 640- x 200-pixel (columns by rows) resolution in monochrome, or four colors with 320 x 200 pixels.

### Macintosh II

Apple's video card, providing 640 x 480 resolution and 256 colors from a palette of 16.7 million.

In addition to providing video signals, some circuit boards also perform graphics processing, a task usually left up to the computer itself. By processing graphics, the board increases the speed with which the graphics are produced on screen. These boards are called video coprocessor cards or video controllers, and the IBM 8514/A graphics card, another current standard, is an example. Though 8514/A has become a standard, it has not replaced VGA, and will most likely be replaced itself by XGA as the standard of choice.

Increasing numbers of microcomputer displays support more than one graphics standard, allowing them to operate with a variety of standard video cards and software. For example, many Super VGA-compatible displays also support VGA, EGA, and CGA. These displays are often called multiple frequency displays. See comparison columns for examples. Different vendors refer to their multiple frequency displays by many different names, including multisync, multifrequency, multiscan, auto synchronous, and auto tracking. Users should make sure that the display they are considering supports their desired standards.

Displays designed for specialized applications such as desktop publishing and CAD/CAM support no industry standards but rely on proprietary technology requiring special (proprietary) video cards. For example, the RasterOps ClearVue/II Monochrome Graphics System requires a RasterOps monochrome graphics card. The terms *graphics system* and *display system* refer to a product consisting of a display and graphics card to be used together.

## IBM's Extended Graphics Array (XGA)

In November 1990, IBM announced the Extended Graphics Array (XGA) video subsystem for the new high-end PS/2 models. XGA was designed to provide higher speed VGA (supporting existing VGA modes) and extended function 1024 x 768 graphics resolution. The XGA Display Adapter/A, introduced in conjunction with the XGA subsystem, is a Micro Channel upgrade for PS/2 models without video on the system board. XGA has several advantages over previous formats:

- XGA maintains 8514/A DOS application program compatibility at the adapter interface level, which means users can still run Microsoft Windows, OS/2 Presentation Manager, and programs like AutoCAD and WordPerfect written for the 8514/A.

- XGA's two mutually exclusive operating modes (VGA and Extended Graphics) provide a great deal of versatility and compatibility with existing VGA applications.

- To improve upon VGA's performance, IBM employed video random access memory (VRAM) technology, providing dual-ported, multiprocessing capability (simultaneous video buffer updates and screen refresh).

- XGA's new 16-bit-per-pixel *Direct Color* function delivers 65,536 colors at 640 x 480 resolution.

- In 32-bit BusMaster mode, XGA can access both system memory and video display buffer memory. Images can be constructed in system memory for display later, at which time the bus master transfers the images from system memory to the video buffer for immediate display.

- A hardware sprite (stable cursor) overlays background video data without changing, saving or restoring that data, increasing performance and screen quality of front-screen applications.

Some systems (e.g., the NEC MonoGraph System) incorporate gray scale emulation, an added feature that allows them to operate with software compatible with recognized industry standards, in addition to desktop publishing or CAD packages.

### Desktop Publishing Displays

Desktop publishers and graphic artists today have a wide variety of applications-oriented DTP displays to choose from, with many different sizes, shapes, and compatibilities. Desktop publishing (DTP) applications have become the strongest force behind the evolution of display technology.

Larger, high-resolution screens geared specifically for page layout and graphic applications have been the most innovative products this year. Radius is a true leader in this area, providing double-page screens, and even a pivoting function to change a document from portrait to landscape mode. E-Machines and SuperMac are also heating up the competition with large screens and competitive

prices. Screen sizes are now big enough to create actual-size posters and advertisements.

Color displays are also making strong headway in DTP. Color graphics can be input and manipulated, and multi-color, camera-ready pages can be output to various types of color printers. Radius has emerged as an innovator in the color market also. The Radius PrecisionColor Calibrator enables users to select gamma correction and color temperature settings, making output colors exactly as they appear on screen. Monochrome models are still strong in DTP, providing more gray levels for graphics. Displays with the What You See Is What You Get (WYSIWYG) feature can produce the image of a full page at actual size.

### Configuration

The Apple Macintosh line has become the standard for desktop publishing, and most publishing displays are designed for this platform. The Macintosh Plus and the Macintosh SE feature a nine-inch, monochrome display built into the system unit. However, users requiring a larger screen or different features may attach an auxiliary display to the machine. The Macintosh II product line features one color and three monochrome displays, including the 1152-x 870-pixel, 21" Two-Page Monochrome Monitor, a standard for double-page displays. Apple's new 24-bit graphics card has also set a standard for publishing.

### Orientation

The screen on a portrait or full-page display is higher than it is wide, enabling users to view an entire page of a document or a sizable portion of programming code. A landscape or dual-page display is the exact opposite of a portrait display, suitable for viewing two pages at once. Although most general-purpose displays are landscape, publishing displays with this orientation are larger and wider. Landscape models usually cost more than portrait models.

### Limitations

For all their functionality, publishing displays have major shortcomings. Most of them include software drivers only for desktop publishing software, requiring users to purchase another display for general-purpose tasks. Their large screens produce a great deal of glare and take a long time for the CRT to refresh. Finally, these displays cost hundreds of dollars more than conventional models.

## Trends and Issues

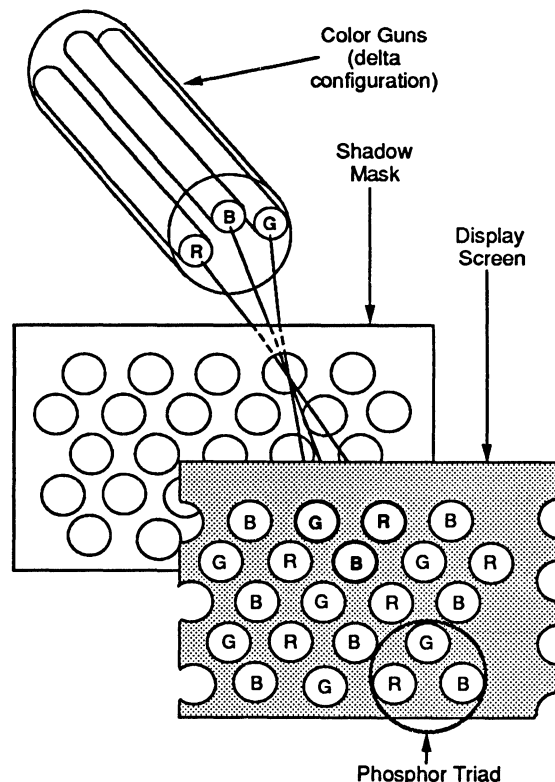
### Large Screens

Larger screens are constantly in demand for desktop publishing, graphics, and engineering applications. This category includes portrait-oriented models, which can display an entire page of text or programming code, and landscape-oriented models, which can display two full pages at once. Large screens produce glare, have comparatively slow refresh rates, and require much higher resolutions than normal displays to achieve the same image sharpness. These displays are still very expensive, compared with conventional models.

### High Resolution

Resolution is another area of vast improvement. New graphics adapters are providing screen resolutions that were unheard of only a year ago. Graphic designers and desktop publishers are ecstatic about the many new 1,024-x 768-pixel displays and graphics adapters. E-Machines

Figure 3.  
Color CRTs



*Color CRTs utilize three electron guns to target a phosphor triad, which consists of one red, one green, and one blue phosphor, representing the primary colors. Beams from the guns illuminate the phosphors, mixing the colors to produce other colors. The shadow mask prevents the beams from illuminating the incorrect phosphors.*

currently markets four displays with a resolution of 1,280 x 960 pixels. Electrohome Ltd. has four displays with 1,280-x 1,024-pixel resolution, all of which are compatible with eight different adapters. Vectrix Corp. now makes a 1,600 x 1,280 display that is compatible with IBM MCGA and up to 1,600-x 1,280-pixel adapters.

### Standardization

Another force driving display technology is the push for standardization in the microcomputer industry. In 1988, nine vendors of graphics products (ATI Technologies, Genoa Systems, Orchid Technology, Renaissance GRX, STB Systems, Tecmar, Headland Technology [formerly Video Seven], Western Digital Imaging/Paradise Systems, and NEC) formed the Video Electronics Standards Association (VESA). Led by NEC Home Electronics, VESA is accepted by the industry as a standard-setting organization. The group resolves compatibility problems and attempts to eliminate confusion in the market. Currently, VESA is exploring the confirmation of IBM's proposed XGA standard.

### The Role of Television

Television plays a significant role in the development of display technology. Since television monitors and microcomputer displays are so closely related, advances in television often apply to displays. One such advance is high-definition television. Invented by Japanese firms, HDTV

uses twice as many horizontal scanning lines as conventional television transmissions, producing wide, crisp, high-resolution pictures proportioned like those of a movie screen. Many products besides television sets will support HDTV signals, including medical imaging equipment, defense radar systems, and, of course, microcomputer displays.

### **Touch Screens and Projection Displays**

Another trend is the proliferation of touch screens and projection panels. Touch screens are pressure-sensitive devices that enable users to select and manipulate functions and data simply by touching the screen with a fingertip. These products are popular in industrial environments, where workers require easier and faster means of operating their microcomputers than keyboards. Vendors of touch-screen equipment include Applied Digital Data Systems (ADDS), Elographics, and MicroTouch Systems. Projection panels use LCD technology and attach to the display or directly to the microcomputer, allowing users to display computer graphics on an overhead projector. Vendors include Eastman Kodak, Electrohome, and Sharp Electronics.

### **Page-White Displays**

Page-white displays are by no means a new technology. Yet, display vendors are now responding to users who wish to work on a screen that provides a white background and black characters (rather than the reverse), just like standard business paper. This trend makes perfect sense, given the advancements in handwriting recognition and pen-based computing. The white backgrounds are easier to read for some users who spend long periods of time in front of the display.

### **Emissions and Ergonomics**

Emissions are now, more than ever, a hot topic for displays. The Swedish SEMKO government organization set strict standards for that country's video display terminals (VDTs), and manufacturers who wish to sell in Sweden are forced to conform. The state of California has also pushed hard for new regulations on displays, including time spent in front of the display and VDT emissions.

Though VLF radiation is the chief concern for emissions activists, quite a stir has been raised over display screen glare. Studies have shown that glare is a main cause of discomfort for display users, many of whom operate under fluorescent light. Glare makes the eyes work harder to pick out characters on screen, and has been proven to lead to decreased vision. Many vendors now provide antiglare coatings on their displays. These coatings are either etched into the glass screen or applied externally. Flat screen displays are said to reduce screen glare. See sidebar on emissions for further information.

### **Flat Screen Displays**

Flat screens are growing more popular, as they reduce glare and provide a wider image space. Flat-screen vendors note the higher image precision that flat screens provide. Commonly used, flat-tension masks provide bolder and brighter images than curved screens. Sharp Electronics Corp.'s recently debuted 3" thick flat-screen television may be a harbinger to future computer displays. Market leaders now provide flat-screen models.

### **Product Innovations**

Display technology is one of the most innovative sections of the computer industry. Revolutionary products include the Radius Pivot, which provides immediate conversion from portrait (vertical) to landscape (horizontal) screen formats with a 90-degree screen rotation, without a need to reboot system software. Radius recently added color to its Pivot product line. Zenith's new flat tension mask screen provides excellent resolution in a much smaller space than conventional displays. One new monitor features Moni-term's proprietary *virtual display* technology, which allows simultaneous, hardware-driven scrolling of the screen within a larger virtual screen, providing four times the normal viewing area. CAD-ready displays provide three-dimensional display capability in true perspective or in orthographics projection, providing a variety of applications for engineers and physicians. All of these advancements enable microcomputer displays to convey a greater amount of information in a form that conserves space and promotes ease of use. Users may confidently expect further such advancements to emerge in the future. ■



# An Overview of Scanners

## In this report:

Image Types.....	2
Emerging Technologies .....	4

## Datapro Summary

The market for scanners continues to grow and diversify. Microcomputer users are becoming more interested in the ability of scanners to meet their input needs for desktop publishing and other office applications. Scanner applications differ according to a scanner's physical ability and software compatibility. Many application software programs are currently available; however, many of the scanners marketed today come bundled with software packages. Of course, there are pros and cons to this setup. Pros include bundled prices and easy setup. Cons include wasted software if different software is desired. Many scanner buyers, therefore, look to unbundled scanners or those with optional software. It makes more sense for these users to buy a software program that suits their scanning needs and then purchase a compatible scanner.

## Technology Overview

- Basic scanner components include the document feeder, sampling device, controller, and software (which may be bundled or separate).
- Scanners can be divided into four basic physical types: flatbed feeders are manual but usually accept large originals; sheet feeders are automatic but take only fixed paper sizes; printer-mounted and handheld scanners are slow but inexpensive.
- A scanner's sampling device, usually a CCD (charge-coupled device), passes scanned information to a controller. A high-resolution sampling device yields a high-resolution scanned image.
- Whether located in the scanner or the host computer, the scanner's controller coordinates various parts of the scanner and interprets electrical signals from the sampling device.

- A scanner performs with the help of software, typically either an external application or resides on a card. Sometimes, it is both.
- Scanners differ widely in the types of images they can capture. In ascending order of complexity, image types include the following: line art (completely black and white, such as drawings and text); half tones (simulated gray); grayscale (true gray); and color (the most difficult to capture accurately).

## Technology Basics

Though scanners vary widely in price and size, they have the same basic components and all process one or more of the basic image types (line art, halftones, and color).

### Scanner Components

All scanners have four basic components: a document feeder system, a sampling device, a controller, and software.

### Document Feeder

For scanners, the source of most images is paper—either loose sheets or bound volumes. To scan the image, a sampling device that senses light must be exposed to the

—By *Todd R. Denton*  
Assistant Editor

page. How this takes place varies with the scanner, but most use one of two methods:

In *flatbed* models, a page lies on a glass plate while the sampling device moves down the length of the page, just as in most copiers. Flatbed scanners are ideal for graphics because they allow great flexibility—the original page can be a nonstandard size or bound in a book.

*Sheetfed* scanners move the paper over the sampling device, which is under a narrow glass window. This is ideal for multipage documents (typically text) since the scanner can automatically feed pages. Sheetfed scanners usually cost less than flatbed scanners, so they are often used for graphics, even if this means scanning a photocopy in place of the bound or oversized original (sheetfed scanners usually cannot accommodate large paper sizes or bound material).

A number of vendors have added automatic document feeders (ADFs) to their offerings this year. ADFs are in high demand, and with good reason. Because copiers, fax machines, and printers all use sheet feeders, ADFs will be essential components of multifunction units as they develop.

A growing number of scanners have other input configurations. *Handheld* scanners are contained in a device that the user rolls up, down, or across the page, much like a microcomputer mouse. Handhelds used to be plagued by image distortion problems (caused by uneven motion of the user's hand), but this problem is being solved through software and user practice. Handheld scanners provide portability and can scan information from odd surfaces, such as labels pasted on boxes, business cards, etc.

*Overhead* scanners, which look like overhead projectors, have no feeder system at all. Instead, the sampling device faces down on the page from above, much like an overhead camera. Overhead scanners can scan bound and even three-dimensional originals, such as circuit boards or ceramic tiles. These scanners are used primarily for specialized applications.

### Sampling Device

If a scanner is, as one vendor claims, the "eye of the computer," then the sampling device contains the eye's rods and cones. All sampling devices have certain things in common. First, they all detect the light reflected from the page and convert it into digital data that computers can interpret. Second, they all do this by dividing the page into a grid of dots. More dots per inch (dpi) means a higher *resolution* and a finer image. With the exception of high-resolution models used in graphic arts, most scanners have a maximum resolution of 300 or 400 dpi, which is considered satisfactory for office applications. However, resolution will always be an area of potential growth for scanners.

In order of prevalence, the two main sampling devices used in today's scanners are the charge-coupled device and the laser.

**CCD.** The most common sampling device consists of a semiconductor chip with an array of light-sensing elements. During scanning, a lamp illuminates the page. A lens reflects and concentrates the light onto the chip. As the feeder system moves the paper (or in a flatbed scanner, the entire sampling device), changes occur in the pattern of reflected light. For each dot pixel in the grid (a square measuring 1/300 by 1/300 inch in most cases) the CCD returns a certain voltage, the highest voltage corresponding to white and the lowest to black, with grays in between.

**LED.** Some handheld or sheetfed scanner sampling devices are classified as light-emitting diodes (LEDs), though

these are really a type of CCD scanner. Light that illuminates the page comes from an LED, which is lighter, smaller, and requires less power than a lamp bulb.

**Laser.** In some scanners, the light source is a laser. The laser beam sweeps across the page and reflects onto a photocell. The information received is the same as in other scanners—a grid of darker or lighter dots represented by voltages—but resolution is typically higher. Many graphic arts scanners use laser sampling devices because of their higher resolution.

### Controller

A scanner's controller converts raw scanned data (a set of voltages emitted by the sampling device voltages) into a digital signal comprising binary bits (zero for a black dot and a 1 for a white dot) and then passes the stream of bits to the computer, along with instructions for where one row ends and the next one begins.

The controller passes this information on to the computer for processing. It also controls the other parts of the scanner, turning the lamp and feeder system on and off. In the past, controller circuitry resided solely inside the scanner, but now functions may be performed by an expansion card that plugs into the computer. In either case, a cable (often using the SCSI interface) links the scanner to the computer.

### Software

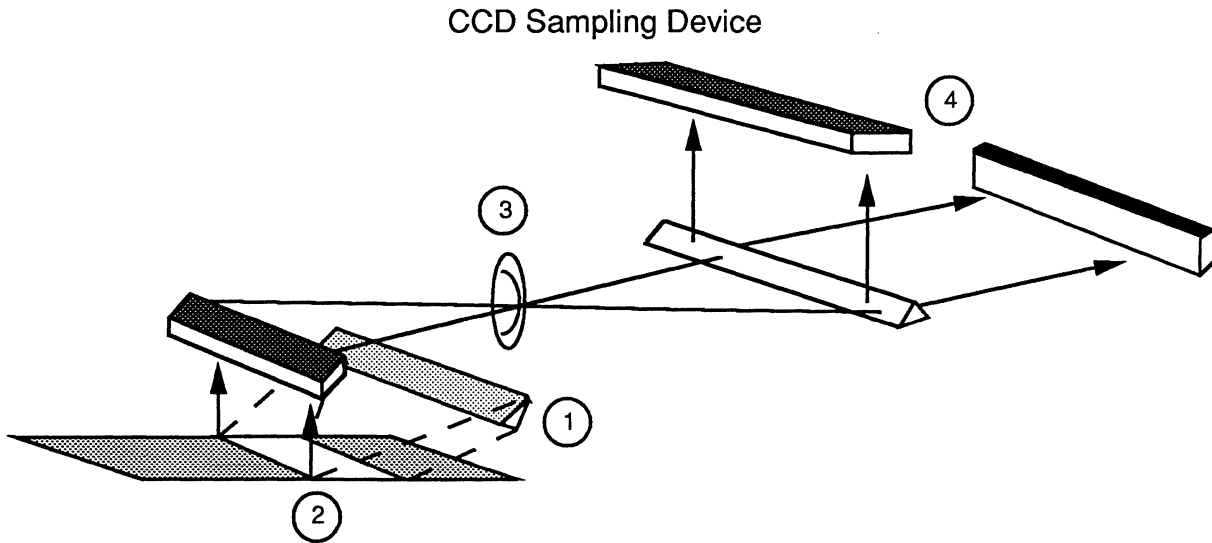
Once the computer receives the image, it must store it in a form that can be recalled and manipulated. Scanner software brings applications programs to the scanner. In graphics scanners, the image can be incorporated into a drawing program, an image manipulation program, or directly into a publishing program that combines images with graphics. On the newest scanners, the user controls the scanner through the software, defining the area to be scanned and the brightness and contrast settings.

Scanner software may reside in four main places: in the CPU, in the operating system, in the scanner itself, or as an external application to the scanner. Compression/decompression software normally resides in the CPU. Scanner driver functions typically reside in the operating system. Editing is normally an external application; however, it may be bundled in flatbed scanners. Optical character recognition (OCR) software currently exists as both a bundled and an external application—we found that about half of the scanners in our survey this year have bundled OCR, a trend that should continue in the years to come. Image processing scanners differ from OCR in that their software resides primarily as an external application. We expect image processing to increasingly become a function of the operating system in years to come.

### Image Types

Until now, we have discussed images in terms of black-and-white dots that the scanner translates into bits. This is fine for the simplest type of image, called *line art*. Line art includes graphs, charts, drawings, text, and anything else that is entirely black and white with no shades of gray. Yet many more image types exist, such as continuous tones in photographs or halftones (shades of gray). In fact, every scanner can detect grays, but in line art scanners, the controller interprets light grays as white and dark grays as black. More sophisticated graphics scanners can pass the grays on to the computer. Color is another image type, although typically not a requirement for document image processing for business applications.

Figure 1.  
CCD Scanner



Most scanners now use a charge coupled device (CCD) to receive images. In the illustration above, a lamp (1) reflects light from a document (2) onto a set of mirrors and lenses (3) that direct it to two CCD chips (4). The CCDs convert the image into electrical impulses, which the scanner puts into digital form and passes on to the computer.

### Halftones

Halftones are an efficient way to replicate shades of gray. A close look at a newspaper photo reveals a barely visible grid of dots, known as halftone dots. Throughout the photo, the dots are spaced evenly at about 65 to an inch. In the dark parts of the photo, the dots are larger; in the light parts, they are smaller. This creates the illusion of gray on a printing press that has only black ink.

Graphic artists traditionally produced halftones by photographing the original photo through a special, transparent screen. Scanners obtain the same effect by dividing the page into a coarse screen, usually with a number of dots that divides evenly into its maximum resolution. For example, if the scanner has a maximum resolution of 300 dots per inch (the most popular resolution currently), it divides each horizontal inch evenly into 75 cells. Each cell contains four dots ( $300 \div 4 = 75$ ). During the scan, the scanner *rearranges* the four dots into an approximation of the halftone dots used in a newspaper. The result looks like gray when viewed on a display or printed on a laser printer or typesetter.

Though the resolution is low, halftones look like photos, because the original photo (or other continuous tone) usually has gentle gradations of gray that run together. Most graphics scanners can distribute halftone dots in a variety of patterns. Each pattern is best suited to a different kind of image and gives varying results, depending on how the image is to be used. Unlike line art, halftone images cannot be modified extensively by the scanner—they can only be cropped and printed.

### Grayscale

Grayscale scanners take the halftone process a step further. In grayscale, the scanner records each dot sampled, not as black or white, but as a value that tells its shade of gray. All scanners use the following pattern, which can be extended to much larger numbers:

2 bits = 4 levels of gray  
4 bits = 16 levels of gray  
6 bits = 64 levels of gray  
8 bits = 256 levels of gray

Thus, a four-bit-per-pixel scanner could indicate that a dot was 8 on a scale of 16 levels of gray, or halfway between black and white. Since the information is stored as a number instead of a halftone pattern, it can be displayed at the full resolution on-screen, or converted to a halftone and printed. The advantage is that, unlike halftones, gray scale images can be made lighter, darker, edited, or otherwise modified at any time, even if the original is no longer available. They can be printed and displayed in different sizes too.

The disadvantage to gray scale images is the large amount of data space needed: each square inch of line art at 300 dpi contains 90,000 dots. At 4 bits per pixel (16 levels of gray), the same 90,000 dots take up 45,000 bytes. Increase the grayscale to 8 bits per pixel, and the image takes 90,000 bytes. A full letter-size page contains at least 70 square inches, or about 3.2M bytes of data at 8 bits per pixel.

Such large amounts of data take a long time to transmit and soon fill a hard disk. The scanner's controller gets around this problem by *compressing* the image by up to 20:1. Even so, grayscale images are unwieldy.

### Color

The hardware of color reproduction is a long way from perfection. A color photograph input on a color scanner will not look the same when it is output to a color printer. This is because the input, manipulation, and output of color images involves several difficulties. Color scanners separate the images they scan into three color *channels*. Depending on the model, these are either red, green, and blue, or cyan, magenta, and yellow (the colors used in a printing press). The sampling device is usually a CCD with color filters or color lamps. Color images have disadvan-

tages: they take up at least triple the disk storage space of monochrome images and are expensive to reproduce on paper because of the high cost of color printers. Still, scanned color images can be cropped, sized, edited, and otherwise manipulated. The most sophisticated color scanners can capture an original from a slide, alter it in ways that would be difficult or impossible in a darkroom, and then record the result on a film recorder.

In a basic sense, color is problematic because people see color in different ways. What looks blue to one person may look green to another, and so forth. Also, graphics artists want the ability to manipulate colors in every way imaginable. All that hardware and software developers can do is try to invent programs that allow incredible flexibility in color input, manipulation, and output. One of the scanner industry's goals is to have scanners producing larger *color spaces* (ability to handle more colors; 10- to 12-bit CCD per plane) than other peripherals in the color process by 1995. By the year 2000, the industry hopes to standardize color technologies and have color calibration as a dedicated operating system function.

## Emerging Technologies

### Multifunction Products

Multifunction or *hydra* products are a hot topic in the scanner industry right now. Office products such as scanners, fax machines, copiers, and printers are destined to be combined into multifunction units. The reason is simple: office space is an expensive commodity, and office machinery takes lots of space. The demand for single-unit machinery that will perform a variety of office functions has arisen. It is no surprise that scanner manufacturers have been among the first to experiment with these multifunction products.

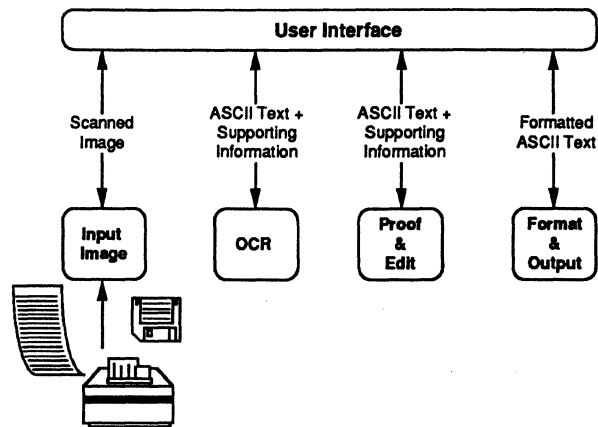
Multifunction products will not suit big business, where office products like copiers and scanners may be used constantly by a large staff. However, these products are needed in small businesses where few employees need a wide variety of functions.

We do not expect that multifunction units will replace single-function units and take over the industry, but they will be manufactured and will find their niche. Technologically, scanners, fax machines, copiers, and printers are not dissimilar. Moreover, the technology to combine such products into multifunction products is here now.

This year, we list a multifunction product in our comparison columns—*Scanafax*, made by Bona International Systems, which uses thermal (fax) paper. This product combines scanning and facsimile technology, and according to Bona, it will incorporate plain-paper printing and copying functions by the Summer of 1991. Plain paper will be crucial to its success. Typically, the average cost of sending a page by facsimile is 10 cents, versus the 29-cent postal rate, so facsimile is a much more practical alternative, especially if plain paper can be used.

Several obstacles will need to be passed before multifunction units will be widely available. For example, multifunction products will require sophisticated software, which could be developed quickly but will take time to refine. We expect multifunction units to eventually be accepted as a useful office tool, but not for some time yet.

Figure 2.  
The OCR Process



The diagram above illustrates the basic components of an OCR process: image input, actual character recognition, error correction, format, and output.

Source: Calera Recognition Systems, Inc.

### Scanners in the Office

Scanners help reduce paper in the office by converting printed information to an electronic format. However, though it would be ecologically beneficial, the world will not soon give up its paper. For this reason, scanners will not soon reach their full potential as a vehicle to convert information's format.

However, scanners still have tremendous potential in the office environment. The ability for scanners to handle high-resolution graphics in large-scale formats still must be refined. OCR developers need to find more efficient and faster ways of recognizing text and figures in all their shapes, sizes, and colors. Color scanners must be standardized and integrated into much more sophisticated systems where color integrity can be preserved from scanning to output.

Multifunction peripherals will incorporate scanning with copying, facsimile, and printing technologies. These have the potential to become the best answer for small business office needs.

### Scanners in the Home

Judging from the tremendous success of handheld scanners already, it is inevitable that this technology will bring the larger world of scanning to home users. As prices drop and units become more compact, they will become a sensible option for use in the home. The proliferation of personal computers in the home will contribute greatly to this process, but the need to input text and images will outweigh that fact. Household records will be input by scanners. Students will use scanners to input information for homework assignments. Even photo albums will be input by scanners and stored electronically.

We expect to see multifunction units, which will include scanners and facsimile, to make a strong push into the home by the year 2000. Considering the importance of home video today, we also expect video capture to become a key function of scanners in the home. ■

# An Overview of Expansion Cards

## In this report:

Selection Guidelines.....	5
Emerging Technologies.....	6

## Datapro Summary

Expansion cards are the means by which users tailor microcomputers to fit their individual needs. Many manufacturers, uncertain of specific user needs, include expansion slots on the system board inside the computer. This allows users to add features as their needs dictate. The result is lower system prices and greater flexibility for the user. Users are neither locked in to purchasing features they don't need nor prevented from purchasing features they do need.

## Technology Overview

### Technology Highlights

- Many new expansion card types in the market
- Many new cards have multimedia applications
- Most cards (55%) are IBM PC/XT/AT-compatible
- 44% of cards use ISA bus
- Most cards (53%) are full-length in size
- 16-bit architecture is still the industry favorite
- Speed of accelerator cards is increasing

Given the immense size of the microcomputer hardware market, it is not surprising that the expansion card market is a burgeoning one. Over a hundred vendors provide expansion cards with a full gamut of features. Users can purchase cards to increase user memory (RAM), enhance graphics, increase display resolutions, add processing power, add print buffers, control peripherals, and more. The available expansion card applications, like microcomputers themselves, seem limited by imagination alone.

Microcomputers enable third-party manufacturers to tie into the computer at

its most basic level; and if microcomputers are to expand or adapt, it is usually via expansion cards.

## Technology Basics

### Accelerator Cards

Today's microprocessor technology enables microcomputers to process information faster than ever; the fastest now run at 33MHz. However, processor speed remains a very important factor for time-sensitive applications and heavy workloads. Applications software has become more complex, and data files have become larger, demanding increased performance. Windowing applications and graphical user interfaces now require tremendous processing power.

The two main types of accelerator cards are the replacement processor and the residential processor. Replacement processors substitute for the original system processor, while residential processors work in tandem with the original. A residential card is easier to install because it fits into a ready-made expansion slot. Residential cards are also a more flexible alternative because they are more numerous and varied. Cards that combine both features are called hybrids.

An accelerator card upgrades the processing power of a microcomputer. For example, an accelerator card for the IBM PC XT may increase the performance of a

**Table 1. Video Graphics Standards**

Standard	Resolution (pixels)	Simultaneous colors	Colors in palette
Hercules Graphics Card (HGC)	720 x 350	N/A	N/A
IBM Color Graphics Adapter (CGA)	320 x 200	8	16
IBM Enhanced Graphics Adapter (EGA)	640 x 350	16	64
IBM Professional Graphics Controller (PGC)	640 x 480	256	N/A
IBM Video Graphics Array (VGA)	640 x 480	256	256,000
Super VGA	800 x 600	256	256,000
IBM 8514/A	1,024 x 768	256	256,000
Apple Macintosh II	640 x 480	256	16.7 million

10MHz 8086 processor to the level of a 25MHz 80386. Likewise, an accelerator card for the PS/2 could increase the performance of a 16MHz 80386 processor to the level of a 25MHz i486. For the Macintosh, an accelerator card may boost an 8MHz 68000 processor to the performance level of a 25MHz 68030.

#### Replacement and Residential Cards

Vendors market two types of accelerator cards: replacement processors and residential processors, also called tandem processors. A replacement processor consists of a circuit board with a cable attached. The accelerator card inserts into an expansion slot, while its cable plugs into an empty socket. A residential processor consists of a card that sits in an expansion slot and operates in tandem with the original computer processor.

Both types have advantages and disadvantages. The replacement processor connects directly to the components of the CPU, but is more complicated and time consuming to install than the residential type. Replacement processors also operate more slowly than residential ones, because they are limited by the existing system memory. Although the residential processor does not connect directly to the CPU, it is easier to install, and faster because it includes its own RAM.

Sometimes called a co-processor, a residential card can do complicated processing while delegating routine house-keeping chores (such as mediating input/output operations) to the original processor. Some cards of this type feature a software utility that allow users to switch back and forth between processors, a useful feature for handling potential compatibility problems.

Although replacement processors are limited to the memory of the microcomputer system, they use memory schemes such as cache memory and RAM disks to increase processing speed. Cache memory consists of a high-speed buffer filled from the main memory. Programs and instructions found in the cache memory can be called at higher speeds to avoid loading another segment from main memory. A RAM disk partitions a *virtual* disk drive structure that resides wholly within system memory. This structure can be used as if it were actually a disk to store and retrieve files.

Some accelerator cards combine the features of a replacement and a residential card. These may be called hybrid boards. For example, some products replace a processor with a cable. Users may then connect the cable to the accelerator board, allowing them to switch back and forth between the cards for compatibility reasons.

The original Apple Macintosh and the Macintosh Plus computers did not carry expansion slots, so accelerators could not be installed in the customary fashion. However, the card may attach directly to the Motorola 68000 processor, which allows the faster chip to control the system.

#### Applications

Users buy accelerator cards for several reasons. Three-dimensional spreadsheets, expert systems, relational databases, and desktop publishing applications all run painstakingly slow on first- and even second-generation processors. Graphical interfaces for applications software and operating environments, such as Microsoft Windows, require more speed and power. These applications may severely task the resources of older machines. By adding an accelerator card, users get state-of-the-art functionality protecting their hardware investment.

However, bus and I/O limitations create expansion card compatibility problems in many cases. Though some cards are compatible with a large variety of microcomputer models, these usually suffer in performance, whereas more specific cards are better performers. Users who need improvements with a variety of applications may be unsatisfied with even the most complex expansion cards. Even a multipurpose expansion card will not make your XT a 486.

Accelerator cards also bridge the gap between the industry-standard MS-DOS and newer, more powerful operating systems such as Unix or OS/2. OS/2 and IBM Presentation Manager, its graphical interface, are designed for Intel's 80386 and 80486 processors. Users wishing to run advanced software on an 80286 or 8088 machine are prime candidates for accelerator cards. They need an OS/2-compatible card with OS/2-compatible Basic Input/Output System (BIOS) chip and the capability to switch between real and protected modes of the new processor.

#### Memory Cards

One of the most important considerations microcomputer users face is whether their machines contain enough Random Access Memory (RAM). As users' tasks become more advanced, they require more memory than initially provided. Today, operating systems and environments require substantial memory—Microsoft Windows 3.0 requires 2M bytes and OS/2 can require a hearty 11M bytes. Graphics applications also require huge amounts of memory — 25M bytes of storage is required to digitize an 8½-by-11 inch color image using a 16.7 million-color palette.

Users increase the amount of RAM in their systems by adding a memory expansion card. User-installable on

most microcomputers, a memory card is a circuit board that fits into an expansion slot inside the microcomputer's system unit. The card enhances the performance of the computer when carrying out memory-intensive tasks. Cards are available for different types of computers, operating systems, and memory standards.

System Memory has three distinct considerations: size and type of the chips, access method (conventional, expanded, or extended memory), and operating speed.

### Memory Standards

Memory access is a confusing issue. The MS-DOS operating system can directly address only 1M byte of conventional memory, 384K bytes of which are required for system use. This leaves a maximum of 640K bytes for applications. Not surprisingly, this amount has proven inadequate for most users. Consequently, Lotus, Intel, and Microsoft (LIM) came up with an Expanded Memory Specification (EMS) that used memory beyond the 640K-byte threshold through a sophisticated mechanism known as bank switching. Essentially, unused portions of the system-memory area were used to hold one segment or another of memory for application use. The specification, known as LIM EMS, allowed programs to access not only conventional memory (up to 640M bytes), but extended memory as well (from 640K bytes to 1M byte).

It did not take long until 1M byte of RAM also became inadequate. AST, Ashton-Tate, and Quadram then developed the Enhanced EMS (EEMS) standard to overcome the inefficiencies of EMS. EEMS switched up to sixteen 64K-byte banks of data among the 1M byte available to MS-DOS and supported a maximum of 16M bytes of RAM (extended memory). It also provided faster access to data because its data banks were four times larger than those offered by LIM EMS. Furthermore, EEMS was a superset of EMS; any application written for an EMS memory card could run on an EEMS memory card.

Not to be outdone, the LIM consortium later announced EMS 4.0, a new version of the EMS that unites the EMS and EEMS standards. EMS 4.0 supports a maximum of 32M bytes (more extended memory) to better handle multitasking applications. It also includes other complicated features like multiple page-mapping, dynamic memory allocation, and the naming of data handles. Current software applications running under EMS 3.2 or EEMS are upwardly compatible with EMS 4.0 but will have to be equipped with new drivers.

### Memory Chips

Memory chip technology is continuously evolving. 1M-bit chips are popular, and manufacturers often include them on a ribbon-size card with a single row of pins. 4M-bit memory chips are currently in use in Single In-line Memory Modules (SIMMs), but IBM has yet to commercially use the 4M-bit chips. In the SIMM configuration, an entire row/card must be replaced or added as a unit.

### Video Cards

In recent years, personal computer graphics have become more popular than ever, due to the growth of such applications as desktop publishing, computer-aided design (CAD), computer-aided manufacturing (CAM), presentation graphics, and multimedia applications. Users create graphics by utilizing application software to generate images on a display, allowing users to view the images and modify them on-screen. Interactive Multimedia applications, which combine high-fidelity audio and full-motion

video with text and high-resolution graphics, have become very hot in the past year. Intel leads these applications with its ActionMedia 750 image capture and delivery boards, which enable the recording and playback of motion video.

A hardware interface receives instructions from the software and translates them into electronic signals intended for the display. The display then generates the intended image on the screen. The interface consists of video circuitry that resides inside the system unit of the personal computer. On some microcomputers, the circuitry comes standard, taking the form of a video chip that the manufacturer installs on the system board. The Apple Macintosh SE and IBM PS/2 are examples of this.

### Digital Graphics Standards

As in many areas of computing, IBM has set the standard in personal computer graphics. A standard consists of specifications for resolution, color, and other video card features. These specifications become standards when adopted by the majority of the industry, including software developers, hardware developers, and users. A number of standards have existed over the years. Described below in chronological order are the digital graphics standards that transmit electronic signals in digital form to compatible displays.

**IBM's Monochrome Display Adapter (MDA)** was the video card used with the original monochrome IBM personal computer (PC) systems. It supported text at a resolution of 80 columns by 25 lines in one color with no graphics capability. MDA is no longer a viable standard.

Hercules computer technology overcame the problem presented by IBM MDA by releasing the **Hercules Graphics Card (HGC)**, a monochrome video card that supports graphics. The HGC produces graphics at a resolution of 720 by 350 picture elements (pixels). Though very outdated, the Hercules Graphics Card is still available for \$299.

Other Hercules video cards, including the Graphics Card Plus (also \$299), contain Ramfont mode, which allows users to display up to 3,072 software-defined characters and to mix graphics with text and foreign language alphabets. All of Hercules' cards are low-end, 8-bit models, ranging in price from \$89 to \$299.

**IBM Color Graphics Adapter (CGA)** was IBM's first color graphics standard, with a resolution of 320 by 200 pixels and 8 simultaneous colors from a palette of 16. Though CGA-only cards are no longer available, many of the newest cards are backward-compatible, retaining a CGA graphics mode.

**IBM Enhanced Graphics Adapter (EGA)**, the successor to IBM CGA, provides a maximum resolution of 640 by 350 pixels in addition to 16 simultaneous colors from a palette of 64 colors. After IBM EGA appeared, users noticed that in order to upgrade their video cards they would have to replace their displays, because displays designed for CGA did not support EGA. To resolve the compatibility dilemma, in 1988 NEC Home Electronics introduced the MultiSync Color Monitor, which supported EGA and all previous graphics standards. Two of NEC's three current MultiSync models still support EGA.

Different types of video cards transmit electronic signals to the display at different frequencies. Instead of accepting signals at one frequency, the MultiSync automatically adjusts itself to the signal being received. Once again, users could upgrade their video cards without having to replace their displays. They could also move MultiSync displays from one computer to another without having to

also move video cards. Today, a number of vendors besides NEC manufacture MultiSync-compatible or multiple frequency displays.

### Analog Graphics Standards

Digital video cards work like on-off switches to create color images, by providing two intensities of eight colors (red, green, blue, cyan, magenta, white, black, and brown). Operating in this fashion severely limits the number of colors available to users who require greater color capability to produce advanced graphics, such as photo-realistic images. Utilizing analog signals resolves this problem. Analog video cards work like dimmer switches to provide an infinite number of intensities of the eight colors mentioned above. This technique enables users to take advantage of an almost limitless number of colors. Analog graphics standards are described below.

**IBM Professional Graphics Controller (PGC)**, also referred to as the Professional Graphics Adapter (PGA), was the first graphics standard to utilize an analog signal. It provided a maximum resolution of 640 by 480 pixels, in addition to 256 simultaneous colors. Designed for CAD/CAM applications, IBM PGC never quite caught on, perhaps because of limited software support for the card. Still, several video cards continue to support this mode. Most users requiring high-resolution graphics for CAD-type applications have adopted the IBM Video Graphics Array (VGA) standard or the IBM 8514/A standard.

When IBM announced its PS/2 computers in April 1987, it also announced support for two new graphics standards: **IBM Multicolor Graphics Array (MCGA)** and the **IBM Video Graphics Array (VGA)**. Entry-level PS/2 models, such as the model 25 and model 30, provided MCGA graphics, while the midrange and high-end models provided VGA graphics. VGA provides a maximum resolution of 640 by 480 pixels, in addition to 256 simultaneous colors from a palette of 256,000 colors. MCGA is basically an advanced version of IBM CGA that supplies more video modes than CGA but fewer modes than VGA. The VGA standard incorporates not only analog signals but also backward-compatibility. VGA cards work with software that supports IBM MDA, CGA, and EGA. VGA cards only work with analog displays.

The PS/2 contains MCGA and VGA graphics circuitry in the system board. PS/2 users do not have to purchase a separate video card because the graphics capability comes standard. In other words, IBM eliminated the intrusion of third-party video cards and regained the lost revenue. Other personal computer vendors, such as Compaq, followed suit and now offer VGA on the system board. IBM does offer a separate VGA card—the Personal System/2 Display Adapter, which brings VGA graphics to the IBM PC/XT/AT and compatibles, and entry-level PS/2 machines. However, built-in VGA draws screens at a faster rate since the display adapter must route its signals through the system bus of the machine, which the integrated circuitry does not have to do.

A problem arises when using certain VGA cards with the IBM PC/XT/AT and compatibles. Many original VGA cards could not guarantee complete compatibility with the IBM standard because they offered only BIOS compatibility. BIOS stands for Basic Input/Output System and is the part of an operating system that links the software to the specific hardware devices. All VGA cards available today are register-compatible.

Although IBM VGA has become the most widely used graphics standard, it is still not enough for some users. In

response, vendors introduced extensions to VGA: **Super VGA and IBM 8514/A**. Championed primarily by NEC Home Electronics, the Super VGA standard provides a maximum resolution of 800 by 600 pixels and the same colors as ordinary VGA. NEC's hope that Super VGA would replace IBM VGA have not come true. Our survey showed 46 percent support for VGA, and only 27 percent support for Super VGA.

Directly in NEC's path sits IBM, promoting its 8514/A standard. The Personal System/2 Display Adapter 8514/A provides a maximum resolution of 1,024 by 768 pixels and the same colors as VGA and Super VGA. Unlike Super VGA cards, the 8514/A contains a chip that processes graphics, increasing the speed with which the images are displayed. Most boards featuring such high resolution are intended for advanced graphics (CAD/CAM) use, therefore utilize some kind of graphics processor. Our survey showed very weak (12 percent) support for IBM's 8514/A standard.

The **Apple Macintosh II** graphics standard applies only to the Macintosh II family of microcomputers. The Macintosh II High-Resolution Display Video Card provides a resolution of 640 by 480 pixels, along with 16 simultaneous colors from a palette of over 16.7 million colors.

Other Macintosh models include the Macintosh Plus, Macintosh SE, and Macintosh SE/30. These machines feature a 9-inch diagonal, monochrome video display that is built directly into the computer's system and cannot be removed. Users who require a larger screen, color graphics, or both must have an authorized dealer attach an external display to the computer. The Macintosh Plus has no expansion slots, so an external display must be attached via a cable to the system board. Both the Macintosh SE and the Macintosh SE/30 feature one expansion slot, which may accommodate a video card. The expansion slots of the Macintosh SE, SE/30, and II, however, are not compatible with each other. Users must insure that the video card they are considering does indeed work with their machine.

The graphics capability of Macintosh microcomputers has made them the standard for advanced video applications such as desktop publishing and multimedia. Initially the Macintosh provided what is called eight-bit color, supporting 256 colors; then 24-bit with 16.7 million colors. Now Apple offers 32-bit color with a 16.7-million color palette, which, along with the Mac's 72-dpi screen resolution, creates an excellent environment for graphics applications. The 32 bits consists of 24 bits for color information plus 8 bits for special video effects, such as text overlays and three-dimensional modeling. Several vendors now offer video cards compatible with 32-bit QuickDraw, including RasterOps, Radius, SuperMac, and Truevision.

With 32-bit color in place, the stage is set for amazing new interactive multimedia applications, all of which depend on high-powered video cards. Our survey results indicate that multimedia management cards are on the rise, and by 1991 we expect to see a variety of new products for these applications.

Intel has made the most significant advances in multimedia with its ActionMedia 750 product family, which includes the ActionMedia 750 capture and delivery boards and two development software programs: Production Tools and Software Library. The \$2,150 Intel capture board captures two-channel hi-fi audio, high-resolution still images, and motion video from live and recorded sources. It then converts the analog audio and digital video



signals to digital data. Intel's delivery board then compresses and decompresses the video with a microprogrammable video processor. The delivery board also includes one megabyte of VRAM, VGA and analog RGB output, and a SCSI interface to CD-ROM drives. At \$1,995, this board also improves motion video resolution through line doubling and pixel interpolation.

Though currently DOS-based, Intel plans to port its DVI software to OS/2, Windows, and UNIX in 1991. By 1992, Intel expects DVI to be a standard feature on motherboards, providing multimedia video conferencing for the desktop.

## Selection Guidelines

### Accelerator Cards

Users who require faster machines have a number of options available. First, they can buy systems that incorporate more powerful processors. Most expensive, this option requires the purchase of entirely new machines and the sacrifice of investment in original equipment. Rather, users may increase system memory by adding random access memory (RAM) and by using memory schemes such as disk caching and RAM disks. Though this option costs less than an entirely new system, it increases speed by reducing disk access time, not by increasing processor speed.

Advantages of using accelerator cards include preserving the user's hardware investment by giving the current equipment the capability to run more sophisticated software. Disadvantages include possible compatibility problems with the system bus, mass storage devices, and existing software. The user must consider these factors and choose an accelerator card that is best suited to an individual system when deciding how to upgrade. In some cases, mass storage devices and software must be replaced by compatible options.

Also called a co-processor card or a turbo board, an accelerator card is a circuit board with a processor that runs at greater speed than the current system. An accelerator card costs less than a new computer system and offers much better performance than a memory scheme.

However, despite their benefits, accelerator cards do not match the performance of an entire new system. For example, an IBM Personal System/1 with an 80386-based accelerator will *not* equal the performance of a new 80386 PS/2. Some system components are not sped up at all by adding an accelerator card. These components include the disk drives, video display, and system bus width. Only an entirely new system provides a wider bus.

Macintosh microcomputers are very difficult to upgrade. In order to satisfy the terms of the warranty, only authorized service people can install an expansion card. Macintosh processors are not completely compatible due to programming shortcuts, so product choices are extremely limited. The additional power demands of an accelerator card may also require upgrading the computer's power supply.

### Memory Cards

New microcomputer applications are demanding more system memory than ever, and many users find memory cards an easy way to increase system RAM. Most memory cards are user-installable and compatible with a wide range of microcomputer types, operating systems, and memory standards. Memory cards help the system accommodate larger data files and more memory-demanding graphics

applications. Often, graphics applications use so much memory that they cannot be loaded into a system without additional memory.

The first IBM Personal Computer, introduced in 1981, contained just 64K bytes of dynamic random access memory (DRAM). Today, the average PC has eight times that amount (512K bytes), with demand increasing at an unprecedented rate. It is predicted that the average PC will incorporate about 10M bytes of DRAM by 1992 and a staggering 20M bytes by the end of the decade.

Digitizing images requires substantial memory. One AutoCad image, for example, can require 5M bytes of storage. For animation, this much is needed for every fraction of a second. With more color and the amount goes up exponentially. Imagery, although it will revolutionize the way we use microcomputers, will have a dramatic effect on memory usage.

Additional memory also enables the support of new operating systems. The MS-DOS/PC-DOS operating system occupies only 384K bytes of system memory. Although DOS remains dominant today, more powerful systems such as Microsoft Windows, OS/2 and UNIX are taking over with high-end features that require more memory. OS/2 requires 3M bytes, while both Microsoft Windows 3.0 and UNIX specify a minimum of 2M bytes, but recommend 4M bytes or more.

Because of their additional storage space, memory cards also facilitate multitasking (multiple simultaneous processing). Multitasking requires a number of applications and related files in the system RAM. Today, an increasing number of operating systems include multitasking capabilities, and with these come greater demand for memory cards.

### Video Cards

A drastic increase in the use of high-resolution computer graphics has meant tremendous success for the video card industry, even though many systems now include built-in graphics standards that are backward compatible. Microcomputer users now have sophisticated applications for video cards in desktop publishing, computer-aided design (CAD), computer-aided manufacturing (CAM), and presentation graphics design.

Video cards are now available with many graphics standards, and users can find an answer for almost any graphics requirement. Selecting the correct video card depends on the applications desired and the display type in use. Most video cards are user-installable; however, some require dealer installation.

Compatibility is a diminishing issue with video cards. Though most video cards are backward compatible, the wide variety of graphics standards still make some cards useless with certain types of microcomputers and displays. In particular, Macintosh computers are difficult to upgrade because they lack expansion slots.

The IBM PS/2 contains MCGA and VGA graphics circuitry built directly into the system board. Thus, PS/2 users do not have to purchase a separate video card because the graphics capability comes standard with the machine. Built-in VGA draws screens at a faster rate than a VGA card since the display adapter must route its signals through the system bus of the machine, which the integrated circuitry does not have to do.

A problem arises when using certain VGA cards with the IBM PC/XT/AT and compatibles. Many original VGA cards could not guarantee complete compatibility with the

IBM standard because they offered only Basic Input/Output System (BIOS) compatibility.

Unlike Super VGA cards, the 8514/A contains a chip that processes graphics, increasing the speed with which the images are displayed. Most boards featuring as high a resolution as the 8514/A are intended for advanced graphics use (CAD/CAM) and therefore utilize some kind of graphics processor.

---

### Emerging Technologies

A number of technological advances in video cards are on the horizon. These include new architectures for new applications such as multitasking and networking. IBM's Micro Channel Architecture (MCA), used by its midrange and high-end PS/2 models, is now integrated in the video card market. IBM considers MCA superior to the industry standard architecture currently used in the IBM PC/XT/AT and compatibles. However, expansion cards designed for the ISA standard do not operate with MCA machines.

Opposing IBM are the developers of the Extended Industry Standard Architecture (EISA). Nicknamed the *Gang of Nine*, this group includes AST Research, Compaq Computer, Zenith Data Systems, and other vendors of pc-compatible systems. Members of the EISA coalition contend that the current architecture is a good one and merely

needs updating. EISA machines currently support expansion cards designed for the ISA standard.

Also on the horizon is bus mastering, also called Direct Memory Access (DMA). Bus mastering technology enables expansion cards to access the system bus without going through the CPU. DMA cards are not accelerator cards, as they do not increase the speed of the microprocessor. Instead, they handle background processing during multitasking, freeing up the CPU and increasing the speed of the application running in the foreground. To avoid confusion, IBM calls these cards applications accelerators.

Bus master cards can send and receive facsimile transmissions in the background, allowing graphics programs in the foreground to operate much more quickly. This technology will take further advantage of the 32-bit bus of MCA and EISA machines in the coming year. Only MCA bus master cards are currently available.

The growing market for multimedia hardware and software is beginning to impact expansion card technology. Multimedia applications can include full-motion video, high fidelity audio, text and high-resolution graphics. Current multimedia hardware includes video capture and frame grabber boards, which capture images from a video camera and digitize them for use in computer applications. Already the markets for conventional video and multimedia equipment are merging, resulting in cards that will be able to perform these high-end multimedia applications. ■

# An Overview of Modems

## In this report:

Digital versus Analog .....	2
Transmission Characteristics .....	4
Compatibility.....	8
Selection Guidelines.....	10

## Datapro Summary

For a device afflicted in recent years by a gloomy prognosis, the modem shows remarkable resiliency and staying power. Instead of retrenching under the threat of the comprehensive, but as yet unrealized, capabilities of ISDN, modem vendors have fortified their products with features that consistently fill users' needs. Faster speeds, data compression, and advanced error correction are keeping the modem industry alive, well, and growing. Not only can users choose from a range of sophisticated devices, but they are facing the happy prospect of paying less for their purchases. Former users of leased lines are now realizing significant economies by switching to high-speed dial-up modems that incorporate data compression and error control techniques.

## Technology Overview

Technological advances are extending the life span of the modem. Although viewed for many years as commodity devices, modems still command a sizable market and will continue to do so until digitalization becomes complete. That event, however, appears to be many years away. ISDN has not struck like lightning and obliterated the modem industry. Taking advantage of the slow progress of ISDN, modem vendors are refining error-checking and data compression techniques and revving up their products to achieve the highest speeds that the telephone lines can bear.

To enhance the reliability of modems, Microcom developed the Microcom Network Protocol (MNP) for controlling data errors in asynchronous transmissions. MNP Classes 2 to 5 are commonly incorporated into modems, activated by software in the modem or in the application in which the modem is participating. The CCITT's V.42 standard, adopted in 1989, differs from the various MNP levels by performing

error correction in asynchronous communications through hardware.

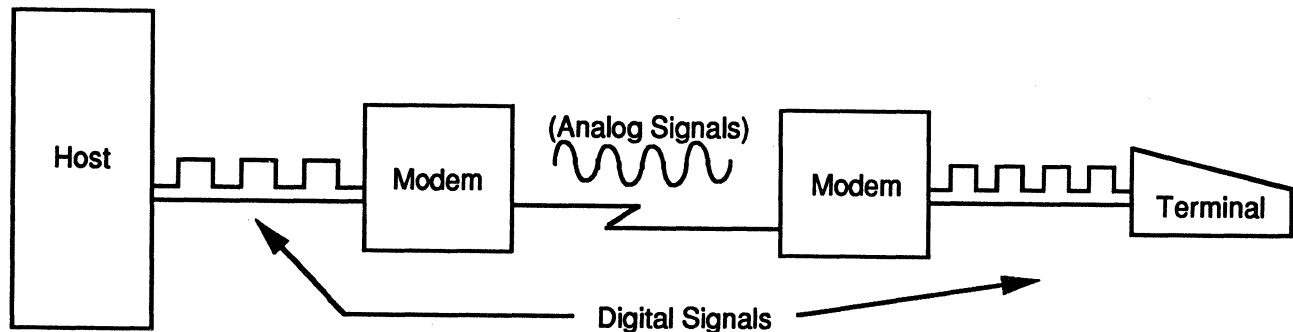
To increase throughput, vendors are adding data compression to their products. This technique deletes superfluous or redundant bits from a datastream by transforming encoded data into words smaller than the eight bits required by ASCII. Microcom pioneered the MNP Class 5 data compression protocol, which enables software conforming to its requirements to reduce files to half their original sizes. Geared to hardware, the protocol prescribed by the V.42bis standard compresses data over 30 times greater than MNP 5. The V.42bis standard also handles previously compressed files more adeptly than MNP 5. Microcom has also developed MNP 7, a more powerful form of data compression, but as of now, MNP 7 has not penetrated most modem product lines.

## Modems: Common Elements

Although there are many different types of modems, all units share common elements: power supply, transmitter, and receiver. The power supply usually takes 120 or 220 V AC and transforms it into the DC voltage

—By Barbara Callahan  
Associate Editor

Figure 1.  
Conversion Process



A modem accepts digital signals from the sending device and converts them into analog pulses that are sent over the public telephone network. At the receiving end, a corresponding modem converts the analog signals back into digital ones.

required to operate the modem's internal circuitry. The transmitter modulates the digital data into analog form; the receiver demodulates analog signals and reverts them to their original digital format (see Figure 1).

Modems designed for asynchronous or synchronous operation differ in one major aspect: A synchronous modem contains a clock source and phasing circuits; an asynchronous modem does not. Asynchronous modems do not need clocking sources because data is transmitted at irregular intervals. In synchronous transmission, data is sent continuously in regular, clocked intervals.

Additional components of modems are:

**Data Encoder:** Part of the transmitter, this unit determines the modulation changes to the carrier frequency at each sampling segment.

**Modulator:** Also part of the transmitter, this component changes the carrier frequency, as determined by the encoder.

**Filters:** Part of the transmitter and receiver, these circuits transmit signals of frequencies within one or more bands. They also screen signals of other frequencies, thus eliminating noise and other impairments to the information stream.

**Line Amplifier:** Part of the transmitter providing the connection between the modem and the carrier, an amplifier boosts the strength or amplitude of a signal. In a modem, the line amplifier is actually a variable-gain device, usually adjustable in two-decibel steps to accommodate the modem's transmit signal to the requirements of the communications line. The amplifier is configured so that its impedance matches that of the line over the represented frequency range.

**AGC amplifier:** Part of the receiver, this unit supplies automatic gain control (AGC), allowing the modem to compensate for amplitude variations on the line.

**Equalizer:** Generally incorporated into modems transmitting data at 2400 bps and higher, this element is basically an inverse filter whose amplitude and phase characteristics are the inverse of those presented on the telephone line.

The equalizer corrects amplitude and delay distortions that can lead to interference during transmission.

**Demodulator:** Part of the receiver, this unit retrieves data from a modulated carrier wave and passes it on to the decoder.

**Decoder:** In conjunction with the demodulator, the decoder formats received data into a serial, binary pattern and sends it to the receiving data device. This component is also part of the receiver.

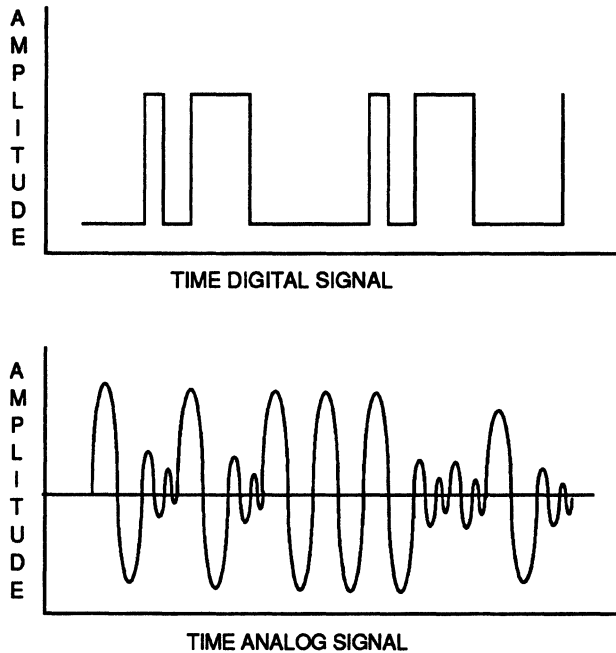
### Digital Versus Analog

Modems are essential components of a data communications network because they allow communication between *digital* devices transmitting information over the public or private telephone network comprised of *analog* facilities. Without conversion, soundless digital signals cannot be sent over an analog facility, which is designed to carry sounds. Figure 2 illustrates the differences in the ways in which digital and analog signals are formed.

Analog signals are waves, and digital signals are a series of pulses with very short rise (leading edge of signal) and fall (trailing edge) times, culminating in a squared-off signal pattern. Technically, these quick transitions result from high-frequency, harmonic signals that are often over 10,000 cycles per second. In a digital datastream, information sent in a series of pulses represents a binary 1 or 0—a rise represents a binary 1, and a fall represents a binary 0 (see Figure 4, top diagram). Since the telephone system is designed to carry the human voice within a frequency range from 300 Hz to 3300 Hz, it cannot accommodate these high-frequency signals. Any attempt to send such information down an analog facility will result in a blurring or distortion of the digital pulse, making it lose its squared-off appearance. Once this happens, it is impossible to see where a digital pulse begins and ends.

Other factors affect the integrity of digital transmissions. A loss of signal strength occurs as the distance between modems increases. Transmission speed also affects data integrity—the higher the data rate, the more pulses sent. As more pulses are sent, they grow closer together and become more prone to distortion. Random noise, caused by molecular vibrations on a communications circuit, generates a low-level mixture of electromagnetic waves at different frequencies, producing a “hiss” on the line. If the transmission signal power falls below an acceptable level,

Figure 2.  
Digital Versus Analog Signals



the digital bit stream may become corrupted. Noise from other sources, such as electrical equipment or atmospheric conditions, also affects the digital datastream.

**Modulation**

Since the public telephone network carries analog signals, digital signals must be converted into analog form by a process called *modulation*. Modems modulate the digital signals used by computers and other types of data equipment into audio tones that can be carried by the analog network. Once the modulated signals reach their destinations, they are *demodulated* to their original digital formats and sent to receiving data equipment. Figure 1 shows modulation and demodulation on a point-to-point communications link.

Modems support modulation techniques to fulfill the requirements of various applications. Simple, slow-speed devices use a technique called frequency-shift keying (FSK); higher-speed devices of 9600 bps and above use a technique called quadrature amplitude modulation (QAM).

The terminology and mathematical theory describing electrical signals are based on the concept that a signal is composed of multiple, simple signals called sine waves. Most modems transmit a continuous sine wave, defined by two parameters: amplitude and time. Once defined, the wave possesses two properties convenient to mathematical manipulation: frequency and phase. Any combination of three of the parameters—amplitude, frequency, and phase—can serve as the basis for a modulation technique. The three basic types of modulation techniques are *amplitude modulation*, *frequency modulation*, and *phase modulation*.

**Amplitude Modulation (AM)**

The simplest technique, amplitude modulation generates a single carrier frequency signal. If the resultant wave is of

high amplitude, it denotes a binary 1; if it is of low amplitude, it denotes a binary 0. AM is highly susceptible to line interference. Quadrature amplitude modulation (QAM), a combination of amplitude modulation and phase modulation, is essentially a four-phase technique. This method uses two signals at the same frequency, but they are 90 degrees out of phase with each other. For each signal, four possible levels of amplitude can be applied (A1, A2, A3, and A4). By combining the two signals that are 90 degrees out of phase, 16 different conditions, each representing four bits of information, can be generated. It is also possible to represent 32 different states, generating twice as much information.

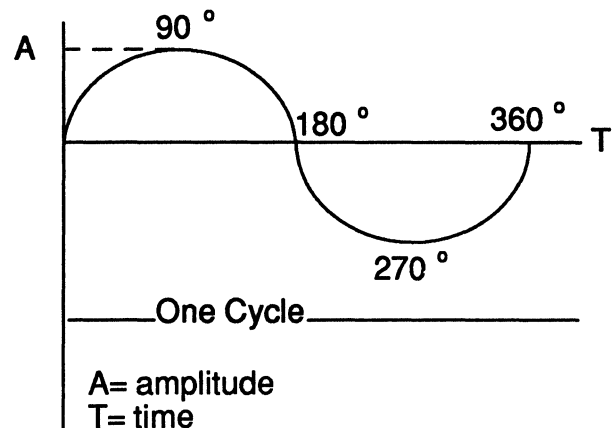
**Frequency Modulation**

The most common frequency modulation is *frequency-shift keying (FSK)*, a two-level technique used on AT&T 103 and 113 Series modems. FSK modulation represents changes in the binary bit pattern by alterations in the frequency of an audio tone. This line is assumed to be in a steady binary 1 or “mark” state when it is idle, represented by one frequency of tone. When the data bit value 0 is sent, the modem changes to another tone frequency, causing a unique, almost musical, effect during the sending of data. FSK modulation works well for relatively low speeds, but as the speed of the digital signal increases, the time allocated to shift frequencies is reduced. Both the production and detection of audible changes become more difficult.

**Phase Modulation**

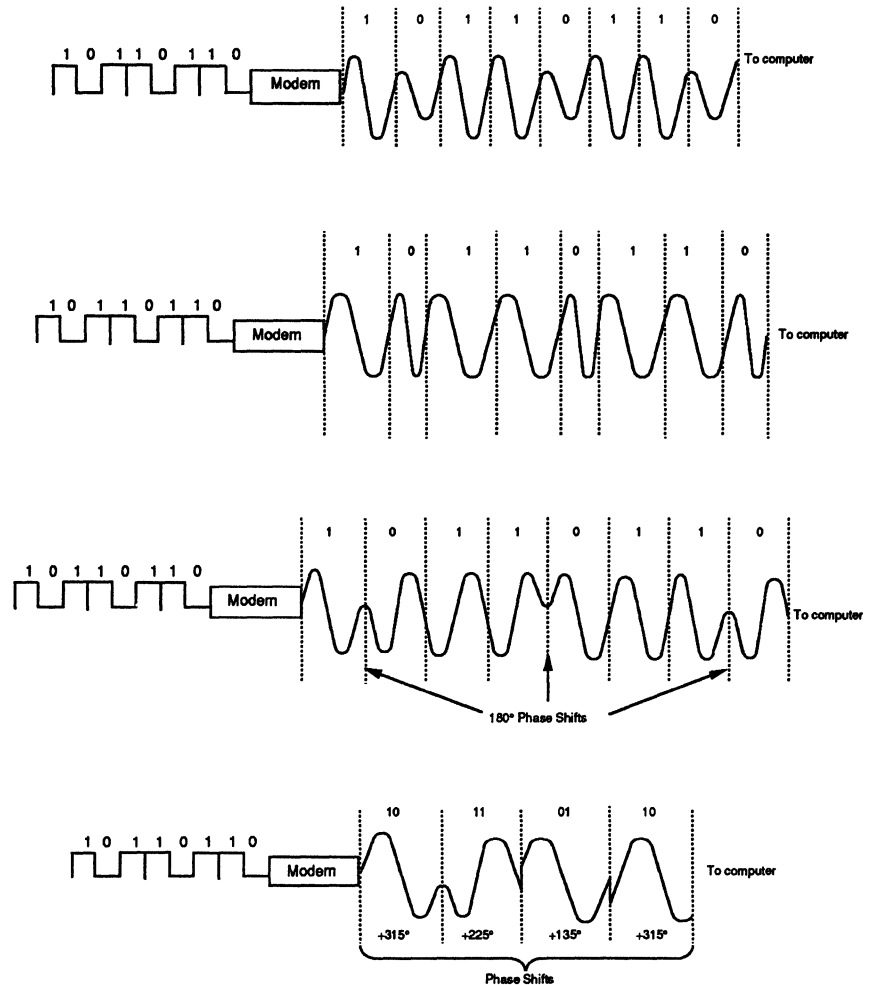
*Phase-shift keying (PSK)* uses changes in the phase of a signal or its timing relationship to a fixed reference to indicate a change in the bit pattern. A reference oscillator determines the phase angle change of the incoming signal, which, in turn, determines which bit or dibit is being transmitted. *Differential phase-shift keying (DPSK)* compares the phase angle of the incoming signal to the previously received dibit. One change in phase is interpreted as a binary 0 if the preceding phase has been interpreted as a binary 1. This method does not require a separate reference wave, thereby reducing the amount of circuitry in the modem. PSK is used in many medium-speed modems and is combined with amplitude modulation in high-speed applications to form *quadrature amplitude modulation (QAM)*, the technique for 9600 bps and higher.

Figure 3.  
Sine Waves



Analog signals are transmitted over the public network in sine waves, depicted here.

Figure 4.  
Basic Modulation Techniques



### Communications Facilities

Primarily because of cost, most data communications applications use the public switched (dial-up) network or equivalent leased lines. Many other facilities exist, however, such as radio links, direct cable connections, fiber optic links, infrared links, lasers, and other common-carrier facilities. The public telephone network and equivalent leased lines are referred to as voice grade lines. Lines of lower capacity are called narrow-band lines; ones of higher capacity are called broadband or wideband lines. Communications line describes a path over which a particular connection is established. The exact physical nature of the connection can vary, and several different types of links can form one connection.

In general, a narrow-band line accommodates up to 300 bits per second, a voice grade line up to 9600 bits per second, and a wideband line considerably more. Cables and other types of facilities generally have large capacities, but are limited by technical considerations.

*Dial-up modems* operate on the public switched telephone network, while *leased-line* units operate on private facilities. *Limited distance* modems operate on private lines in local applications, while *line drivers* and *modem eliminators* operate on the customer's premises. Figure 6 shows the relationship between the communication facility and modem types.

### Bandwidth

Bandwidth refers to the information-carrying capacity of a transmission facility. It actually defines the range of frequencies, measured in hertz (cycles per second), that a transmission medium can accommodate without interference or signal loss. The greater the range of frequencies a medium can handle, the greater its information-carrying capacity. Most modems transmit data within a 300- to 3000-Hz frequency range in the middle of a bandwidth.

Although signal characteristics are usually optimal in the middle of a bandwidth, transmission limited to the middle of the band restricts the amount of bandwidth available for data. To compensate for this factor, conventional modems use sophisticated, multiple-bit encoding algorithms to squeeze as much data as possible over one carrier in each direction. The downside of this solution, however, is an increase in data lost during line hits or other error-inducing conditions on the transmission medium. The goal of much modem design work is to minimize data loss while transferring greater amounts of data.

### Transmission Characteristics

#### Data Rates

Depending on other transmission characteristics, modems operate at a fixed speed or at the speed of the sending device—up to a specified limit. Some modems can be

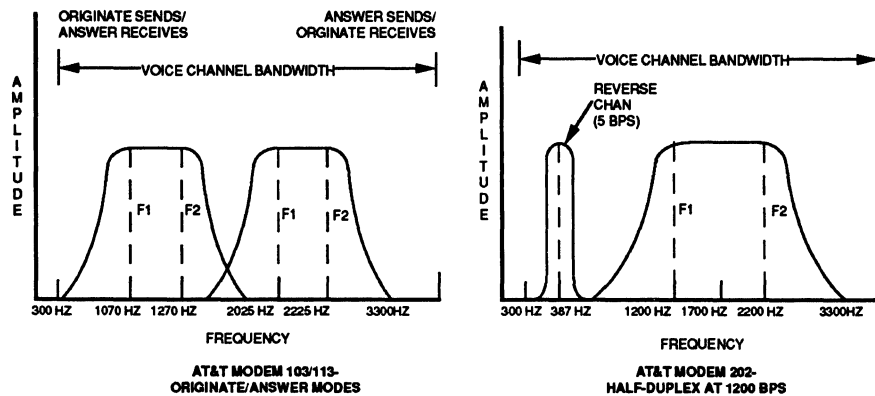


Figure 5.  
Modem Bandwidth Usage

equipped to operate at several different speeds. Users can adjust operation speeds via software commands or changes in components, wires, or switch settings.

A modem's *fallback rate* refers to its capability to detect poor line conditions and to lower transmission speeds to prevent errors. For example, a modem operating at 9600 bps may fall back to 7200 bps or 4800 bps when line conditions deteriorate.

As the data rate increases in limited distance modems, the distance over which transmission is effective decreases.

Data transmission speed is usually measured in *bits per second (bps)*, but it is not uncommon for a data rate to be expressed in *baud*. Although often used synonymously, the terms bps and baud rarely mean the same thing. The bps designation expresses the data signaling rate, while baud measures modulation rate.

### Line Conditioning

The telephone network is not a perfect electrical path, even for voice communication. As the speed of data exchange increases, it becomes increasingly difficult for modems to manage the precise signaling required for accurate information transfer. A number of problems affect the performance of a circuit, such as line failure, electrical interference, or random noise.

Variation in the strength of the signal itself, independent of noise, can cause a problem called *amplitude distortion*. To combat amplitude distortion, modems must adjust the properties of the communications line to prevent the signal from getting too far out of shape. Circuit problems also arise from changes in the analog signal, caused by uneven propagation of low and high frequencies, a condition known as *phase jitter*.

To minimize problems in leased circuit transmission, users can pay a premium for a line that is specially *conditioned*. The telephone company will either select a path that meets the more stringent requirements of the conditioning or it will service the path to meet necessary standards.

### Synchronization

To ensure an orderly data flow across a communications facility, a time relationship, known as synchronization, must exist among the bits that make up the messages. The two basic forms of synchronization are asynchronous and synchronous. PC-to-PC transmissions via modem usually occur in asynchronous mode. Synchronous mode accommodates communications from personal computers or terminals to mainframes.

*Asynchronous* or stop-start transmission refers to data transmitted in an irregular manner, one character at a time as it is being keyed. The time interval between sequentially keyed characters varies in duration because of the impracticality of keying consecutive characters at precise intervals. With asynchronous transmission, each character (byte) is delineated with one start bit, one or two stop bits, and an optional parity bit. The parity bit is part of a simple error detection scheme known as parity checking.

*Synchronous* transmission occurs in a continuous stream. Data transmitted from files on storage media can be handled as a continuous datastream. Electronic or mechanical methods can accurately control the flow. Since only a few control characters are needed to delineate large blocks of several characters (bytes), communications overhead is lower and more time can be spent transmitting useful data. Circuitry to establish the timing of the datastreams is more complicated, however, making synchronous modems more expensive than asynchronous models.

Asynchronous modems can usually transmit data at any rate up to the specified maximum rate of the modem. Synchronous modems operate at fixed transmission rates, established by internal or external clocking sources.

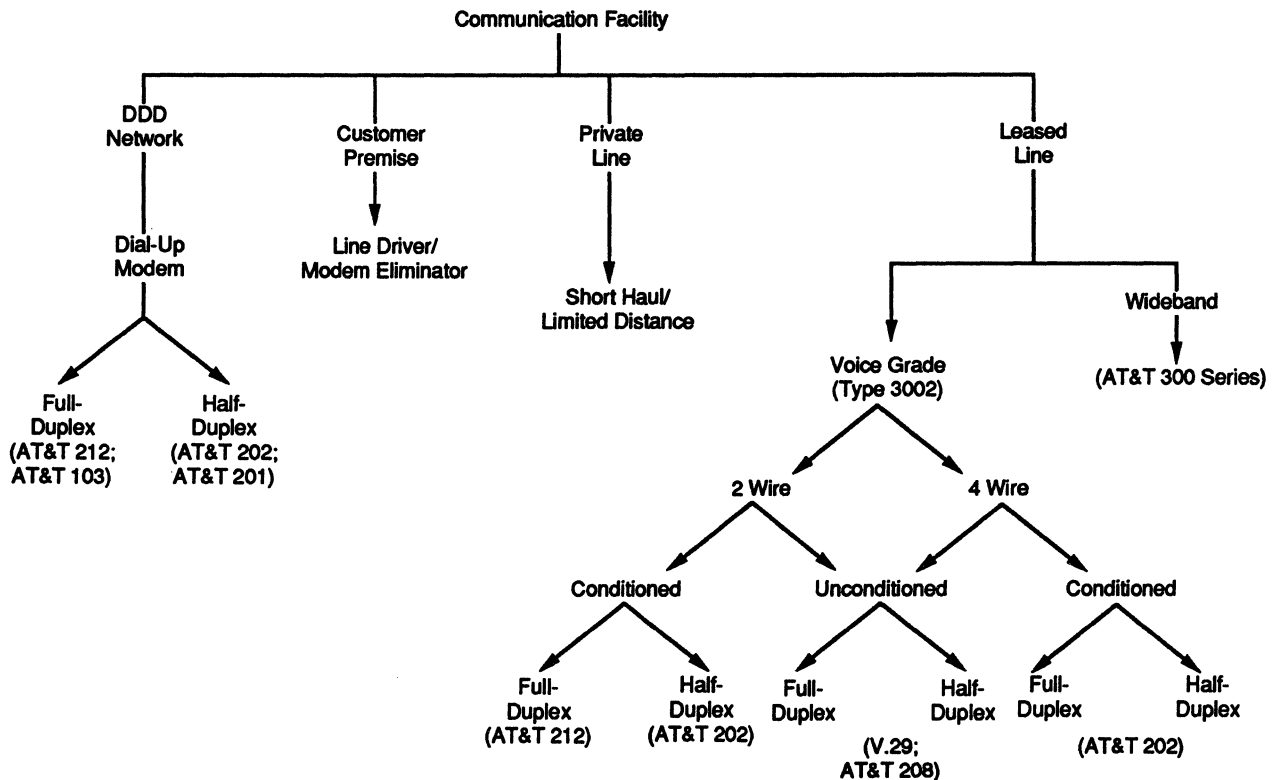
Modems operate in these transmission modes: simplex, half-duplex, or full-duplex. The simplex mode supports unidirectional data transmission, i.e., data is transmitted only or received only. The half-duplex mode supports data transmission in either direction, but not simultaneously. In full-duplex mode, data is transmitted in both directions simultaneously. Many modems operate in both half- and full-duplex modes.

### Equalization

Equalization neutralizes the undesirable electrical characteristics of a communications line that distort transmission, increase the error rate, and degrade operating performance. This technique matches line conditions to take full advantage of the line's data-rate capability. The faster the modem, the greater the need for equalization and the more complex the equalizer. Most low- and medium-speed modems contain fixed (nonadjustable) equalizers.

*Fixed equalization* handles the average group-delay distortion for all telephone lines and may be of no use if the actual connection does not fall within this range. *Manually adjustable equalizers* are preset at the time of installation to match the line and are normally used for leased or private service. *Compromise equalization* compensates only

Figure 6.  
Communications Facility Versus Modem Type



for a fixed distortion on a specific, standard telephone line. Most AT&T 212A-compatible modems use compromise equalization.

Modems built around microprocessors offer distinct advantages: improved equalization, increased reliability, added flexibility, and faster speeds. The powerful capabilities of a microprocessor support the implementation of sophisticated algorithms that enable equalizations to be performed automatically on changing line conditions. Many newer medium-speed and nearly all high-speed modems offer *automatically adaptive equalization*, which samples the line several times a second to determine the parts of the transmitted signal that need enhancement to be readable. Many new modems check the need for signal enhancement before they transmit, a technique known as *forward compensation*.

### Error Correction

One of the chief problems of data transmission is the varying quality of the analog voice-grade telephone lines. In addition to the usual limitations of the analog line, many errors can enter the bit stream from transient noise, harmonic distortion, phase jitter, and other signal disruptions. Current solutions to these problems include expensive line conditioning and/or the selection of modems that perform sophisticated line equalization functions.

Forward error correction (FEC) techniques enable a processor to put a bit stream through a series of complex algorithms before transmission occurs, resulting in a rearranged bit sequence with extra bits added to the original block of data. At the receiving end of the communications circuit, another processor decodes the bit stream. The bits inserted at the transmitting end determine if the block was received correctly. These bits correct any blocks received

incorrectly. All these adjustments occur without retransmission of any part of the original data.

To accommodate the growth in the use of microcomputer dial-up links, error correction protocols have been introduced to ensure file-transfer data integrity. On-line information services, electronic mail facilities, and packet networks all require different protocols. Furthermore, PC-to-host links have their own protocol specifications. No single protocol meets all communications needs, but users have several options. Some facilities have circumvented the protocol compatibility issue by implementing error-correcting modems that allow a host to communicate with various ASCII-based systems without disturbing the host software or communications port.

**MNP:** Microcom Networking Protocol (MNP) conforms to the International Organization for Standardization (ISO) Model for Open Systems Interconnection (OSI). Microcom claims that MNP offers several advantages over xmodem, the leading software-based protocol. According to the vendor, MNP supports five of the seven OSI layers, while xmodem supports one.

Direct support of the OSI model allows MNP to adapt to different systems. The model's layered design and its protocol separate different functions, such as electrical connection and data link. Modifications do not require total redesign. MNP is a full-duplex protocol, and xmodem is half-duplex. In full-duplex mode, acknowledge characters (ACKs) and negative acknowledgment characters (NAKs) sent along with the data block do not interrupt the flow of data, thereby increasing throughput. In half-duplex transmission, data transfer is halted because transmission occurs in only one direction. MNP also shortens the data



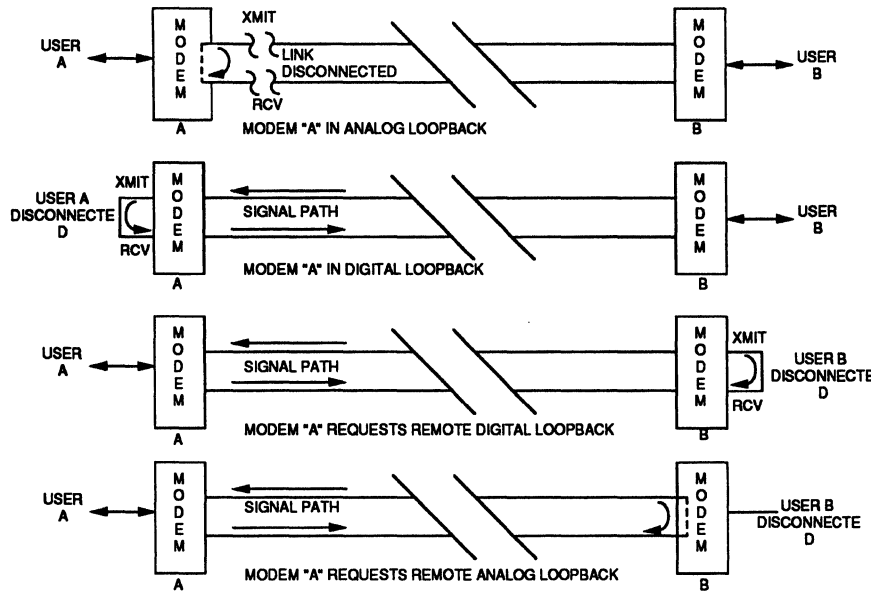


Figure 7. Analog and Digital Loopback Tests

block size on noisy lines and sends packet acknowledgments as part of the frame.

MNP is divided into Classes 1 through 10. *MNP Classes 1 to 3* packetize data and protect data integrity via CRC, a method for detecting packet errors. When necessary, devices equipped with these classes ensure that packets are retransmitted. *MNP Class 4* attains up to 120% link-throughput efficiency, automatically adjusts packet sizes to meet line conditions, and reduces protocol overhead. *MNP Class 5 Data Compression* dynamically adjusts to the type of data being sent. A combination of MNP Class 5 and Class 4 increases link throughput by up to 200%. *MNP Class 7 Enhanced Data Compression* offers higher compression efficiencies. When combined with Class 4, Class 7 improves throughput by delivering efficiencies up to 300%.

*MNP Class 9* includes features such as Enhanced Universal Link Negotiation, Piggy Back Acknowledgments, and Multiple Selective Negative Acknowledgments. Enhanced Universal Link Negotiation supports the connection of MNP- and non-MNP-compatible modems at the highest performance levels. The Piggy Back Acknowledgments feature integrates Packet Acknowledgments into data frames, thereby lowering link overhead. The Multiple Selective Negative Acknowledgments feature dispenses with unnecessary retransmissions by rejecting data frames erroneously received.

In November 1989, Microcom announced *MNP Class 10* as part of its QX/2400t modem. MNP Class 10 optimizes data transmission over poor lines, including cellular connections. According to Richard Sterry, vice president of marketing, "MNP 10 releases the land line terminal from the office and opens the door for a variety of exciting mobile applications in the news media, delivery, and public safety arenas." MNP 10 compensates for signal fade and interruptions, and raises performance to levels approaching land line conditions.

The Adverse Channel Enhancements (ACE) feature of MNP 10 improves connections over low-quality lines and triggers certain performance-enhancing features when line quality deteriorates. ACE includes the Robust Auto Reliable feature, which enables the modem to make multiple attempts to overcome channel interference at linkup. This

feature also supports backward compatibility with non-MNP modems. The Aggressive Adaptive Packet Assembly feature improves link performance in unfavorable channel conditions by changing adaptive packet sizes over the link to accommodate levels of interference. The Dynamic and Negotiated Speed Shifts feature enables the modem to increase speed as line conditions improve and to increase the chances of successful connection at the highest possible speed.

*LAP D and LAP M:* The major rival to MNP is Link Access Protocol D (LAP D), a member of the High Level Data Link Control (HDLC) protocol family. As the error-control protocol in CCITT Recommendation X.25, LAP D is slated for use in ISDN. LAP D offers the advantages of being HDLC based, logical, and reliable. Many DTE manufacturers already support LAP D, and many test instruments exist for it.

Proponents of MNP point out that MNP has an installed base of over 200,000 units and a large number of vendors supply it on their products. Initial problems with MNP on dial-up lines have been solved. Like SNA, MNP is on its way to becoming a de facto standard, despite LAP D's rise to prominence among vendors and users.

In 1989, the CCITT defined the V.42 asynchronous error correction standard, which outlines the primary protocol, Link Access Procedure for Modems (LAP M). Communications between two modems conforming to V.42 specifications occurs via LAP M.

*Trellis Coding Algorithm:* The preferred error-correcting technique for private-line modems is currently the Trellis Coding Algorithm, which derives its name from the trellis-like signal constellation created by this type of signal modulation. Trellis encoding, part of the V.32 standard, is an extension of quadrature amplitude modulation, which allows a modem to tolerate more than double the amount of noise tolerated by a conventional unit at the same block error rate. Basically, this scheme compares redundantly transmitted data and rejects distorted bits. At an equivalent signal-to-noise ratio, a Trellis-coded modem enhances throughput by reducing the error rate almost three decibels. The Trellis modem may require a retransmission only

**Table 1. AT&T Modem Specifications**

Data Rate (bps)	AT&T Name	PSTN		Private 2-Wire		Private 4-Wire		Synchronization		Line Configuration Conditioning
		Hdx	Fdx	Hdx	Fdx	Hdx	Fdx	Async	Sync	
300 (1)	103 Series		•					•		P
300 (2)(3)	108			•	•	•	•	•		P
300	113		•					•		P
1200 (4)(5)	202 S, T	•		•			•	•	•	P
1200 (6)(7)	212, 212A		•					•	•	P
2400 (8)	201B			•			•		•	P
2400 (8)	201C	•							•	P
4800	208A			•	•		•		•	P, M
4800	208B	•							•	P
9600	209			•			•		•	P, M, D1

• = Higher data rates now available from AT&T with Dataphone II modems  
 bps = bits per second P = Point-to-point  
 Hdx = Half duplex M = Multipoint  
 Fdx = Full duplex D1 = D1 line conditioning  
 Async = Asynchronous Sync = Synchronous

(1) 103J is originate/answer with auto answer.

(2) 108F, G have RS-232-C interface and two- or four-wire line connections.

(3) 108H, J have 20 mA current loop interface.

(4) 202S supports 1800 bps on private line with C2 conditioning. Auto answer and synchronous capability standard. Reverse channel optional.

(5) 202T is the same as 202S, but with manual operation.

(6) 212A adds scrambler-kicker to avoid occasional data lockup experienced with earlier 212 modems. Includes 300 bps operation in 103J-compatible mode.

(7) Defines EIA interface to be similar (functionally compatible) with CCITT Recommendation V.24.

(8) PSK modulation.

Courtesy of Concord Data Systems.

once in every 10,000 blocks, while another type of modem may require retransmission after every 10 blocks. Trellis coding is usually incorporated only on high-speed, synchronous modems on private circuits or on CCITT V.32 dial-up modems.

### Compatibility

The term "handshaking" describes the exchange of control signals that establish a connection between two modems. Standards govern the signals required to set up, transmit, and terminate calls. For a handshake to occur, modems must be compatible.

Mixing modems from different suppliers in one network is a common practice because data communications networks are frequently implemented in stages, with ongoing equipment procurements over long periods of time. When organizational changes occur, suppliers that first provide network modems may not continue to do so. In addition, in organizations without central network management, several individuals may be purchasing equipment for one integrated network.

Since most equipment in a network is replaced as it begins to age or malfunction, users can easily switch suppliers. Data communications involves both geographical and corporate distance, which can lead to a division of responsibility and to different restrictions on the communications facilities.

### AT&T Compatibility

Just as IBM has traditionally set the standards for peripheral media, AT&T has set the standards for communications. Although this situation is changing in the postdivestiture environment, many older AT&T modems (often referred to as Bell System modems) are still well known, widely used, and frequently duplicated by other vendors. Modems compatible with a particular AT&T modem specification are likely to be compatible with one another, even if manufactured by different vendors. Manufacturers of AT&T-compatible modems, however, tend to embellish original AT&T specifications with convenience features to distinguish their products from those of competitors.

Compatibility between modems depends greatly on the exactness with which their modulation techniques conform to the AT&T specification. A vendor who has benefited from refining an AT&T standard may alter a specification to such a degree that the unit is incompatible with the original AT&T product.

For the most part, however, manufacturers claiming AT&T compatibility need not be regarded with suspicion. Table 1 outlines standard AT&T modems and related specifications.

### CCITT Compatibility

Many dial-up and private-line modems conform to CCITT standards, the most popular being V.22 and V.22bis for medium-speed dial-up modems, and V.32 for 4800 bps and 9600 bps dial-up units. Other CCITT recommendations in North America are V.26, V.27, V.29, V.33, and their alternatives.

CCITT V.XX modem specifications call for newer, faster modems to be backward compatible with their earlier counterparts. For example, a 2400 bps V.22bis modem will be compatible with its 1200 bps V.22 counterpart at the V.22bis fallback operating speed of 1200 bps. It also will be operationally compatible with most AT&T 212A modems operating at 1200 bps since the standards are closely related.

CCITT V.XX recommendations are directed mostly toward European telephone facility specifications, but U.S. vendors are increasingly incorporating them into their products. Since signaling conventions and equipment differ from continent to continent, compatibility problems can arise when modems designed primarily for European facilities are used on North American networks. For example, a V.22bis modem made for U.S. operation can generally initiate a call to a European-based V.22bis modem because it will tolerate the minor differences in facilities encountered while establishing the call. The same call cannot be completed, however, if the European modem is the originator.

### Hayes Compatibility

Another type of compatibility, indigenous to microcomputer modems, involves the Hayes *AT* command structure. Primarily, Hayes compatibility relates to modem commands, modem responses, and the ability to provide settings that are compatible with various communications programs. For example, modems compatible with the Hayes Smartmodem 1200 obey the 22 modem commands and 16 register settings stipulated for that device. Hayes compatibility also involves compatibility with the AT&T 103 operating standard at 0 to 300 bps and the AT&T 212 standard at 1200 bps. Hayes-compatible products have speakers for audibly following calls in progress; front-panel status lights on external units; and auto dial, auto answer, manual dial, and manual answer features for asynchronous operation.

### Standards Issues

New high-speed modems use modulation schemes compatible with, or similar to, CCITT Recommendation V.32, which facilitates synchronous, full-duplex transmission over two-wire switched circuits at 9600 bps, with a fallback speed of 4800 bps for deteriorating line conditions. V.32 specifies the use of Trellis-coded modulation to overcome noise problems on dial-up lines. Because V.32 has been difficult to perfect, modems designed to V.32 standards were originally expensive. Production difficulties centered around effectively implementing echo-cancellation, which allows high-speed signals traveling in opposite directions to coexist on the same two-wire switched line in the same frequency band. V.32 is presently a growth area in the modem market, effectively handling microcomputer-to-host file transfers.

Currently, the CCITT is working on draft standards for V.32bis, a recommendation that will allow modems to operate at 14.4K bps with fallback rates of 12K, 9.6K, 7.2K, and 4.8K bps. The V.32bis standard will eliminate the need for users to reestablish connections to regain higher-speed operations, a situation that occurs when modems have had to lower their transmission speeds to adjust to deteriorating line conditions.

The CCITT has assigned an accelerated status to V.32bis. The accelerated process bypasses the former CCITT stipulation that formalization of standards take

place only at plenary sessions, which convene every four years. The next CCITT plenary session is scheduled for 1992.

The V.33 recommendation for point-to-point transmission over leased lines is very similar to V.32. Transmission is synchronous, half-duplex over two-wire, and full-duplex over four-wire, leased, voice-grade lines. The specification calls for primary transmission rates of 14.4K bps, with fallback to 12K bps. CCITT M.1020 line conditioning, equivalent to AT&T C1 specifications, is recommended. Modems using implementations of the V.33 scheme are now achieving 14.4K and 16.8K bps transmission rates.

In April 1988, the CCITT approved the V.42 recommendation for specifying error correction for asynchronous modems. V.42 error-correction protocol enables modems to conduct asynchronous-to-synchronous conversion between the DTE and DCE. Incorporated within the V.42 modem, Link Access Procedure for Modems (LAP M) handles error correction. LAP M is an extension of LAP B, the error-correction protocol specified as part of the X.25 packet-switch network standard.

CCITT Study Group XVII approved V.42bis in September 1989. The V.42bis recommendation specifies the data compression techniques required to increase the throughput of modems that feature LAP M error-control protocol. Based on the Lempel-Ziv compression algorithm developed by AT&T Bell Laboratories, the V.42 bis standard also includes enhancements suggested by IBM, British Telecom, and Hayes Microcomputer Products.

### Diagnostics

Diagnostic capabilities range from basic functions, such as local and/or remote loopback, to sophisticated forms of testing that analyze the quality of the received signal.

The most common form of diagnostics is the *loopback test*, in which the user completely checks the data link from one site. The local modem, remote modem, and interconnecting telephone facilities are included. The test returns or loops back the output of the transmitting device to its input. This return occurs via the local modem interface (local digital loopback), the local modem itself (local analog loopback), the opposite end of the communications line (remote analog loopback), or the remote modem (remote digital loopback). Figure 7 illustrates the various types of loopback tests on modems.

Microprocessor-based modems frequently feature advanced capabilities for diagnosing communications failures. A self-test feature determines the operational status of the modem before connecting it into the network. Random pattern generation (511 test pattern) diagnoses errors resulting from random noise, harmonic distortion, and phase jitter—problems that influence the quality of the transmitted signal.

### Network Management

Network management capabilities for modems can cover configuration management, monitoring, restoration, and testing. Via configuration management, users can set the times and dates of the modems, establish the characteristics of ports connecting to various equipment, define transmission modes, and indicate data rates.

Network monitoring alerts an operator to the occurrence of a fault in the network. After reporting a problem, the system pinpoints its cause. Testing capabilities include loopbacks, which check data continuity and measure error rates. In multipoint, polled lines, a polling test determines

problems and actions. In analog lines, tests cover several types of problems, such as jitter and distortion.

Restoral can take the form of dial backup in point-to-point and multipoint lines. Dial backup maintains communications within the network while failures are being corrected.

---

## Selection Guidelines

The highly competitive nature of the modem market forces manufacturers to add new features to their products. A few years ago, modems with comprehensive diagnostics and microprocessor-based capabilities represented new advances in technology. Today, these capabilities are fairly common. Advancements in modulation, error-correction, data compression, higher-speed transmission, and fiber optics represent the most current trends in the industry at present.

### Data Compression

Introduced in the mid-1970s, data compression delivers higher data speeds at lower transmission costs. Modems with data compression use an algorithm that eliminates repetitious characters from high-speed data transmission and encodes regularly recurring character groups. The words *the* and *it*, for example, would fit into a single character. The data is transmitted in this reduced form and reconstructed at the receiving end of the transmission line, resulting in a transmission rate 20% to 50% higher than the basic transmission rate.

Data compression enables data that would otherwise have to be transmitted over a high-speed line with D1 conditioning to be compressed and transmitted at a lower speed over an unconditioned line. In addition, compressed data has a built-in safeguard against eavesdropping and unauthorized system intrusion because individuals with such an inclination would need identical equipment to be successful.

### Reverse and Secondary Channels

Many modems support a reverse channel that relieves the primary channel on four-wire circuits of the burden of carrying acknowledgment data. This reverse channel operates simultaneously with the primary channel, but at a much slower speed. A secondary channel, used in a variety of

applications, can serve as a path for high-speed and low-speed terminals at the same time, eliminating the need for an additional line. A secondary channel can also function as a reverse channel. Reverse channels operate on both two- and four-wire circuits, while secondary channels operate only on four-wire circuits.

### Integrated Multiplexing (Multiport Modems)

Units with this capability contain a limited-function, time-division multiplexer (TDM) that allows the user to transmit more than one synchronous datastream over a single transmission line. The multiport modem's TDM uses the modem's clock for synchronization, thus eliminating the need for one of its own. Less complex than a standalone unit, the integral multiplexer is also less expensive; however, it handles only synchronous data.

### Multiple-Speed Selection (Dial Backup)

Modems operating on leased lines must switch automatically to dial-up lines when the dedicated circuit fails or degrades. Since leased-line operation generally occurs at a higher speed, users can ensure backup by lowering the speed of the modem so that it can operate on the dial-up line until the leased line is restored.

### Security Features

Modem security features include password protection in which the modem accepts or rejects user access codes; call-back capabilities through which the modem returns a user call to verify access; and data encryption in which the datastream is coded.

### Voice/Data Capability

Modems equipped with this feature usually have an adapter to accommodate voice communications over the same line used for data transmission. Although vendors advertise this capability as simultaneous, on most units, the voice and data transmit alternately.

### Network Management

Many modems operate in conjunction with a network management system, through which large networks of modems are monitored from a central site. In this type of environment, the modem monitors its own status, as well as that of its received line signal, reporting conditions automatically to a central control panel. ■

---

# An Overview of PC-to-Fax Boards

---

## Datapro Summary

A PC-to-fax board consists basically of an expansion board installed in a PC expansion slot and a telephone line connected directly to the board by a modular jack. The package also contains software including the main communications program for sending and receiving and several utility programs for converting ASCII documents and PC-generated graphics to a facsimile format. A PC-to-fax board enables users to manipulate data prior to transmission. For example, text from one file can be merged with graphics from another file and then sent via facsimile to another facsimile unit or PC.

---

## Technology Basics

A PC-to-fax board consists of an expansion board installed in a PC expansion slot and a telephone line connected directly to the board by a modular jack. The package also contains software, including the main communications program for sending and receiving, and several utility programs for converting ASCII documents and PC-generated graphics to the facsimile format. A PC-to-fax board enables users to manipulate data prior to transmission. For example, text from one file can be merged with graphics from another file and then sent via facsimile to another facsimile unit or PC.

Although not readily apparent, facsimile machines and microcomputers share a few characteristics. The facsimile machine is really a graphics scanner and printer connected to a dedicated modem. In *send* mode, it converts pages of text or graphics, line by line, into analog or raster images and transmits them over telephone lines. In *receive* mode, it accepts data line by line and prints the images on paper. Microcomputers, although geared primarily to ASCII file formats, can also handle raster images—as bit maps. For example, most

business graphics programs create charts and graphs by this method.

In its most basic application, a PC-to-fax board sends graphics and/or an ASCII file that has been converted into a fax-compatible form; it also receives pages of text and/or graphics from conventional facsimile machines or from PCs with fax option boards and saves them on disk. It is thus a facsimile transmitter and receiver unit without the traditional facsimile hard copy interface. (Note: It cannot use hard copy input without a digital scanner and cannot produce hard copy output without a graphics printer.)

To be the functional equivalent of a standalone facsimile machine, a PC with a fax board must have an attached image scanner and a dot-addressable printer, such as a high-resolution dot matrix or laser device. Except for the input and output paper document interface, a PC-to-fax board performs all of the necessary functions of a conventional facsimile machine. Even if a PC with a fax board is not equipped with these optional peripheral devices, it can still function as a facsimile transmitter and receiver terminal, in which the input and output forms are somewhat nonstandard from the standpoint of traditional facsimile. Without the optional image scanner, a PC-to-fax board can receive text files and computer-generated graphics; without a

---

—By Raymond Falls  
Senior Associate Editor

graphics printer, a PC-to-fax board can send to a mass storage subsystem, a standard facsimile machine, or an electronic-mail network.

Remember that scanning is not necessary for data input. Numerous business applications programs such as spreadsheet, database management, graphics, word processing, and desktop publishing packages, help a PC create a variety of documents, including any combination of charts, graphs, and formatted text. Also, with today's varied and flexible peripheral sharing arrangements, a PC with a fax board does not necessarily have to interface directly with a hard copy output device; a PC attached to a LAN can use the network printer. Furthermore, a PC attached to a server with one of the new LAN-based PC-to-fax boards can share facsimile capability among the nodes.

The primary advantage of a PC-to-fax device lies in its operation as a multifunctional communicating device. It can handle documents with a variety of information content and formats—text, graphics, and halftone image (or any combination), meaning its appeal is not as a substitute for a standard facsimile machine, but as a way to combine file formats and avoid hard copy.

### Standards and Compatibility

Today's facsimile equipment adheres to standards set in 1981 by Study Group XIV of the International Telegraph and Telephone Consultative Committee (CCITT). The CCITT facsimile standards include specifications for the direction of scanning; the size of the scan line; the number of scan lines per millimeter; and phasing, synchronization, and modulation techniques. The standards address three primary machine Groups—1, 2, and 3. Group 3 facsimile machines support grayscale images, encoding an average of 16 gradations between black and white, and is by far the most popular. It transmits a standard page in a minute or less. In addition to the three sets of standards, a fourth (Group 4) has been proposed to define high-resolution digital equipment operating in the range of 56K bits per second (bps). Resolution must be at least 240 by 240 lines per inch (lpi). Although no specifications have been formalized for a standard modem interface for the use of Group 4 machines on public switched telephone network (PSTN) facilities, proposed standards for Group 4 include specifications for digital networks using packet switched and/or circuit switched transmissions. Many industry observers believe that CCITT Group 4 facsimile equipment will not have wide appeal until the Integrated Services Digital Network (ISDN) becomes more widespread.

Almost without exception, today's PC-to-fax boards are compatible with the CCITT Group 3 standard, with a few providing additional Group 2 compatibility.

### Transmission

The digital signal from the PC is converted into a tone series to be sent over the telephone line and reconverted to a digital signal at the other end. Conversion is accomplished on the PC-to-fax board just as it would be in a modem: the data is modulated at transmitter end, and demodulated at the receiving end.

The common types of signal modulation for transmission are amplitude modulation (AM) and frequency modulation (FM). In addition, the boards are classified according to their capacity for handling binary (on and off) information: 2400, 4800, 7200, 9600 bits per second (bps). Usually all binary ones are transmitted at one frequency and all binary zeroes as a tone of another frequency. At the

destination, the receiving board differentiates between the tones and reconverts them into a bitstream, which the PC can use.

### Selection Guidelines

PC-to-fax boards consolidate the power of a microcomputer and sophisticated facsimile machine capabilities in one desktop unit. The most obvious benefit of PC-to-fax over a standalone facsimile machine is that the receiving user can read the document and either delete it, store it, forward it to another party, or print it on a dot matrix or laser printer. Standard fax machines do not provide this level of flexibility. Users, however, should have a measure of computer competency to take full advantage of the boards' capabilities. Standalone facsimile machines are much easier to operate; users do not need microcomputer operational skills. More than one user has lost data because they forgot to save the file before switching to facsimile mode.

The selection of PC-to-fax boards, more so than the selection of conventional facsimile devices, requires a careful evaluation of features and capabilities. PC-to-fax boards can interface with multiple devices and systems, as well as local area networks, and also have a variety of software options. Various equipment features and external factors must be considered before selecting the appropriate unit.

As with any communications system installation, the user organization must first consider its current and future information handling needs. Consideration must be given to the type or class of information to be handled; whether the devices will be connected to an information network; and whether the volume of usage suggests that many departments, not just a few individuals, will need them.

Based on these considerations, selection of a PC-to-fax board can be as simple as adding any PC option card, or as complex (and full of pitfalls) as any communications network installation, where facsimile becomes part of a message delivery system on a large information network.

A PC-to-fax board must offer the basic features that make it fit its intended use. The buyer should consider whether it will be used to supplement existing facsimile machines, or to provide all the facsimile capabilities for the organization and whether the fax board will be used in a standalone configuration or as a local area network (LAN) gateway.

Depending on the application, here are some features to consider.

### Unattended Operation

Unattended sites require an auto dial/auto answer feature and automatic save-to-disk capability. After-hours transmission and reception of documents when communication rates are lower makes fax use more economical, although the computer must be left on.

Many models are also able to send a document to multiple locations automatically.

### Data Compression

Data compression cuts communications costs by eliminating the transmission of redundant information, like spaces. The transmission characteristics of PC-to-fax boards, especially transmission speed, determine its compatibility with competitive products. Document transmission time is important if the Direct Distance Dialing (DDD) switched network is used.

### Background Operation

Many products are able to transmit a document in the background without requiring the CPU, meaning the user doesn't have to switch out of the active application. This capability is only possible if the board has an on-board processor and memory to off-load facsimile processing chores, thus adding to the fax board's price.

The ability has as many proponents as detractors. Whether the PC is used just for facsimile operations depends on the application and volume of message transmission. If the volume of message traffic at a LAN installation requires a dedicated gateway system anyway, then background operation of the fax board is unnecessary. If the volume of message traffic is intermediate and the PC must be used for other applications, then background operation becomes a necessity. If the volume of message traffic is sporadic, or can be handled fully after-hours, then the background operation feature could be unnecessary.

### Image Resolution

The user should determine the type of information that will be transmitted and/or received; i.e., whether it will be straight text, drawings, photographs, or data with several shades or tones of gray. The resolution required is important because usually the greater the resolution, the more expensive the PC-to-fax board.

### Printing Technique

Selecting the correct printing technique to match the image resolution required is also important. For standard text printing, almost any dot matrix printer suffices. For users requiring high image resolution, a more sophisticated printing device is needed. Low resolution is adequate for alphanumeric data equivalent to typed copy or line drawings without fine detail. Higher image resolution is required for fine print or highly detailed images. Even higher resolution is required for photographs or data with several shades or tones of gray. (For example, a scanner with a horizontal resolution of 96 lines per inch can easily read standard 0.0125-inch-high letters. Data with gray tones requires a resolution of at least 200 lpi, since an expanded number of points must be scanned per unit distance.)

### Price

As noted above, price is proportional to resolution, but transmission speed is inversely proportional to resolution. The greater the resolution, the more expensive the PC-to-fax board and the slower the transmission speed, unless the scanning rate is increased to compensate for increased resolution. ■





# Microcomputer Sound Capabilities

## In this report:

Libraries of Sound .....	4
The Final Score .....	5
In Sync .....	6
Multimedia Masters .....	7
Special Software .....	7
The New Music .....	7

## This report will help you to:

- Review the evolution of computer-created music.
- Understand how to use a microcomputer as a “studio” to compose and play music.
- Evaluate and select music software that will suit your audio/video needs.

Over the past few years, the advances in microcomputer graphics have been obvious to even the most casual observer: Higher-resolution images, full-motion digital video, and state-of-the-art animations can be seen in everything from business presentations and logos on the evening news to multimedia extravaganzas.

But what of the sounds that accompany those images? Are we stuck with the pitiful beeps and buzzes of the early microcomputer days? Hardly. Thanks to the advent of hardware and software that rival those of the computer graphics explosion, producers of audio/video presentations—everything from computerized slide shows to scores

for motion pictures—can have professional sound capabilities available right on their desktops (See Figure 1).

## MIDI Made the Difference

The merging of music and microcomputers began innocently with the establishment of MIDI, the industry standard for music synthesizers. MIDI is a protocol for sending digital information over serial lines between electronic musical instruments and equipment, including computers. The information includes note on and note off, key velocity (speed of keystroke), aftertouch (pressure applied after keystroke), pitch bend, modulation wheel, foot pedal, and sound changes. With MIDI, a musician can play a single keyboard and simultaneously trigger a roomful of synthesizers. MIDI guitars, MIDI woodwind instruments, and even

This Datapro report is a reprint of “Sounds of Success” by Dean Friedman, pp. 429-442, from *Byte*, Vol. 15, No. 9, September 1990. Copyright © 1990 by McGraw-Hill Inc. Reprinted with permission.

MIDI acoustic voice trackers are also used to control the array of equipment.

MIDI's potential became obvious when programmers began writing applications that took advantage of its ability to interface synthesizers directly with personal computers. The first and most important of these applications was the music sequencer. Today, there are six categories of music software:

- Sequencers
- Editors/librarians
- Notation programs
- Pattern generators
- Film score utilities
- Everything else (for lack of a better term)

---

### The Sequence of Things

In computer music parlance, a sequencer is a program that records the events and gestures of a musical performance. The electronic equivalent of a player piano roll, music sequencers record the events of a musical performance—but not the actual sounds. The sequencer can play back the instructions to the appropriate synthesizer or sound module, telling it exactly when to trigger its sounds (see Figure 1).

Unlike tape, the audio playback from a sequencer never degrades in quality caused by generations of overdubs or by tape wear. The playback from a sequencer is always first-generation. Unlike tape, a sequenced performance can be sped up or slowed down without transposing the music's pitch, and, conversely, a sequenced performance can be transposed without altering its tempo. With analog tape, pitch and tempo are permanently intertwined; changing tempo automatically changes pitch, and vice versa.

One other fundamental difference between a sequencer and tape (analog or digital) is that, while tape permanently records all the elements of a performance, including the sound that was actually made during the performance, a sequencer only records the skeletal outlines of a musical performance; that is, the timings, durations, and numeric values of the keys that were played. With this information, a musician can experiment with different sounds in the same musical piece, auditioning a

variety of timbres until he or she finds the one that works best.

The sequencer didn't replace multitrack analog or digital recording formats. Those are still necessary for handling acoustic instruments like voice and guitar, and for mixing down to a final two-track master audio format. But the sequencer allows anyone with a personal computer and a few synthesizers to prepare, and in some cases even master, a finished album in his or her own virtual studio.

This basic concept spawned a musical revolution. It had the effect of turning the recording industry on its head. In the process, it created a billion-dollar-a-year market for MIDI devices and peripherals. Suddenly, a new world of polyphonic (more than one note), multitimbral (more than one sound), multitrack, digital music systems became available to anyone who could cough up the price of a cheap PC clone (or an even less-expensive Commodore 64) and a few synthesizers.

Recordings that previously would have cost upwards of \$100,000 are now made for only a few thousand dollars, and even less as MIDI software and hardware evolve.

Dedicated music sequencers had existed prior to MIDI, but it was only when the power of a microcomputer with graphics was added to the MIDI network that sequencing packages flourished into elegant and intuitive music-making programs.

### Sequencing Software

Although the basic job of sequencers remains the same, the extra features in today's computerbased sequencers bring joy to the work of making music. Four of the five main personal computer platforms have high-end professional music sequencers: the Amiga, Macintosh, Atari ST, and IBM PC. The fifth platform, the NeXT machine, has the best standard hardware with its compact disk-quality audio, but it suffers from a dearth of commercial music software.

The list of top-notch, high-end professional music sequencing software packages includes Vision by Opcode, Performer by Mark of the Unicorn, Cubase and Pro24 by Steinberg/Jones, Master Tracks Pro by Passport, Beyond and KCS by Dr. T, Sequencer Plus by Voyetra Technologies, Personal Composer by a company of the same name, and Cakewalk by Twelve Tone Systems. Used regularly in recording studios around the

## Items Discussed

### AmigaVision

Commodore Business  
Machines  
1200 Wilson Dr.  
West Chester, PA 19380  
(215) 431-9100

### Bars & Pipes

Blue Ribbon Bakery  
1248 Clairmont Rd.  
Atlanta, GA 30030  
(404) 377-1514

### Beyond

#### Hitman

#### KCS

#### MIDI Mouse

#### MT-32 Editor/Librarian

#### Tiger Cub

#### X-or

Dr. T  
220 Boylston St.  
Chestnut Hill, MA 02167  
(617) 244-6954

### Cakewalk

#### Sound Globes

Twelve Tone Systems  
165 Bedford St.  
Burlington, MA 01803  
(617) 273-4437

### Clicktracks

#### Master Tracks Pro

Passport  
65 Miramontos St.  
Half Moon Bay, CA 94019  
(415) 726-0280

### Cue

#### EasyVision

#### Galaxy

#### Vision

Opcode  
3641 Haven Dr., Suite A

Menlo Park, CA 94025  
(415) 369-8131

### Cubase

#### Pro 24

Steinberg/Jones  
17700 Raymer St., Suite  
1001  
Northridge, CA 91325  
(818) 993-4091

### Director

MacroMind  
410 Townsend St., Suite  
408  
San Francisco, CA 94107  
(415) 442-0200

### Drummer

Cool Shoes Software  
P.O. Box 391  
Burlington, MA 01803  
(617) 229-9942

### Dynamic Drums

New Wave Software  
P.O. Box 438  
St. Clare Shores, MI  
48080  
(313) 771-4465

### Easy Performer

#### Performer

#### Performer 3.4

Mark of the Unicorn  
222 Third St.  
Cambridge, MA 02142  
(617) 576-2760

### Finale

#### MusicProse

Coda  
1401 East 79th St.  
Minneapolis, MN 55425  
(800) 843-2066

### GenEdit

Hybrid Arts  
8522 National Blvd.  
Los Angeles, CA 90232  
(213) 841-3048

### HookUp

HIP Software  
117 Harvard St., Suite 3  
Cambridge, MA 02139  
(617) 661-2447

### Jambox

#### M

#### Upbeat

Intelligent Music  
116 North Lake Ave.  
Albany, NY 12206  
(518) 434-4110

### Mandala

Very Vivid, Inc.  
P.O. Box 127, Station B  
Toronto, Ontario,  
Canada M5T2T3  
(416) 686-7850

### Master

CMS  
382 North Lemon Ave.  
Walnut, CA 91789  
(714) 594-5051

### MIDIBASIC

Altech Systems  
122 Faris Industrial  
Park Dr.  
Shreveport, LA 71106  
(318) 226-1702

### MIDI Quest

Sound Quest  
1573 Eglinton Ave., W,  
Suite 200  
Toronto, Ontario,  
Canada M6E2G9  
(800) 387-8720

### Monitor

Bartleby Software  
P.O. Box 671112

Dallas, TX 75367  
(214) 363-2967

### Notator

C-Lab/DigiDesign  
1360 Willow Rd., Suite  
101  
Menlo Park, CA 94025  
(415) 327-8811

### Personal Composer

Personal Composer  
2448 76th Ave. SE  
Mercer Island, WA 98040  
(800) 446-8088

### Sequencer Plus

Voyetra Technologies  
333 Fifth Ave.  
Pelham, NY 10803  
(914) 738-4500

### Showmaker

Gold Disk  
P.O. Box 789, Streetsville  
Mississauga, Ontario,  
Canada L5M2C2  
(416) 828-0913

### Super Librarian

Pixel Publishing  
1573 Eglinton Ave.,  
Suite 3  
Toronto, Ontario  
Canada M6E2G9  
(416) 785-3036

### Sybil

Scorpion Systems  
175 Fifth Ave., Suite 2624  
New York, NY 10010  
(415) 864-2956

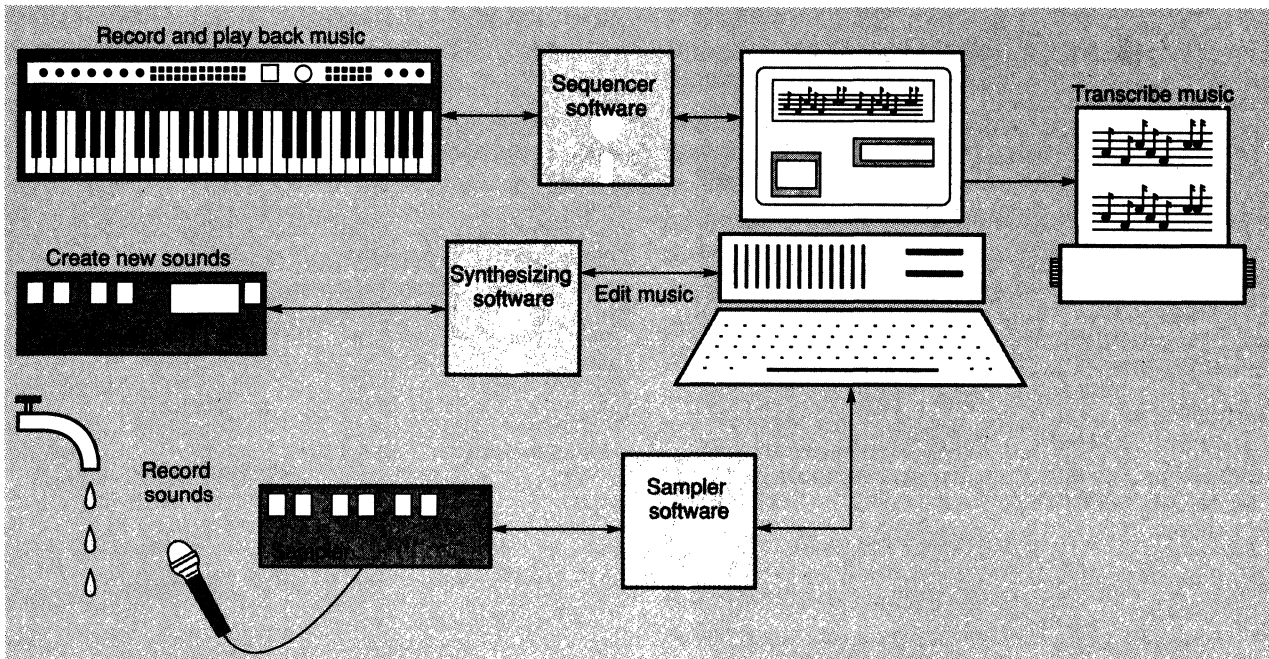
### Synthia

The Other Guys  
P.O. Box H  
Logan, VT 84321  
(800) 942-9402

world, they all provide the means to do graphics editing, since your music sequence is depicted graphically, either in standard musical notation or as rectangles on a pitch/time grid.

The graphics editing tools include the same kinds of cut-and-paste, copy, insert, and delete commands that you'd find on any word processor. The analogy is a good one: Music sequencers let

Figure 1.  
From Concept to Composition: A Micro Music System



The versatility of the personal computer and its graphics screen make it an ideal front end for any electronic music system. Software that performs sequencing, synthesis, sampling, and notation have made it possible to have an entire studio in a box.

you manipulate and manage music in much the same way a word processor lets you move and manipulate text.

Today's sequencers also offer graphics tools for editing other MIDI parameters, such as aftertouch, key velocity, pitch bend, and modulation wheel moves, as well as a host of continuous and intermittent controller messages. Controller messages include additional and often simultaneous performance parameters such as volume-pedal changes, sustain-pedal events, and patch or sound changes.

In a program like Vision, for example, a pitch-bend movement can be graphically edited by simply drawing or reshaping a curve under the affected note or notes. This type of graphics controller editing represents a vast improvement over the first generation of sequencers, in which the musician had to edit a list of hundreds of numbers.

One of the newest implementations of graphics controller editing is *automated fader mixing*. This feature displays a row of 16 or 32 mixing faders, which the musician can assign to control any of

the common MIDI continuous controller parameters (e.g., volume pedal and modulation wheel), as well as additional internal sequencer parameters (e.g., tempo). As an illustration, if the musician assigns faders to the volume levels of all his or her MIDI instruments and then records the sequence of individual fader moves, he or she can program the volume mix as part of the sequencer. The automated mix will keep all relative volume levels in balance throughout a dynamic piece.

### Libraries of Sound

Synthesizers and samplers (a close relative) store the instructions used to create a sound in memory. These instructions are called patches or programs. They contain the voice settings that define each individually stored sound. With the pitifully small display on most keyboards and sound modules—30 to 80 characters—it is very difficult for the musician to edit or design a sound. A typical synthesized sound may have as many as a hundred different variables to describe its harmonic content, filter and amplitude envelopes, pitch, volume,

and so on. It's hard enough to be aware of all the relationships of these variables. It is extremely difficult if you can't see the values all at once.

Voice-editing software displays all of a synthesizer's internal voice parameters simultaneously on a single monitor. What's more, editors can depict complex voice parameters graphically far better than the built-in LCDs of the synthesizer. It's far easier to understand an envelope when you can see its shape than when all you see is a list of numbers.

Some voice editors use the computing power of the computer to manipulate voicing parameters in ways the synthesizer can't. For example, a random-voice-generator option can produce variations on a sound by manipulating the parameters automatically within a specified range.

Once sounds are defined, you can store them using a companion to the voice editor, known as a librarian. This utility is simply a database for storing and retrieving patches or sounds. Many synthesizers can store no more than 100 different sounds internally, but a computer with a librarian can store thousands of patches on a single disk. You can also sort, search, copy, and delete patches.

In the past, owing to the unique architecture of each synthesizer, editors/librarians were instrument-specific. Every time a new synthesizer was released, software houses scrambled to be the first to sell software for it. The recent trend is toward developing more generic editors/librarians. Such programs either come bundled with multiple editors or include a MIDI toolkit that you can use to design your own editor or librarian templates. Some examples are MIDI Quest by Sound Quest, GenEdit by Hybrid Arts, X-or by Dr. T, Master by CMS, Galaxy by Opcode, and Super Librarian by Pixel Publishing.

---

### The Final Score

Since the first hint of MIDI's phenomenal success, electronic musicians have fantasized about being able to sit down at a keyboard, play a performance, and instantly print out a music manuscript in perfect standard musical notation. Easier said than done. It's easy enough to teach the computer to recognize pitch, time, duration, and other objective values. But such decisions as how measures should be divided, where bar lines should go, which notes

should be beamed together, and what kinds of expression markings are needed are all subjective. Such decisions are generally too daunting for the computer to make on its own. Therefore, some editing is usually necessary before a manuscript can be considered complete. It's this post-input step that has given notation packages a reputation for being a real bear to learn.

Some notation programs require input in the form of sequencer files, rather than using live keyboard input. The notation package needs to compile these files before it can produce the music notation. Some of the notation packages have you perform your music directly into the program, and then, after a brief compilation period, they display it in notated form. Several notation programs (e.g., C-Lab's Notator, and Coda's MusicProse and Finale) can actually display your performance in notated form as you perform it.

There are some combination sequencer/notation packages on the market. Personal Composer was the first successful sequencer for the PC, and it has always offered notation as an integral part of its package. Notator and Mark of the Unicorn's Performer 3.4 have fluid, high-end sequencers that can display notation in real time.

The problem with some of the most powerful notation packages is that they are painfully difficult to learn. A professional notation package like Finale is one of the most feature-packed notation packages on the market, but it might take an experienced computer musician up to three months to become proficient. For some users, this learning curve can be justified in view of its enormous power and flexibility.

Fortunately for those of us who barely take time to read the introduction to the manual, the newer notation packages are designed with friendlier and more considerate user interfaces. MusicProse provides a simpler user interface than its predecessor but has less powerful features.

Releasing scaled-down versions of complex programs is becoming something of a fashion in music software applications. While offering reduced features, these versions maintain file compatibility with their more powerful counterparts, but they have a much flatter learning curve. Some examples are EasyVision, Opcode's introductory version of Vision; Easy Performer, Mark of the Unicorn's introductory version of Performer; and Tiger Cub, Dr. T's introductory version of Beyond.

One drawback of composing with computers is that there is no way to derive inspiration by bouncing musical ideas off other musicians. A slew of music applications referred to as Random Pattern Generators and Compositional Aids respond to this shortcoming. While they don't necessarily replace a live musician, random pattern generators do succeed in turning the personal computer into a contributing partner in composing and performing. These applications apply random pattern algorithms to thematic source material. This source material is altered according to parameters established by the composer. Being able to apply the randomizing power of the computer within defined musical limits enables a composer to creatively generate new textures and forms within a theme. In the hands of a musician, this powerful tool can yield impressive results that would not have developed by more conventional means.

Some good programs are Sybil by Scorpion Systems; Jam-box, Upbeat, and M by Intelligent Music; Sound Globes by Twelve Tone Systems; and MIDI Mouse by Dr. T. Sybil is unique in that it offers dynamic real-time performance features. M is one of the finest examples of an interactive pattern generator. It has an appealing and intuitive user interface for modifying musical material.

Some recent sequencer programs have random pattern generators, although they tend to be less powerful than the standalone packages. A standout is the creatively designed music sequencer called Bars & Pipes by Blue Ribbon Bakery. It has a number of built-in randomizing tools, as well as the means to create your own pattern-generating tools from scratch.

## In Sync

To score soundtracks to film and video, it is critical to accurately lock a sequencer's timing onto the pulse of an audio- or videotape machine or another sequencer. Early sequencers accomplished this by locking onto an analog pulse on one of the tracks of the tape, a method known as FSK (Frequency Shifted Keying). It worked well but was prone to dropping out of sync, and it required that both the sequencer and the tape machine always started from the beginning of a performance.

Today, the preferred method of synchronization is via SMPTE (Society of Motion Picture and Television Engineers) coding. As is obvious from

the name, this is the time code standard used in TV and film. It specifies time in frames per second. Most current sequencers address SMPTE time code either directly or indirectly via another standard, MIDI time code, that divides time into beats per quarter note.

The synchronous lock-on is achieved with an SMPTE-to-MIDI converter, which translates the musical timing divisions of MIDI time code into the video- or film-frame realm of SMPTE. Synchronization is accurate to the resolution of a frame. It can begin anywhere instead of requiring a simultaneous start from the beginning of a piece.

All high-end music sequencers have this important synchronizing ability, but sequencers alone don't address all the specialized, complex needs of a musician who is adding sound to visuals. As a result, a new category of music software known as cueing software has quickly evolved. These programs are designed to work hand-in-hand with a music sequencer to automate as much of the scoring process as possible.

Opcode's Cue, Passport's Clicktracks, and Dr. T's Hitman all share the following features:

- They provide a fast and easy way to identify and tag the beginnings and ends of cues in a film (i.e., the points in a film that require music).
- They help calculate optimum tempos, enabling the composer to catch the maximum number of hits (synchronized musical/video events) during a cue.
- They can perform instantaneous time-format conversions, allowing you to translate a single cue point into NTSC or PAL video frames, film frames, or beats per minute.

Some of these programs have limited sequencing features themselves, allowing you to trigger single MIDI events or even complete sequences created in a dedicated sequencer. All the features help enormously in minimizing the most tedious and time-consuming aspects of soundtrack composing, leaving the composer more time to actually compose music.

---

## Multimedia Masters

There are some powerful software packages that allow you to integrate audio and MIDI note events with video, graphics, and animation—in other words, multimedia. One of the most prominent of these is Commodore's AmigaVision, the multimedia icon-based authoring system that is being bundled with the new Amiga 3000. Another is Showmaker by Gold Disk, a time-line-based multimedia sequencer geared toward video production. Both of these products make full use of the Amiga's native multitasking operating system and offer features that allow you to manipulate musical elements within a multimedia event.

---

*The next music software evolution is just now getting off the ground.*

---

MacroMind's Director is designed specifically for multimedia. A unique program from HIP Software, called HookUp, is a quirky icon-driven programming toolkit that allows you to create interactive animations that trigger or are triggered by MIDI events.

One of the most interesting and unusual programs to incorporate music in a multimedia environment is Mandala by Very Vivid. It allows you to employ video input to control animations and musical MIDI events in real time. Standing in front of a video camera, you can play virtual instruments in thin air while actually triggering audio samples.

---

## Special Software

As music software matures and moves beyond its formative years, enhanced MIDI protocols and a collection of MIDI utilities have evolved to address the needs of music and MIDI within the personal computer environment. MIDIBASIC by Altech Systems consists of 12 additional MIDI-related BASIC commands, enabling BASIC programmers to more easily create music applications.

A MIDI utility by Bartleby Software called Monitor allows you to examine, send, and store MIDI data as hexadecimal, decimal, or binary numbers, or as text.

With the proliferation of voice editors that provide features for conventional synthesizers, it wasn't long before developers exploited the audio circuitry in computers themselves. Synthia by The Other Guys uses the four-channel digital-audio hardware of the Amiga. It is essentially a synthesizer/sampler on a disk. New Wave Software's Dynamic Drums turns the Amiga into a polyphonic drum machine. Drummer by Cool Shoes turns the PC into the front end of a drum machine, triggering (via MIDI) any remote drum module or synthesizer.

---

## The New Music

Exotic random pattern generators and new MIDI utilities will continue to come our way in varying forms, and the necessity of incorporating music into multimedia environments will spawn more applications for integrating music with video, graphics, and other media. But it appears that the basic music software tools—the sequencer, the voice editor/librarian, and the notator—are already mature.

The next music software evolution is just now getting off the ground. It is occurring in tandem with an evolution in music hardware: software/hardware packages for digital multitrack recording direct to disk. Opcode and DigiDesign are working together to create a system that will access and control Digidesign's Soundtools system—a two-track, 16-bit, CD-quality digital recording module—from within Opcode's Vision sequencer. The merging of sequencing and digital recording represents the birth of the music workstation, and the ultimate realization of the home recording-studio-in-a-box fantasized by every musician that's ever yearned to produce master recordings.

In a world that is increasingly defined by how well we learn to communicate with and relate to computers, being able to work with them as fluid musical instruments and creative partners capable of warmth and subtlety of expression offers hope that our futures might not be as sterile and unfeeling as many sometimes fear. ■





---

# PC Mice

## In this report:

Mouse Anatomy .....	2
The Resolution Revolution.....	4
Mouse Interfaces.....	4
The Software Perspective.....	5
A Faithful Companion .....	7

## This report will help you to:

- Review the evolution of the PC mouse.
- Understand how a mouse communicates with a personal computer.
- Compare the relative merit of optical versus mechanical mice.

Mice! Suddenly they're everywhere—about a quarter of all PCs users have them. PC mice have grown steadily in popularity since their 1982 introduction. The increased availability of programs that support mice will continue to accelerate this trend. In particular, the phenomenal success of graphical user interfaces (GUIs) for the PC—most notably Microsoft Windows 3.0—is having a dramatic effect on the demand for PC mice. Before long most PCs, like all Macintoshes, will have one scurrying around next to their keyboard.

### The Way it Was

Douglas Engelbart invented the mouse in 1963, at the Stanford Re-

search Institute. At that time Engelbart was exploring various computer input device possibilities. His first prototype mouse was made of wood, with metal disks for rollers that detected the mouse movement. After using the mouse, Engelbart concluded that it was superior to the other alternatives and that it would remain the best pointing device for computer users until something better came along.

Xerox further developed the mouse concept in the early 1970s at its Palo Alto Research Center (PARC), under the direction of Jack S. Hawley. Unlike Engelbart's mouse, which used variable resistors and an A/D conversion circuit, Hawley's was the first digital mouse. Much of Hawley's basic design has been carried into the modern PC mouse.

---

This Datapro report is a reprint of "The Mouse That Roared" by Roger Alford, pp. 395-401, from *BYTE*, Vol. 15, No. 12, November 1990. Copyright © 1990 by McGraw-Hill, Inc. Reprinted with permission.

In 1982, Mouse Systems introduced the first mouse for the IBM PC. With no real software available with mouse support, initial sales of the three-button mouse were primarily to computer users who were curious about the creatures, and to those attracted to the novelty.

Around that time, Microsoft also started seeing the mouse as a device with a lot of potential in the PC marketplace and, being a software company, the company has the wherewithal to encourage mouse use by writing mouse support into its software.

Microsoft introduced its own two-button PC mouse in mid-1983. With the subsequent introduction of such programs as Microsoft Word, and later Windows and Excel, Microsoft showed PC users that a mouse can make working on computers easier and more efficient (and more fun).

When the Macintosh appeared in 1984, sporting a mouse and a user-friendly GUI, users everywhere became even more aware of the benefits of the mouse. Meanwhile, mouse-supporting applications continued to trickle into the PC marketplace.

Mouse vendors further encouraged mouse use by supplying pop-up menus that allowed their mice to work with standard nonmouse applications. Mouse-based PC paint programs also began to appear, and it was common to buy a mouse that included a bundled paint program.

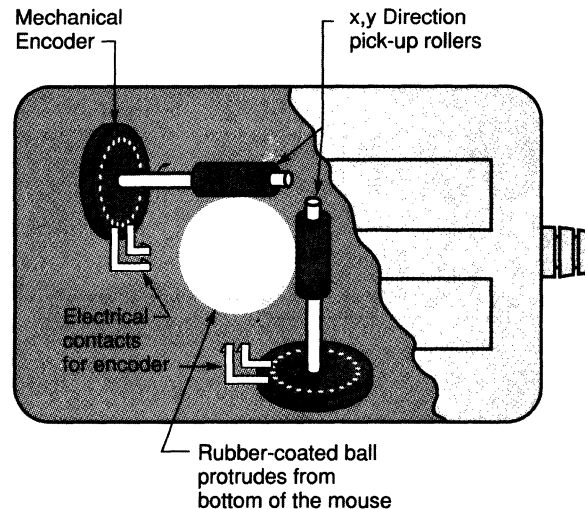
The use of mice on PCs continued to grow. In mid-1988 Microsoft recorded its one-millionth mouse sale and ended the 1990 fiscal year in June with nearly two million mouse sales—about half of all PC mice sold that year. Other major mouse suppliers have also benefited from the increased popularity of mice, including Logitech, Mouse Systems, and IBM. According to International Data Corp. (Framingham, MA), 1989 mouse sales in the U.S. totaled around 3.2 million units, with worldwide sales for that year of around 5.5 million units.

As Engelbart predicted, the mouse has indeed withstood the test of time. There are far more mice on PCs than any of the alternative pointing devices (i.e., trackballs, graphics tablets, light pens, and touch screens).

## Mouse Anatomy

Mice come in two species: mechanical and optical. Mechanical mice, in turn, belong to two subspecies: electromechanical and optomechanical.

Figure 1.  
Electromechanical Mouse



*In an electromechanical mouse, a rubber ball drives the encoders, which make and break electrical contacts.*

Figure 1 illustrates the operation of an electromechanical mouse. A rubber-coated metal ball protrudes from the bottom of the mouse; as you move the mouse, it turns. Two rollers touching the ball record its movements along the  $x$  and  $y$  axes. As the rollers rotate, encoders make and break electrical contacts that send electrical pulses the computer can use to track the mouse.

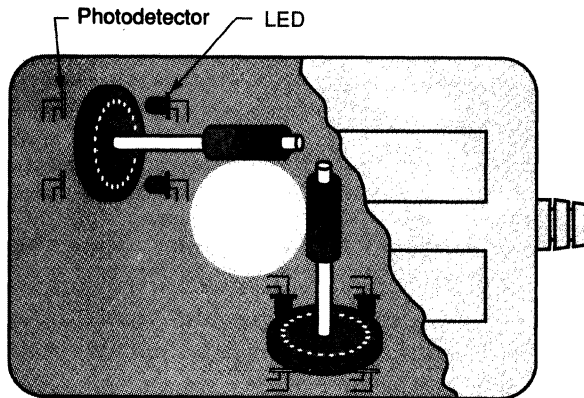
Alternatively, some mechanical mice, like the Manager Mouse from Numonics, don't use a roller ball. Instead, two rollers protrude from the bottom of the mouse to sense the  $x$  and  $y$  directional movements.

The optomechanical mouse illustrated in Figure 2 works differently. LEDs shine through holes in the encoders onto photodetectors. As the rollers rotate, the encoders alternately make and break light beams between the LEDs and the photodetectors. Corresponding electrical signals sent to the computer describe the motions of the mouse.

Note that the optomechanical mouse needs two LED/photodetector pairs in order to determine the direction of rotation. A single LED/photodetector pair can only determine rotational speed.

Figure 3 shows how an optical mouse works. It requires a special reflective mouse pad with a grid of black and blue lines. The mouse has two LEDs that shine onto the mouse pad, one red and one infrared. The reflected light beams reenter the

Figure 2.  
Optomechanical Mouse



As with an electromechanical mouse, a rubber ball inside an optomechanical mouse drives the encoders. In this case, however, LEDs shine through holes in the encoders. The optical encoding scheme eliminates wear on the encoders.

mouse through lenses, and then reflect onto photodetectors. The blue lines absorb the red light, and the black lines absorb the infrared light. As the mouse moves, the pad alternately absorbs and reflects light. The photodetectors detect the “makes” and “breaks,” which the mouse converts to signals that it sends to the PC. As with all species of mice, additional signals tell the computer about push-button events.

Most PC mice have either two or three push buttons (in contrast to the Mac’s single button). Mouse-based PC programs generally require just two buttons, but can often assign a function to a third button. Mouse push buttons can also work in combinations (e.g., two buttons simultaneously) to specify other functions. Some programs support the double-click—two button presses in rapid succession—to specify more functions.

What are the relative merits of optical versus mechanical mice? Optical mouse proponents claim greater reliability for their favorite, thanks to its solid-state, no-moving-parts design. The “opticians” also point out that the optical mouse is maintenance-free, unlike mechanical mice, which require periodic cleaning of the roller ball to eliminate the inevitable build-up of foreign substances. They also claim the optical mouse is more accurate. If an optical mouse moves from one point to another on its mouse pad, then back, the cursor in your screen should be back exactly where it started. In contrast, the mechanical nature of mechanical

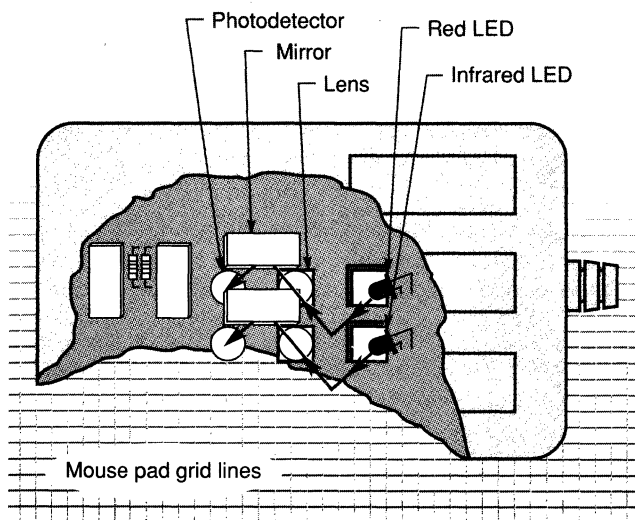
mice makes them more susceptible to slight variations, including minor ball skipping and alterations in the registration of the roller ball to the encoder shafts. Move a mechanical mouse from one point to another and back, and you’ll typically find the cursor slightly off its starting point.

The mechanical-mouse proponents argue that modern mechanical mice have shown no reliability penalty, and that the roller ball rarely needs cleaning—especially when used on a rubber mouse pad. Furthermore, the mechanical mouse doesn’t need a pad, as an optical mouse does. Some users don’t want to give up the desk space, or restrict the mouse to a limited field.

Finally, the mechanical design more readily accommodates higher mouse resolutions. You can cram only so many black and blue lines onto an optical mouse pad before you begin to lose the ability to resolve them.

What about the two species of mechanical mice? Electromechanical mice suffer from a couple of problems that their optomechanical cousins solve. With an electromechanical mouse, the electrical contacts on the encoders can “bounce” a bit. This affects accuracy and requires a compensating circuit design. Electromechanical mice also tend to wear out their encoders, since there are always points of physical contact. The optomechanical design eliminates bounce, and there’s no encoder

Figure 3.  
Optical Mouse



Red and infrared LEDs shine from an optical mouse onto a special pad. Reflected beams pass through lenses, then reflect onto photodetectors.

wear (except at rotational joints). The optoelectronic design of the encoder also supports higher resolution. Most high-resolution mice are optomechanical (although the 350-point-per-inch PC Mouse III optical mouse from Mouse Systems is the exception to this rule).

---

### The Resolution Revolution

The resolution of a mouse refers to the number of points it can detect for every inch of movement. The distance between two adjacent points (the shortest distance the mouse can resolve) is measured in a half-dozen different units. Programmers who work with mice have whimsically coined the unit *mickey*, but the industry is using more common ones, including dots per inch (dpi), counts per inch (cpi), pulses per inch (ppi), and points per inch (another ppi, and the one used for this report).

Early mice, like the original Microsoft mouse, had a resolution of 100 ppi. Most of today's mice have a 200-ppi resolution, as did Microsoft's second- and third-generation mice. Some newer high-resolution mice register between 320 and 400 ppi, including Microsoft's latest 400-ppi entry. There has been some debate over the necessity of resolutions as high as 400 ppi, but some users claim smoother mouse operation on high-resolution screens when using a high-resolution mouse.

---

### Mouse Interfaces

In what form do the signals enter your PC, and how does the PC process them? That depends. Three primary types of mouse interfaces are common in the PC world: bus, serial, and special port.

The earliest mice were bus mice. They came with a half-size interface board that plugged into one of the PC's expansion bus slots; the board drew its power from the expansion bus. The board processed signals from the mouse, and periodically generated interrupts to pass mouse movement and button-press information to the mouse driver.

Microsoft made a substantial contribution to the PC mouse market when it introduced a serial version of its mouse in 1984. The serial mouse could plug into a standard COM1 or COM2 RS-232C serial port. It didn't need a bus interface board or any other external circuitry. The mouse

included a small controller that sent packets of information to the PC via the serial port. The controller required so little power that it could operate without an external power source, simply by drawing its power from the RS-232C request-to-send (RTS) handshake line. This became a trend in the mouse industry, and now most mice are of the serial variety.

I should mention one caution concerning the use of serial mice with laptop computers. Since these mice draw their power from the serial port itself, they expect to see the typical PC voltage of around +12V on the RTS handshake line. When laptops are operating on battery power, however, a lower voltage is often used to generate the serial-port signals. This prevents many serial mice from working properly with the system. If you use the laptop's AC adapter, of course, there won't be a problem.

If you don't count Microsoft's brief flirtation with its Mach 10 PC turbo board (using the company's proprietary InPort mouse interface), IBM was the first company to include a mouse port (aka, "pointing device" port) on its systems. The mouse port on IBM's PS/2 systems (Models 50 and up) is essentially a bus-mouse interface built into the system motherboard.

Some of the newer bus mice have taken a different approach to implementing the PC/mouse interface. Rather than offer two different mice—one serial and one bus—some manufacturers combine the two into a single serial mouse. The "bus interface" in this situation is functionally little more than a standard serial port that maps to an I/O address other than COM1 or COM2.

---

### How Serial Mice Communicate

Serial mice send multiple-byte packets of information to the PC to indicate the directional movement of the mouse and the status of the mouse push buttons. A couple of packet formats have emerged as the predominant standards in the industry. Most applications, however, don't need to worry about them; the mouse driver hides the packet formats.

The two-button Microsoft packet format is the most popular format in use. The packet comprises 3 bytes; only the 7 low-order bits of each byte are significant. The first byte includes the 2 high-order bits or both the *x*- and *y*-position values,

and the status of the two push buttons. The second byte contains the remaining 6 low-order  $x$ -position bits, while the third byte contains the remaining 6 low-order  $y$ -position bits.

The 8-bit binary position values are in two's-complement format (ranging from -128 to +127), with a negative value indicating movement left or up, and a positive value indicating movement right or down. The mouse sends the packet only when there's change of state, such as a movement of the mouse or a press or release of a button. The  $x$ - and  $y$ -position values sent in the packet indicate the number of points the mouse has moved in each direction since the last packet.

Transmitting only an 8-bit value for each direction isn't a limitation—even for high-resolution mice—because the values indicate only the *change* in mouse position since the last packet was sent.

For example, a typical serial mouse operates at 1200 bps. That means each byte needs about 7.5 milliseconds to pass from the mouse to the PC (7 data bits, 1 start bit, and 1 stop bit), and each 3-byte packet takes about 22.5 ms. Each packet can specify a maximum position change value of 127 (in each positive direction), so the mouse can specify a position change of up to 5644 ( $127/0.0225$ ) points per second. Even with a 400-ppi mouse, this scheme allows for movements of over 14 inches per second.

Of course, the baud rate can always be increased if this becomes a limitation. At 9600 bps, a serial mouse using the 3-byte Microsoft packet format can support a velocity of up to 112 inches per second.

The three-button Mouse Systems packet format comprises 5 bytes. The first byte reflects the current state of the three buttons. The second byte specifies the "first"  $x$ -position value and the third byte specifies the "first"  $y$ -position value. The fourth and fifth bytes are similar to the second and third, but specifying the "second"  $x$ - and  $y$ -position values instead of the first; that is, the change in the  $x$ - and  $y$ -positions since the readings sent in the second and third bytes. This can, for example, be helpful in determining mouse velocity.

As with the Microsoft packet format, the  $x$ - and  $y$ -position values are in two's complement format. A positive value indicates movements right or up; a negative value indicates movement left or down.

## The Software Perspective

It is probably obvious that Microsoft has set the standard for PC mice. You'd be hard-pressed to find one that doesn't tout "Microsoft Mouse compatibility."

DOS applications generally access the mouse movement and button information by making calls to a mouse driver. Virtually every PC mouse includes a mouse driver that emulates the Microsoft Mouse driver to make the mouse look like a Microsoft Mouse to the application. Many mice also come with a driver to emulate a Mouse Systems PC Mouse.

Interestingly, the mouse driver interacts directly with the video adapter to control mouse cursor movement. The driver must therefore include support for the video adapter you use to ensure proper operation on your system. Naturally, all current mouse drivers support the standard video adapters, including MDA, CGA, EGA, and VGA, but if you are using something a little newer (like an 8514/A adapter) or something out of the ordinary, the mouse driver may not support it. Check if you are unsure.

Microsoft's mouse driver supports 35 function calls (see Table 1). The driver offers a lot of flexibility to the mouse programmer. While it is not possible to describe all the functions in detail here, I will briefly describe some of them.

The Mouse Reset and Status function (0) sets several mouse parameters to default values (e.g., the mickeys-per-pixel ratio), and returns the current status of the mouse; that is, whether or not the mouse has been found, and which mouse buttons, if any, are currently pressed. This function also hides the mouse cursor on the screen if it is displayed.

The Show Cursor and Hide Cursor functions (1 and 2) control whether or not the mouse displays its cursor on the screen. A counter value determines when to display the cursor. When the counter is 0, the cursor appears, otherwise it does not. The counter decrements with each Hide Cursor call and increments with each Show Cursor (although it cannot be incremented past 0). Thus, it takes three Show Cursor calls to undo three Hide Cursor calls.

The Get Button Status and mouse Position function (3) returns the current status of the mouse buttons and the current cursor position on the screen. Beware, however, that the mouse driver

**Table 1. Microsoft Mouse Driver Function Calls**

Function Number	Description
0	Mouse Reset and Status
1	Show Cursor
2	Hide Cursor
3	Get Button Status and Mouse Position
4	Set Mouse Cursor Position
5	Get Button Press Information
6	Get Button Release Information
7	Set Minimum and Maximum Horizontal Cursor Position
8	Set Minimum and Maximum Vertical Cursor Position
9	Set Graphics Cursor Block
10	Set Text Cursor
11	Read Mouse Motion Counters
12	Set Interrupt Subroutine Call Mask and Address
13	Light Pen Emulation Mode On
14	Light Pen Emulation Mode Off
15	Set Mickey/Pixel Ratio
16	Conditional Off
—	
19	Set double-speed threshold
20	Swap interrupt subroutines
21	Get mouse driver state storage requirements
22	Save mouse driver state
23	Restore mouse driver state
24	Set alternate subroutines call mask and address
25	Get user alternate interrupt address
26	Set mouse sensitivity
27	Get mouse sensitivity
28	Set mouse interrupt rate
29	Set CRT page number
30	Get CRT page number
31	Disable mouse driver
32	Enable mouse driver
33	Software reset
34	Set languages for messages
35	Get language number
36	Get driver version, mouse type, and IRQ number

uses a “virtual screen” matrix for determining the position of its cursor, and that virtual screen isn’t always the same as the physical pixel array on the screen.

In the case of a medium-resolution graphics screen with a 320- by 2-pixel matrix, the mouse’s virtual screen would be 640 by 200 pixels. The virtual screen concept is intended to simplify mouse programming. You can address the virtual screen (which is always a minimum of 640 by 200 pixels) and allow the mouse driver to translate the addressed position to the correct location on the display, based on the current video mode. For some high-resolution EGA and VGA modes, the virtual screen expands to 640 by 350 or 640 by 480 pixels, but for all other modes, the virtual screen remains at 640 by 200 pixels.

Function 15, Set Mickey/Pixel Ratio allows you to adjust the mouse sensitivity by selecting the number of mickeys, or points, required to move the mouse cursor eight pixels on the screen. You can set the value to anything between 1 and 32,767, inclusive. Another way to adjust the mouse sensitivity is to use Function 26, Set Mouse Sensitivity.

Function 36, Get Driver Version, Mouse Type, and IRQ (interrupt request) Number, returns the mouse driver version, the mouse type (e.g., bus, serial, InPort, or PS/2), and the IRQ number. This information can help determine if the current mouse and driver is compatible with the application.

An application can access a mouse driver in a couple of ways. One option is to link a .LIB file containing the driver with the application program. That way, the application supports the mouse directly. More commonly, however, users install the driver by way of the CONFIG.SYS file (DEVICE=MOUSE.SYS0 or the AUTOEXEC.BAT file (MOUSE.COM), and the application accesses the driver functions by making calls to software interrupt 33 hexadecimal.

The *Microsoft Mouse Programmer’s Reference* (Microsoft Press, 1989) fully describes the operation of the Microsoft Mouse driver.

MOUSE.SYS and MOUSE.COM work well enough in the DOS world (although I have seen incompatibilities), but the whole picture changes when you switch to a protected-mode operating system. OS/2 and Unix can’t use a standard MOUSE.SYS driver to allow a mouse to emulate a Microsoft mouse, because such drivers won’t work in protected mode. If these operating systems don’t include support for your mouse, you’ll need a special driver. Generally, you’ll have more options

with a serial mouse—particularly one that supports the Microsoft Mouse packet format.

---

### **ICBM: Infinitely Configurable Ballistic Mice**

An increasing number of mice support a feature known as *ballistic tracking* (or variable acceleration). At times, you may need to use your mouse for some detailed cursor movements at one part of your screen and then move clear across the screen for some further detailed work. Operating at high resolution, the trek across the screen can take a long time, and require several repeated movements of your mouse.

With ballistic tracking, the mouse can detect when you move it faster. As its velocity increases, it automatically changes the number of points per inch to allow faster travel across long distances. As it slows down, it reduces the number of points per inch to again allow more detailed cursor movement.

Ballistic tracking can be implemented with an on-board controller or in the mouse driver software. Although most who have tried it like ballistic tracking (it beats repeatedly pounding your desk with your mouse to get the cursor across the screen), some find it irritating. If you're unsure, make sure your mouse has the option to disable the feature.

---

### **A Faithful Companion**

The mouse has come a long way in the past five years, but in terms of technology, little has changed. The basic mouse design remains essentially the same, with increments in resolution being the only real thing to show for the longevity of the mouse.

Experience has shown that mice in general are quite reliable, most operate basically as well as others, and resolution is often not a big concern. Some users prefer optical mice because there are no moving parts and nothing to clean; the mouse pad, however, takes up a chunk of your valuable desk space. Other users prefer the mechanical mouse to avoid the optical mouse pad; but the roller ball gets dirty and must be cleaned periodically.

A serial mouse or a bus mouse? All other things being equal, it depends on whether you can more easily spare a serial port or an expansion bus slot. The final decision usually comes down to whether you like the size, the style, the color, the length of the tail, the number of push buttons, and the price.

PC mice will continue to grow in popularity. Continuing evolution of the PC mouse will be in the area of ergonomics; I doubt resolution will push much beyond 400 ppi. Other pointing devices, especially trackballs, will gain some ground but the mouse is not likely to give away very much of its cheese. ■





# An Overview of Laser Printers

## In this report:

Image Processing Technologies.....	2
Page Description Languages.....	3

## Datapro Summary

Laser printers are widely used with computers of all sizes, from desktop PCs to mainframes. Today's laser printers are smarter than ever, and cover a wide range of speeds. Many also offer duplex (double-sided) printing. Other types of page printers use light-emitting diode (LED), liquid crystal shutter (LCS), ion deposition, and magnetographic technologies.

## Technology Basics

By definition, laser printers and other page printers produce text and graphics one page at a time. Page printers use one of three printing technologies. The most common by far is electrophotography, which encompasses laser, light-emitting diode (LED), and liquid crystal shutter (LCS) printers. Next comes electrography, represented by ion deposition printers. Magnetography, the third technology, is currently available from only one vendor. We will explain electrophotography in the greatest detail and then how it differs from electrography and magnetography.

### Electrophotography—Laser, LED, and LCS

Laser printers and other page printers produce hard copy in three stages: processing, imaging, and developing.

**Processing.** Electrophotography and other page technologies begin when the printer controller converts text and graphic images into a raster pattern or a grid of dots. The host computer transmits the text or graphics as a stream of ASCII or EBCDIC codes, allowing the printer controller to translate these codes into font patterns or graphics. The translation requires a raster image processor (RIP) that resides inside the printer, in the host computer, or in a separate housing. When the printer controller has received and processed a page of text or graphics, it sends the pattern to the imaging device.

**Imaging.** The imaging device writes the raster image onto the electrophotographic drum (or, in some cases, a belt). Whether laser, LED, or LCS, this process requires light. The electrophotographic drum closely resembles the print mechanism of a copier. But whereas copiers expose a light-sensitive drum with a lamp, electrophotographic printers expose the drum with a precisely controlled light source. In laser printers, the source, a laser beam, sweeps across the surface of the drum, making (in most models) 300 passes to the vertical inch. Mirrors focus the beam on the rotating drum and control the sweeping motion, while the printer controller turns the laser on and off in a controlled fashion. As the beam shines on the charged drum, it exposes it and transcribes the image as a grid pattern of static electric charges. The laser writes a dot by discharging the surface of the drum.

The same static electric patterns can be created with other sources of light. In the case of the LED, a row of tiny light-emitting diodes stretches across the surface of the drum, with 300 diodes to the inch. The printer controller turns each diode on or off up to 300 times as each inch of drum passes. If the diode is on, it writes a spot on the drum by discharging the surface. If it is off, the spot remains charged.

A few printers use LCS technology. This resembles the LED printer—a liquid crystal display stretches across the drum. Like the numbers on the face of a liquid crystal

display watch, the shutter consists of segments that can be made black or clear. Each inch of the shutter has 300 such segments, and each segment becomes black or clear on the command of the printer controller. The difference between these shutters and light-emitting diodes is that the light comes from a fixed source. The shutters simply block the light or allow it to pass. As with the laser and LED technologies, the light selectively discharges the surface of the drum in a grid pattern.

**Developing and Transfer.** After the laser, LED array, or LCS component selectively exposes the drum to light, the drum continues to rotate and passes near a brush loaded with toner. Like the toner in copiers, printer toner consists of tiny particles that bear a positive static electrical charge. The charge pulls the toner particles to the parts of the drum that were exposed to light. These areas print black. Areas not exposed to light are already positively charged, so they repel the toner and do not print.

At this point, the image still resides on the drum. To transfer it to paper, the printer feeds the paper between the rotating drum and a corona wire. The negatively charged corona pulls the positively charged particles of toner off the drum and holds them to the paper. The paper then passes between two heated rollers that fuse the toner particles together and bind them to the surface of the paper. In place of the heated rollers, some printers use only heat or only pressure. Regardless of the fixing method, the page is now printed.

Some printers use the write-white approach. In this case, the imaging mechanism exposes the spots that will print *white*. A different type of toner adheres to the charged parts of the drum—those that will print black. The results are almost identical, except that large areas of white are contaminated with fewer flecks of black than usual.

### Electrography—Ion Deposition

Ion deposition printers resemble electrophotographic printers in that they use a printer controller, drum, and toner to print an image. Processing is the same as that of electrophotographic printers. Ion deposition printers differ mainly in that instead of creating a static charge on a light-sensitive drum, an ion-generating cartridge writes a pattern of negatively charged ions directly on the surface of the drum. In the developing stage, these negatively charged spots attract toner particles; in the fixing stage, pressure (without heat) binds them to the paper. Because of their high speeds, ion deposition models usually serve large, centralized computers, but some new models can be used in microcomputer or workstation networks instead.

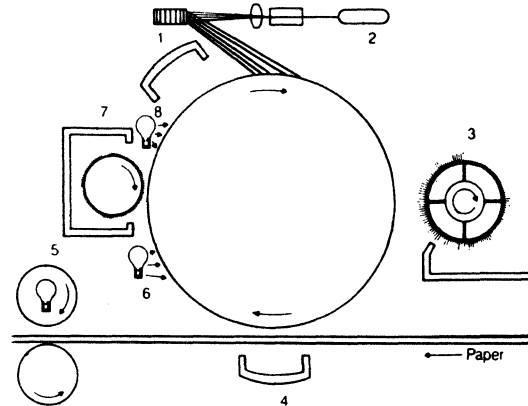
### Magnetography

Magnetographic printers complete the same three stages as electrophotographic and electrographic printers to create a printed page. Once again, processing is identical. In the imaging stage, however, magnetic heads, similar to those in a tape recorder, write the pattern of dots on the drum as magnetic charges. In the developing stage, these magnetic charges attract the toner, which contains iron. Pressure, heat, or a combination of the two fuses the toner and binds it to the paper. Because of their high speeds, magnetographic printers generally serve large, centralized computer systems, but in principle, they can be used in networks of microcomputers and workstations.

### Enhanced Laser Printers

Some recently introduced laser printers print true halftones, crisper text, and higher resolutions. Typically, laser

Figure 1.  
How the Laser Printer Works



The basic components of large and small laser printers are similar: 1. Charge corona; 2. Laser, acousto-optic coupler and rotating polygonal mirror; 3. Toner brush; 4. Transfer corona; 5. Heat/pressure fixing rollers; 6. Light source to discharge drum; 7. Cleaning station to remove residual toner; and 8. Light source to discharge residual static charges.

printers form text and graphics only by putting a dot of toner in a specific pixel location on the page, or by leaving the pixel location blank. Usually, they cannot make one dot larger than another or place it closer to another dot.

With adaptations to the controller, however, some newer laser printers can print a larger or smaller than normal dot, or can shift the location of the dot by less than the usual 300th of an inch. Intel and Hewlett-Packard, for example, have both explored this technology. Intel's Visual Edge card enables the Canon LBP-SX printer engine (the heart of Hewlett-Packard's LaserJet II and III) to print halftone photos by making selected dots smaller or larger. Also, Hewlett-Packard's LaserJet III prints crisp characters and smooth diagonals by making selected dots smaller and shifting their location. Both vendors achieve these effects by modulating the laser beam. In addition, other printers in this report, such as Eicon Technology's Eicon-Laser, use the same method to print at resolutions of 800 dpi and higher. Such printers, however, are rare, require lots of memory and processing power, and work best with special paper and special toner.

### Image Processing Technologies

Page printers differ not only in how they print images, but in how they interpret the information they receive from the host computer. So while the printing technology determines the appearance of the toner on the paper, the types of paper that can be used, and the printing speed, processing technologies determine how versatile the printer is. Two issues dominate: the fonts and the page description language.

### Fonts

Different page printers are suited to different material, and fonts are one of the factors that separate the different models. Knowing how type is classified, what fonts are appropriate for what jobs, and how different printers store fonts will help you choose the right printer and fonts.

**Font Classifications.** A font is composed of a set of characters, usually comprising the letters of the alphabet, numerals, punctuation marks, and sometimes special symbols. Each font has a specific size, weight, width, and slant, and each belongs to a specific type family, such as Courier or Helvetica. Type families are either *serif* (with “feet” at the end of vertical lines) or *sans serif* (vertical lines are cut clean). The type you are reading now is serif (Times Roman), but the running headers at the top of the page are sans serif (Universe). Roman fonts are upright; italic fonts slant forward. *Bold* fonts are thicker and darker than *normal* and *light* fonts. Fonts for publishing are measured by height, in points (1/72 inch). Fonts for simple office documents (the kind that look typewritten) are measured by width, in characters per inch (cpi), also called *pitch*. (The more points, the larger the font; the more characters per inch, the smaller the font.) Fonts can be *extended* to make them wide or *condensed* to make them narrow without changing the height. Characters from typewriter-like fonts (such as Courier) usually are fixed in width or *monospaced*; typographic fonts vary with the width of the character and are *proportionally spaced*.

**Font Styles.** The thousands of fonts available fall into four groups: text, display, decorative, and symbol. Text fonts are the common, everyday fonts, such as Helvetica, Times, and Century. In small sizes, they are suited to body text; in large sizes, to headings. Only about a dozen text fonts appear in most printed English. Display fonts are a little more exotic and numerous. They are the ones used for headings, especially in advertisements and book covers. Decorative fonts look unusual. For example, those known as *script* resemble hand-drawn calligraphy. Symbol or *pi* fonts have the special figures and signs used in science, engineering, mapmaking, music, and many other fields. *Dingbat* fonts are an assortment of common symbols. Most character sets include common symbols and foreign language characters.

**Font Storage.** Two things are important: where the font is stored—in the printer, in a cartridge, or on a disk, and how it is re-created—from a bit map or an outline. The manufacturer installs the resident fonts at the factory. They are the easiest to print but cannot be changed. Font cartridges are easy to plug into the printer but are usually expensive. Most font cartridges hold only a limited number of fonts; so desktop publishers may have to buy several cartridges and swap them between the printer’s one or two slots. In this case, downloaded fonts make more sense—special software copies the font electronically from the computer’s disk to the printer. Though inexpensive and available from many sources in many styles, downloaded fonts must be reinstalled each time the printer is turned on. To get around this problem, some printers have a built-in diskette or hard disk drive for fonts.

Printers store *bit-mapped* fonts as a matrix of dots and simply convert the pattern into pixels during printing. Bit-mapped fonts print fast, but each size of type and each angle (portrait or landscape) is a different font. *Outline* fonts come as a formula that the printer can reconstruct at different sizes and angles, in italic or roman, bold, light, or regular. The trade-off for translating a single outline into a dozen fonts is costly memory and precious time. Also, whether the printer accepts bit-mapped or outline fonts depends on the page description language.

### Page Description Languages

Laser printers are too sophisticated to rely solely on the host computer to prepare pages for print. After all, each

Figure 2.  
Text Rotation

PostScript  
can rotate text and graphics  
into circles of different sizes

This illustrates how page description languages, such as PostScript, can rotate text. Rotation is possible because the program creates text and graphics from geometric formulas.

page contains about 7 million pixels. Page description languages and their hardware counterparts, raster image processors (RIPs), help software deliver pages to the printer. The simplest printers accept a stream of characters from the host computer. Some of the characters are codes that tell it when to start a new line or to change fonts; others are simply text. Graphics are sent separately as a series of bits that correspond to the pixels on the printed page. This scheme, which requires no page description language, lets the printer do either graphics or text—separately. Page description languages allow printers to do both at the same time, since they combine text and graphics in a single set of commands. For each document, the host computer sends a program to the RIP, which compiles and runs it. The RIP is usually inside the printer, but it is sometimes in a separate housing or in an expansion slot on the computer. The RIP translates the program into the pattern of dots that fill the printed page.

**PostScript versus PCL.** Several page description languages are available, but the two most important are Adobe PostScript and Hewlett-Packard Printer Control Language IV (PCL IV). Both are available on a wide selection of laser printers. PostScript is mainly for complex graphics and sophisticated desktop publishing, while PCL IV is for simple jobs, such as word processing and simple desktop publishing. PostScript, which stores fonts as outlines, can print many sizes of type from a single master font. It can also rotate characters to any angle in a full circle, print them hollow or fill them with different patterns, slant them into italic, and perform other special effects. It treats text as graphics and can also do most of these operations to any graphic design. Naturally, it can combine fonts and graphics in one page, on one line, or one on top of the other. PostScript supports shading for fonts and graphics. Every PostScript printer has a processor and megabytes of memory in the RIP, but PostScript is still slow, especially for complex pages. For example, a page printer that can print eight pages per minute prints six or less in PostScript. PostScript is available either built into the printer or in the computer.

Hewlett-Packard’s PCL IV has a few of the capabilities of PostScript. It stores fonts as a bit map and requires different masters for each size, slant, and weight. Only a few PCL IV printers can rotate fonts (and then only 90 degrees), and none can fill or hollow fonts. Apart from rules

and rudimentary patterns, it cannot combine text and graphics, although it can print them separately on the same page. PCL IV does not do special effects or overlays. It needs less memory and processing power than PostScript and entails no licensing fees.

Hewlett-Packard's most recent version of PCL, Level V, has PostScript-like capabilities but has not become an industry standard. PCL V scales fonts, rotates fonts, shadows characters, prints white-on-black characters, fills characters with graphical patterns, and overlays one pattern on another. As with PostScript, PCL V slows down on complex pages. In addition, PCL V lacks some of PostScript's typographic capabilities. More importantly, even though many of the printers in our comparison columns emulate PCL III or IV, none emulates PCL V.

In summary, page printers use a variety of technologies. One of three printing technologies puts images on paper: electrophotographic, electrographic, or magnetographic.

Each uses a different imaging component—a laser, LED, LCS, ion cartridge, or magnet—but the result is a printed page. Two aspects of the printer controller concern the user: fonts and page description languages. Printers come with different fonts, each of which plays a specific role in the printed document. Depending on the model, the fonts may be built into the printer, stored on optional font cartridges, or downloaded from the host computer into the printer's memory. Each method of storage has trade-offs in convenience and cost. Regardless of the method of storage, the fonts may be either bit maps or outlines, the former being less expensive and the latter more versatile. Translating the electronically stored text and graphics requires a raster image processor with the right protocols. Two page description languages provide the standards: Adobe PostScript for publishing and Hewlett-Packard PCL for most office work. ■

# An Overview of Dot Matrix and Ink Jet Printers

## In this report:

Products .....	4
Future .....	4

## Datapro Summary

Dot matrix printers are fast, inexpensive, and one of the few print technologies that produces carbons. Also, they are the least expensive to operate. Ink jet printers are the most unusual class of printers: they employ tiny nozzles that spray tiny droplets of ink on paper. They print letter-quality text and graphics almost silently. Daisy-wheel printers stamp out the letter, number, or symbol from a fully formed die. Though they have good print quality, they print no graphics, cannot automatically switch fonts, and are slow and noisy. Thermal printers use heated printheads to produce images on chemically coated ribbon or paper. This technology is used almost exclusively in color printers and facsimile machines.

## Technology Basics

The one-character-at-a-time or serial printers covered in this report employ the following print methods.

### Dot Matrix

Dot matrix printers form characters out of patterns of dots. A moving printhead incorporates an array of small pins that strike the ribbon against the paper in the pattern of a given character. Different types of dot matrix printers perform the transfer in different ways, but to print simple text on a page the printer receives ASCII code—a standard way to represent letters and numbers—from the computer and fires the correct dot sequence. The letter I, for example, might be formed with a column of five closely packed dots, and so on.

The maximum number of dots that can be used for any one character is known as the matrix. A 9-by-9 matrix comprises 9 dots across and 9 dots down (81 dots total), and a 24-by-36 dot matrix comprises 24 dots across and 36 dots down (864 dots total). Naturally, the more dots in the matrix, or the higher the resolution, the more defined each character will be. Nine-pin printers, still the most common type, produce coarse characters with rough edges, and are used primarily to produce rough drafts.

Recently, manufacturers vastly improved print quality, with many dot matrix printers now employing such developments as multirowed 24-pin printheads, dense dot matrix capabilities, optional font cartridges, special ribbons, double and triple passing, and rack-and-pinion drives for more precise positioning of the printhead. In fact, many dot matrix printers produce characters that at quick glance can pass for typewriter characters. In addition, many dot matrix printers can be made to alter matrix quality or density, producing different modes of print at different speeds. For example, some are capable of producing a low-density, 7-by-9 matrix for fast draft copy and a high-density, 24-by-18 dot matrix for slower, near-letter-quality copy. The matrix varies from vendor to vendor. (The definition of "letter quality" depends on the source. Datapro defines it as a typewritten appearance. Currently, only daisy-wheel/thimble, ink jet, thermal-resistive ribbon, and page printers meet the criteria. Sales talk notwithstanding, today's dot matrix printers produce *near* letter quality text.)

In a dot matrix printer, the print wires that form the characters can protrude in any number of shapes, these matrix printers are much more versatile than fully formed

character printers, allowing the user to select many different sizes of type and different typefaces (e.g., italics) without changing print elements. The user simply sets the control panel, issues a software command, or manipulates DIP switches on the printer instead of manually changing a print wheel. A single line can contain bold, italic, superscript, subscript, and foreign-language characters. Some printers even allow users to design custom characters. The only other way to approach this kind of versatility is with a page printer.

The graphics capabilities of dot matrix printers are also far superior to fully formed character printers—many dot matrix printers are capable of matching a graphics CRT display dot for dot. Matching the characters dot for dot is called a screen dump. In addition to text, users can print charts, graphs, and other designs. Twenty-four-pin printers can print up to 360 by 360 dots per inch, which exceeds the resolution of laser printers. (Actually, the dots are too large for this resolution—thus, they overlap. The exception is a few speciality printers that are really intended as plotters.) Fully formed character printers, on the other hand, can use the period only to place graphics on paper, which pales by comparison to dot matrix printers.

Increasingly, manufacturers are enhancing the graphics capabilities of dot matrix printers by incorporating color. A single color ribbon incorporates yellow, red, green, and black bands. The printer changes ink color by raising or lowering the ribbon to the appropriate band, just as a typewriter does for red or black. Printers with this capability produce colorful graphics for distinctive charts and diagrams. Color dot matrix printers are no longer prohibitively expensive: the \$500 Citizen GSX-140 printer can be augmented with a \$60 color kit and will produce outstanding color graphics.

Dot matrix printers are also much faster than fully formed character printers. They form all images with the same set of pins and waste no time waiting for the right character to come around on a print element. Each pin is also much smaller than a fully formed character printer's impact hammer (less inertia) and can move much faster. Microcomputer dot matrix printers range in speed from 20 to 1,000 cps.

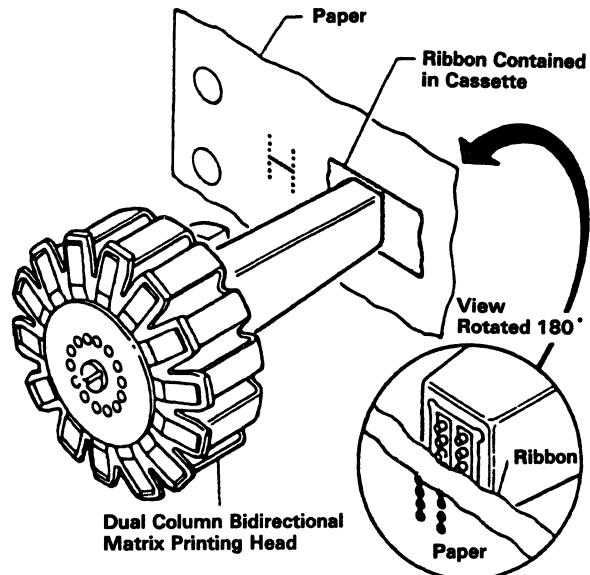
In general, dot matrix printers are faster, more versatile, and produce better graphics than fully formed impact printers. Often they are less expensive, and their print qualities quickly approach those of their fully formed character counterparts. They also incorporate sophisticated features unavailable only a couple of years ago, such as zero-clearance forms tear-off (so a page isn't wasted when tearing off a printed page), intelligent paper-parking, front-panel menus, and optional font cartridges.

### Fully Formed Character Printers

Thanks to the availability of the \$1,000 laser printer, the fully formed character or daisywheel printer is almost a thing of the past. As the prices of competitive technologies come down (PostScript upgrades now available for \$500), look for laser printers to displace dot matrix printers along with the less versatile fully formed character printers. However, for those that want superior character quality for less than the price of a laser printer, the fully formed character printer remains an option.

The fully formed character printer is a direct descendent of the typewriter, using a removable daisy- or thimble-shaped print element to press full-character images through ribbon onto paper. For the most part, the print element has a central stamen-like hub with flat

Figure 1.  
Dot Matrix Printers



*Dot matrix printers produce text and graphics by striking pins (also known as wires or needles) against the ribbon, which is not shown.*

spokes or petals radiating from it. On the end of each of these spokes is one or, at the most, two raised and inverted characters. Thimble printers employ a similar printing method but use a cup-shaped or thimble-shaped element instead of a daisywheel.

The daisywheel or thimble spins around, passing each character by a print slot covered by a ribbon. When the character to be printed is directly in front of the slot, the printer sends a small hammer out to strike the character, pushing it against the ribbon and onto the paper. This requires precise timing.

Due to the time necessary to reposition the printwheel after each character is produced, fully formed character printers are relatively slow—output ranges from about 12 to 90 cps. By contrast, some microcomputer dot matrix printers print up to 1,000 cps, while the world's fastest typist can only type up to 15 cps or 180 words per minute. Fully formed character printers, however, produce characters of the clearest legibility. The printouts from a daisywheel/thimble printer are at least the quality of those from an electric typewriter.

### Ink Jet

Except for the actual process of applying ink to paper, ink jet printers are very similar to impact dot matrix printers. Instead of pins firing out to strike a ribbon, droplets of ink are sprayed onto the paper from nozzles (up to 60) in a specified array. Most print up to 300 dpi resolution, exactly like laser printers.

The majority of microcomputer ink jet printers employ drop-on-demand technology. Here, a set of 4, 9, 12, 30 or more nozzles is attached to an ink reservoir and carried in a printhead. Each nozzle is encircled at the tip with a small piezoelectric crystal, which, when charged with electricity, expands and contracts, which forces out a drop of ink with enough force to fly straight at the paper. Bubble jet printers use an alternative drop-on-demand technology: an electrical element heats the ink in the nozzle, causing a bubble to

## Selection Guidelines

Microcomputer serial printers can print text at a mind-boggling 1,000 cps, produce characters of perfect quality, dazzle with color, produce carbons and multipart forms, and can cost less than \$500. What they can't do are all of these things at once. Fast printers aren't cheap; cheap printers cannot produce quality characters; and color printers aren't fast.

Since trade-offs are inevitable, customers have to choose. First, determine the principle applications the printer will perform—word processing, business forms, color graphics, spreadsheets, or draft copy. If word processing is the principle application, a daisywheel or letter-quality ink jet printer is the best choice. Or, if many applications are of equal importance, including multipart forms, a high-resolution 24-pin dot-matrix printer is best. If straight draft copy is all that is needed, a 9-pin dot matrix printer will do. Color, of course, would require color dot matrix or thermal printers.

### Print Speed

Next, consider speed. Print speeds vary from about 20 cps to 1,000 cps, with speed being a very good indication of printer price. Estimate the volume of printing done each day and divide by prospective print speeds. For example, estimating that each word contains 5 characters and each page about 250 words (1,250 characters), a 55 cps printer would print 198,000 characters or 158 pages per hour. A 225 cps printer would print 540 pages per hour, and so on. Any printer that takes more than four hours, or a half day, to complete the estimated printing is too slow.

### Duty Cycle

Printers have duty cycles indicating the percentage of an 8-hour work day that they are designed to operate, and most printers list 50 percent duty cycles. Also, the rated print speed of a printer is the absolute maximum speed of the printer, not counting blank spaces, carriage returns, tabs, headings, and the like. It is usually an optimistic approximation. As a general rule, the true speed

of a printer is about 70 percent of the rated speed.

### Interface

The interface must match the interface of the computer. Fortunately, there are standards. Printers receive output information from the computer in one of two ways, either *serially*, where information travels over the wire one character after another, or in *parallel*, where several impulses, usually eight, travel over separate wires simultaneously. RS-232-C or RS-422-C are the industry standard serial interfaces, and Centronics parallel is the most common parallel printer interface.

Once the hardware interfaces are matched up, consider software interfaces. Printers have specific commands sets or programming routines that must be addressed if applications are going to run on the printer. The word processing software, for example, must be able to control margin widths, line spacing, carriage returns, and so on. Most software packages list the printers capable of driving the software, and most printers emulate popular printers. For example if a user's packages are compatible with the IBM Proprinter, and the user's C. Itoh printer emulates it, the package will run. To save a lot of trouble, ensure that the

desired applications you use will run on the printer being considered.

### Comparing Models

Once the printer, type, speed, and interfaces have been determined, it is time to compare individual models. Isolate the models from the products listed in Datapro's comparison columns that fit requirements, and compare specifications and price. Then make comparisons and compare print samples. (Always compare print samples of the same size as the smaller the type face, the sharper the characters appear.) Then run identical work applications on the different models and compare results. The winner should be clear.

### Which Printer to Buy?

With many sources of printers, the leading brands in each technology usually provide a safe choice; however, the off-brands sometimes give more for the money. Some small vendors set themselves apart by meeting specific needs (such as bar code fonts) that the leaders pass by. As a general guide, Hewlett-Packard makes the best ink jet printers, and Epson, Panasonic, Okidata, Apple, and IBM make the best dot-matrix printers.

form. This expels a droplet of ink from the nozzle, and the subsequent vacuum draws more ink from the reservoir to refill the nozzle.

Most of the common problems with ink jet printers, namely clogged printheads and water-soluble ink, have been circumvented. Ink now comes in sealed cartridges that include the nozzles, making them less likely to clog. And the ink compositions today are far superior to the older inks and resistant to smearing.

Ink jet printers are ideal for low-cost color printing because they use separate cartridges for the different color inks. The quality is much better than the color dot matrix printers because of the tendency of the ink to spread a bit on impact.

Ink jet printers have many advantages over impact printers. Foremost are their high resolution and quiet performance. Today's 300-dpi desktop ink jet printers almost match the quality of laser printers. Ink jet printers are virtually silent, especially when compared to the 55 to 60 decibels typical of a dot matrix or daisywheel printer. Ink

jet printers also have low maintenance, since they have very few moving parts. Also, since the printhead never touches the surface, it does not degrade with use. The life of an ink jet printhead is typically estimated at about 10 billion characters, compared to the 200 million-character life of many dot matrix printheads.

In regard to graphics, the dots produced by an ink jet printer are better than the dots produced by a dot matrix printer because ink jet dots tend to spread slightly when they hit the paper, giving them a smoother appearance. The filling of solid areas is more uniform, and a great variety of colors can be created by mixing the inks. Also, ink jet colors remain more stable over time than those of a color ribbon. Ink jet printheads have a 100 percent duty cycle, meaning that they can work continuously at the vendors' recommended production level without failure.

On the negative side, ink jet printers cannot produce carbons, and the less expensive models cannot use all types of paper. Paper with a rough surface spreads the ink unevenly and diminishes the print quality. The less expensive

models cannot produce true letter-quality characters. Also the danger exists that ink can dry out in printers left inactive—to address this problem, manufacturers supply printheads in a replaceable cartridges containing ink.

In general, the ink jet printers used for office work produce resolutions of about 100 to 480 dots per inch and range in price from about \$350 to \$15,000 (specialized models for high-resolution color graphics cost much more). Naturally, the more sophisticated the printer, the more nozzles and higher resolution it will have and the more it will cost. Canon, Epson, and Hewlett-Packard are the most popular vendors active in this area.

### Thermal

Another alternative to impact printing is thermal printing, in which pins in a thermal printhead are heated, causing reactions on the chemically coated ribbon or paper. Thermal printing is a relatively quiet and inexpensive way to produce images, and direct thermal printers (in which no ribbon is required) are used mostly in facsimile machines. Color thermal transfer printers (which require a ribbon) generate graphics similar to those made by plotters.

The major drawbacks of direct thermal and thermal transfer printers are poor print quality and the need for special paper. Even typical copier paper may produce inferior results. For this reason, thermal printers have fared better in the office as color graphics printers than as monochrome text printers. An exception to the need for special paper and the poor print quality is the resistive-ribbon technology found in the IBM QuietWriter Printer. The QuietWriter produces letter-quality copy on plain paper (at up to 160 cps). As the name suggests, the low noise level is the main advantage over impact technologies. Users of the QuietWriter have complained of smudges. Canon has also adopted the resistive-ribbon technology.

### Products

- Dot matrix technology saw its rise with microcomputers. At about \$200 to \$2,000 per printer, it does the most for the lowest cost: text in different fonts, sizes, and attributes, including near letter quality text on the 24-pin models; fast drafts; multipart forms; and bit-mapped graphics. Dot matrix printers are also the fastest way to produce multipart forms, short of large and expensive line printers. Dot matrix technology is widely used in desktop, tabletop, and a few freestanding computer printers. Incremental developments in the technology, such as 27-pin printers and multiple-head printers, go a little further to increase the speed and improve the print quality. It is certain that dot matrix printers will be superseded by laser printers for all applications but those that require carbons, and that dot matrix technology will never produce the resolution of lasers, but for the next year or two dot matrix models will continue to be a viable alternative to nonimpact output.
- Daisywheel and thimble technologies are past their heyday, with the few remaining models still out there

doomed to die a slow death. However, letter-quality characters are still used for correspondence, especially in law firms. The output is virtually indistinguishable from typewriter output.

- Ink jet printers have been around since the early 1970s, but they have really caught on recently. Hewlett-Packard has done very well with its inexpensive, plain-paper, 300-dpi printers. Its DeskJet is already a popular choice over dot matrix and daisywheel printers, and its DeskWriter is a perfect Apple-compatible choice for those that cannot afford the Apple LaserWriter. The new ink jet printers are quieter than either rival and produce better results, almost as good as laser printers. Ink jet printers are also used in specialty printing, such as dedicated envelope, label printing, and packaging.
- Thermal printers are mainly for special applications. Because they are quiet, direct thermal printers are used in cash registers and in most facsimile machines. Some lightweight portable printers also use thermal technology because it consumes little power. The plain-paper resistive-ribbon technology that IBM championed with its QuietWriter has had little success—ink jet does the same thing at a lower cost. Color thermal transfer printers are used in many graphics printers.

### Future

It is clear that the refined technologies will bring marked improvements to print quality and costs, although the non-impact technologies, specifically ink jet and light-duty laser printers, will take market share away from the impact technologies, such as dot matrix technology. For now, ink jet printers cost more and print more slowly than the fastest dot matrix printers, but they are coming up in speed and down in price. The new four-ppm laser printers are already available for well under \$1,000. Falling prices mean that many users who could not afford a laser printer for an individual microcomputer now can. As prices drop, the competition will intensify, and some of the small dot matrix printer vendors will leave the market or find alternate technologies. Those remaining will have to compete on the low cost and multipart forms capability of dot matrix printers. Two of the openings that impact and ink jet printer makers will try to fill are for more portable printers and more printers for the Macintosh. The installed base of portable computers is increasing, and relatively few of today's printers are small enough to transport; fewer still run on batteries. This is one area that is not threatened by the influx of laser printers. Apple has promised more help in Version 7 of its Macintosh operating system for companies developing Macintosh printers, and we expect more models for the Macintosh. Such printers must support complex, scalable QuickDraw fonts. We also expect more fonts, including scalable fonts, for 300-dpi ink jet printers used on IBM microcomputers. ■



# Microcomputer Bus Architectures

## In this report:

Micro Channel Architecture vs. the AT Bus.....	3
Micro Channel Architecture vs. EISA .....	5
Micro Channel Architecture vs. NuBus .....	12
Micro Channel Architecture vs. Futurebus .....	13

## Datapro Summary

The world of microcomputer bus architectures is confusing and complex. The list of contending architectures includes Micro Channel architecture, the Extended Industry Standard Architecture (EISA), 32-bit AT architecture, and Apple Computer NuBus. As 32-bit, multitasking systems equipped with expansion boards become more prevalent, the unique characteristics of these bus architectures will become more evident.

In this report, Micro Channel architecture will be compared to other bus architectures used in microcomputers. The objective of this report is to sort out some of the confusion present in the microcomputer industry. And a very confused industry it is! With a broad range of system designs using the various versions of the AT bus, systems based on the Extended Industry Standard Architecture (EISA), and Micro Channel systems with enhanced features of the architecture, it is no wonder decision makers have their hands full these days.

"Which bus to take?" is a prime topic for discussion in almost every organization, and continues to be the subject of numerous articles in the trade press. This decision will form the basis for the purchase of billions of dollars of equipment in the 90s and will affect almost every aspect of information management in the coming decade. This report is designed to help you answer the question: "What is the best choice for microcomputer bus architecture in our organization and how can we manage the change?"

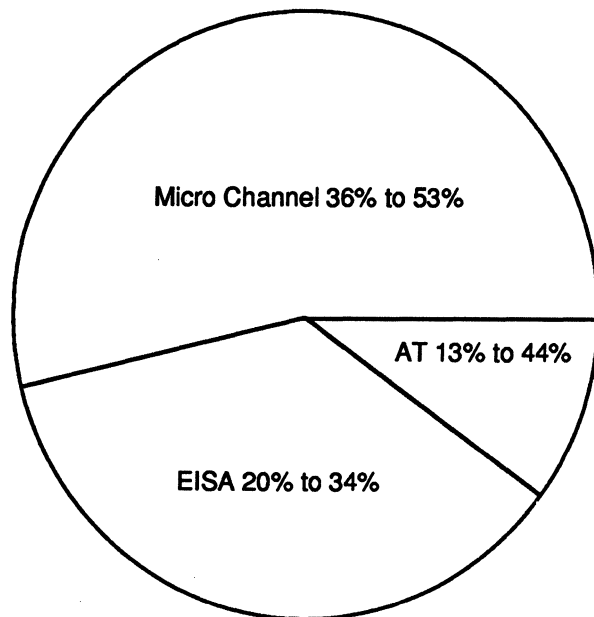
This Datapro report is a reprint of Chapter 6, "Bus Wars: EISA and Other Warriors," pp. 244-271 from *Micro Channel Architecture* by Pat A. Bowlds. Copyright © 1991 by Van Nostrand Reinhold. Reprinted with permission.

## Why Choose?

At this point, you may be asking yourself, "Why even make a choice?" The answer to this question is driven by a number of factors that relate to how computers are used, as well as technology advances. There are a number of forces in the early 1990s that will drive a change from earlier personal computer buses:

1. 386, and later 486, systems will replace 286 and 386SX systems as the entry level workstation. (You don't have to look very far into the future to see this one!)
2. High-frequency (>25 MHz) 32-bit systems will dominate the mid-range, technical workstation, and server arena with increasing presence of RISC processors.
3. Multitasking 32-bit operating systems (OS/2, UNIX) will utilize and require the features of advanced I/O bus architectures, delivering significant performance benefits.
4. I/O-intensive applications such as page printing, multimedia, and image processing will drive 32-bit I/O requirements at the personal workstation level.
5. Migration of function from mainframe systems to "personal computers" (both

Figure 1.  
Composite Estimate of 1993 Shipments Based on Various  
Bus Architectures



standalone and networked) and advanced workstations will impose stringent reliability, data integrity, and fault-tolerant requirements.

These trends are already present to some degree. It is clear that today's purchase decisions should be made with these directions in mind if at all possible. Even though the majority of systems in use today are based on the AT bus, they are simply not capable of addressing these directions adequately. A choice must be made. What the best choice is and when and how it should be made are the real questions.

## The Contenders

The list of major contenders includes Micro Channel architecture, 32-bit AT systems,<sup>1</sup> the Extended Industry Standard Architecture, and Apple's NuBus. Figure 1 is a composite estimate of 32-bit system sales projected by bus architecture for IBM compatibles in 1993. By the numbers, Micro Channel architecture is expected to have the highest share in this environment.

One of the most striking things about the chart is that a significant number of AT-bus systems are still predicted to be sold. There is no question that AT-bus systems will be around for awhile. These systems will continue to be used for traditional personal productivity applications in a single-tasking environment at home and in the office. However, the need for some AT expansion boards will shrink since most of the necessary I/O devices required for the bulk of personal productivity applications will be integrated with the system board.

The conclusion from these projections and other forecasts is that no one bus architecture will dominate the market, but most projections give Micro Channel systems the edge in the 32-bit world. However, these projections may

be more closely related to choice of vendor rather than choice of bus architecture!

Let's get out of projections and examine the offerings of system vendors today. These should give some clues about their directions. Apple is already heavily focused on NuBus. The group of vendors with systems growing out of the IBM AT architecture present a more difficult pattern for projections.

The bus strategy of most of the EISA vendors seems to offer systems based on the AT bus for the traditional PC applications and EISA systems for the high end, positioned closer to the minicomputer and technical workstation market. Compaq, however, is beginning to build lower-priced EISA systems in the midrange and appears to be moving away from the AT bus. Some of the system vendors are building products based on all three buses.

IBM seems to have a strategy of offering the AT bus for entry PC applications, but with much more overlap of Micro Channel systems in the lower price ranges. (There are 16-bit Micro Channel systems but no 16-bit EISA systems offered today.) In the other direction, IBM has extended Micro Channel architecture much further up the price/performance ladder with the RISC System/6000 and ES/9371 systems.

The 32-bit bus market for add-in board vendors is split into three primary segments: Micro Channel, NuBus, and EISA systems. At this time, Micro Channel boards constitute the largest segment and NuBus is second.

In terms of opportunity, there is no doubt that Micro Channel systems offer a more attractive business case for third-party developers than do the other two bus architectures at this time. NuBus boards are offered by fewer vendors than Micro Channel, but many develop only NuBus boards. In the world of IBM and compatibles, it is not unusual to find vendors developing boards for the AT Micro Channel, and EISA buses.

In terms of EISA board development, third-party activity is fairly small at this time but is growing as the installed system base increases. Several of the EISA boards are being developed by the system vendors for their own products and are not available for other systems. Examples include Compaq's intelligent disk array (IDA) controller and Zenith's high performance ESDI disk controller.

You could safely conclude from the preceding discussion that current hardware development activities are focused heavily toward advanced bus architectures. System software development is also targeted toward the 32-bit platforms. What all this points to is that 32-bit systems with 32-bit buses and 32-bit software will rapidly become the desktop and server standard. Because of these obvious directions, it makes sense to plan your current purchases with an eye toward the future.

## Disposable Computers?

One aspect of this discussion that may be bothering you is the rate of obsolescence of equipment. Why not buy the cheapest systems available and "throw them away" in a couple of years? Why plan for the future with today's purchases? You probably already know the answers. Look around and you will see that all that "obsolete" equipment is still with you. You really don't throw it away, you just live with it. That's why it's so important to purchase equipment that complements your future directions instead of inhibiting them.

**Table 1. Bus Architecture Feature Comparison**

Feature	AT (8 MHz)	EISA	NuBus	MC
Superior/Subordinate Structure	Yes	Yes	No	No
Number of Signals	98	188	96	178
Number of Ground Pins	4	21	22	27
Number of Power Pins	6	15	23	18
Address Lines	24	32	32	32
Memory Addressability	16 MB	4 GB	4 GB	4 GB
Data Path Width	8/16	16/32 (EISA Bus)	8/16/32	8/16/32/64
Basic Data Rate	5.3 MB/s	16.7 MB/s	20 MB/s	20 MB/s
Maximum Data Transfer Rate (Published)	8.0 MB/s	33 MB/s	37.5 MB/s	160 MB/s
DMA Channels	7	7	Optional	8 (7 virtual)
16-bit DMA Data Rate (Basic)	1.6 MB/s	4.0 MB/s	N/A	5 MB/s
32-bit DMA Data Rate (Stream/Burst)	N/A	33 MB/s	N/A	10 MB/s
Multiplexed Address and Data Bus	No	No	Yes	No <sup>1</sup>
Bus Master Support	Limited	Yes	Yes	Yes
Arbitration Levels	7	15	# Slots	16
Arbitration	Ctrl	Ctrl	Dstrb	Dstrb
Fairness	No	Round Robin	Always	Optional
Bus Master Priority	N/A	Low	Slot	Flexible
Sync/Async	Sync	Sync	Sync	Async
Interrupt Handling	Edge	Level (EISA Cards)	Level	Level
Interrupt Request Lines	11	11	1	11
Shared Interrupts	No	Yes (EISA Cards)	No	Yes
Audio Bus	No	No	No	Yes
Switchless Setup	No	Yes (EISA Cards)	Yes	Yes
Address/Data Parity	No	Not Defined	Yes	Yes
Adapter ID Number	No	Yes (EISA Cards)	No	Yes
Power Per Slot <sup>2</sup>	N/A	4.5 amp at 5V	2.0 amp at 5V	1.6 amp at 5V
Board Area (sq. in.)	63	63	52	36
Connector Type	Edge	Edge	DIN	Edge

<sup>1</sup>Micro Channel Architecture does not use multiplexed data and address lines to achieve its basic transfer rate. However, multiplexing is used to achieve data rates in excess of 40 MB/sec.

<sup>2</sup>The power per slot is normally a design specification. The values given here are derived from the PS/2 Micro Channel and the Macintosh II system specifications and from the EISA specification. The power specifications for the AT were never published by IBM.

## Micro Channel Architecture vs. the AT Bus

The AT bus was a synchronous design that supported 8- and 16-bit data transfers and was capable (although rarely used) of addressing up to 16 MB of memory (see Table 1). It was designed primarily as an I/O bus to support the 80286 microprocessor of the AT. There was no provision for the accommodation of the i386 microprocessor, such as wider data paths, greater memory addressability, or faster bus speeds.

### Will the Real AT Bus Stand Up?

The AT bus was not really an architecture; it was more of a subset of the AT system design. There were a number of

designs that collectively constituted what is generally called the AT bus. IBM itself had three versions: the 6- and 8-MHz AT systems and the 10-MHz PS/2 Model 30 286. To further complicate matters, very little documentation for the AT was published by IBM. What was published was subject to a broad range of interpretations by board vendors and system manufacturers. Because of this diversity, the compatibility of boards from system to system differed significantly. The widely varying characteristics of the implementations of the AT bus and the incompatibilities from system to system certainly could not be called a bus architecture.

So which version of the AT bus do we compare with Micro Channel architecture? The 8-MHz AT Model 339 probably represents the best case for a comparison. There

were an enormous number of these systems built by IBM and other manufacturers that centered around this design point. It is probably the most representative of the "industry standard AT" if there ever was one. It is the design point that expansion board vendors generally target for compatibility.

### The Fundamental Difference

The fundamental difference between the AT bus and Micro Channel architecture is that the AT bus was designed primarily for a single-tasking single-user environment with a limited number of I/O devices and operations. Performance and flexibility were not primary considerations as they were with Micro Channel architecture. The Micro Channel bus was designed with OS/2 and UNIX in mind, providing many features that support multitasking and multiuser environments. The greater bus band width, enhanced utilization of system resources, improved data integrity characteristics, the method of handling interrupts, more efficient DMA procedures, exception condition procedures, extended address range, and the multimaster capability are examples of Micro Channel architecture features that provide benefits in multitasking and multiuser environments.

One of the most important differences between the AT bus and Micro Channel architecture is the processor-dependent nature of the AT bus. Just as the original PC bus was based on the 8088 microprocessor, the characteristics of the AT bus in IBM's computers were patterned after the 80286. The AT bus signals and frequency were basically the same as the external signals of the microprocessor.<sup>2</sup>

In terms of the structural characteristics, the AT bus is an I/O bus that requires a system board implementation for the system microprocessor. There is no provision for the system microprocessor to reside on an expansion board. It is basically a superior/subordinate structure in which the system microprocessor or the DMA controllers directly or indirectly control all information transfers on the bus. Even though there was a provision for one additional master other than the system microprocessor and the DMA controller, this additional master was not a peer of the other two. In the AT architecture, the DMA controller and the bus master had a subordinate relationship to the system master.

The true bus master capability of the AT bus was rarely used for a number of reasons. It was very difficult, if not impossible, to design an AT bus master that was functional on a number of machines. While the lack of documented timings was a significant problem for AT masters, the absence of a method for changing bus ownership and the lack of a PREEMPT signal were even greater problems. The AT master was required to determine the proper time to release the bus in order for memory refresh to proceed. There was also no effective method for an alternate master to know if a device was requesting a DMA transfer and which DMA channels were available.

The AT's synchronous bus required that transfers to the bus be synchronized to the bus clock rate. This was usually 8 MHz but varied considerably from "industry standard" machine to machine. The result was to substantially slow down bus transfers from an alternate master. Because the Micro Channel bus is asynchronous, an alternate master can transfer data as fast or faster than the system master.

In terms of system resources, the AT bus had two sets of DMA controllers, yielding a total of 8 DMA channels. Four were 8-bit and four were 16-bit channels, one of

which was reserved for use only on the system board. In terms of interrupt capabilities, the AT bus had 15 interrupt request lines, 11 of which could be used by expansion boards.

In terms of data transfer rates, the basic data rate for the 8-MHz version of the AT bus was 5.3 MB/sec, assuming a bus cycle time of 375 ns and one wait state. With no wait states, a maximum of 8 MB/sec could theoretically be achieved. (The basic data rate for the 10-MHz version was 6.7 MB/sec, assuming a bus cycle of 300 ns with one wait state.) Micro Channel architecture defines a data rate that approaches 20 times that of the 8-MHz AT's maximum capability (using the 160 MB/sec streaming-data procedure with a 50 ns cycle time). A comparison of these rates with those of the other bus architectures is shown in Table 1. (The actual data rates are somewhat less due to the cycles associated with memory refresh and arbitration overhead.) Obviously, the data transfer rates of the AT bus are very slow by comparison to all the other buses.

The AT bus used the edge-triggered method of handling interrupts as compared to the level-sensitive method used by Micro Channel architecture. The former is definitely inferior in terms of data integrity, due to the possibility of lost and false interrupts.

The AT bus has some significant problems in terms of its electromagnetic characteristics, a problem later addressed by Micro Channel architecture. Micro Channel architecture contains 4.5 times the number of power and ground pins than the AT. This increase in pins and their strategic connector locations goes a long way toward removing stray RF-induced signals, a common source of errors on the AT.

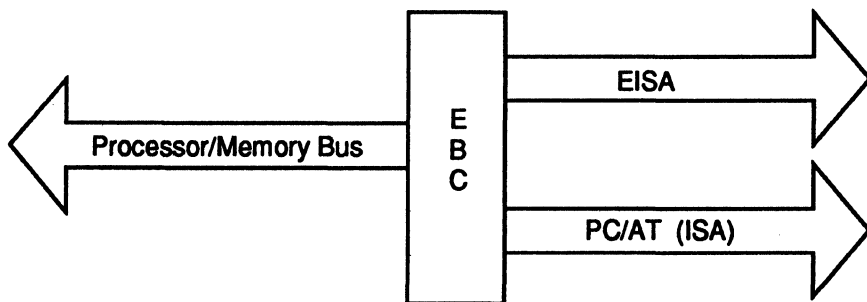
Another significant improvement is Micro Channel's POS function. There is nothing comparable to this feature in AT systems. The DIP switches used for selection of system resources in AT systems (as compared to the "software switches" of Micro Channel architecture) were a source of confusion and error. Several other POS functions are also improvements over the AT bus.

### "32-Bit AT Systems"

The 32-bit AT systems were designed by a number of manufacturers in an attempt to exploit the capabilities of the i386 and i486 microprocessors. These dual bus designs added a separate high-speed 32-bit processor-memory bus while retaining the 16-bit AT I/O bus. Each of these systems is somewhat unique and may require specific memory upgrades from the manufacturer. Unique software may also be required.

Many systems based on these designs have excellent performance capabilities, however. This is especially true in the DOS environment. Why is this? Performance is definitely influenced by the application environment. Single-user, single-tasking applications do not utilize the concurrency features of advanced bus architectures. As a result, DOS performance is most often affected by the raw CPU power of the system rather than I/O bus characteristics. In addition, most of the AT boards rely heavily on the CPU to move information. Because of these factors, these systems have excellent performance capabilities in a DOS environment with AT boards. However, they were not designed for multitasking.

The DOS performance benefit of these systems is their only advantage over the traditional AT design. All the other limitations of the AT bus described in the previous section still apply.



Superior/Subordinate Structure

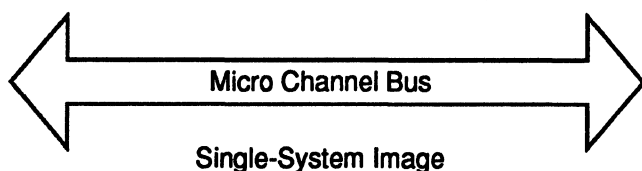


Figure 2. EISA versus Micro Channel Architecture

Even though some Micro Channel systems have a separate processor/memory bus and a system board implementation, it is not a requirement of the architecture.

**Conclusions**

The capabilities of the AT bus are rudimentary in comparison to those of Micro Channel architecture. Its lack of flexibility, performance limitations, limited system resources, and absence of important data integrity characteristics do not place it in the same league with Micro Channel's technical improvements.

Yet there are currently more AT-bus systems sold than any other type, and this is expected to continue for quite some time. An AT-bus system with moderate CPU power continues to be an acceptable choice for the DOS environment for users with modest requirements. However, you are probably not using your money very wisely buying fast 386 and 486 systems based on the AT bus because the I/O subsystem cannot keep up with higher-speed processors.

If your performance requirements are expected to increase in the future, you should consider a Micro Channel system for its bus master and upgrade capabilities alone. A 16-bit Micro Channel system is a good choice because the price is about the same as an AT-bus system from a comparable manufacturer. As CPU-hungry graphical user interfaces become pervasive, even the least sophisticated users will need faster systems. Micro Channel systems are designed for graphical upgrades with powerful adapters, such as IBM's Image Adapter/A.

**Micro Channel Architecture vs. EISA**

In late 1988, a consortium of companies consisting of AST, Compaq, Epson, Hewlett-Packard, NEC, Olivetti, Tandy, Wyse, and Zenith announced support for a 32-bit extension to the AT architecture called the Extended Industry Standard Architecture (EISA). The primary benefits claimed for EISA have been the absence of control by a single manufacturer and the bus's compatibility with AT and PC expansion boards. Many of the features of EISA bear a strong resemblance to those of Micro Channel architecture, including multimaster capability, automatic setup, and high-speed 32-bit data transfers.

In late 1989, the first EISA systems were announced, led by AST, Olivetti, Compaq Computer, Hewlett-Packard, and Zenith. These systems were targeted toward

the high-end server market and high performance desktop applications and have price tags to match. During 1990, many other systems and models by other members of the consortium were announced, a significant number of which featured system boards manufactured by Mylex.

EISA will be described in more detail than the other bus architectures in this report because of its relationship to Micro Channel architecture. It is, of course, software-compatible with previous IBM personal computers. EISA and Micro Channel architecture are viewed as direct "competitors," at least in the 32-bit world. Because of this, an in-depth description of EISA architecture is called for.

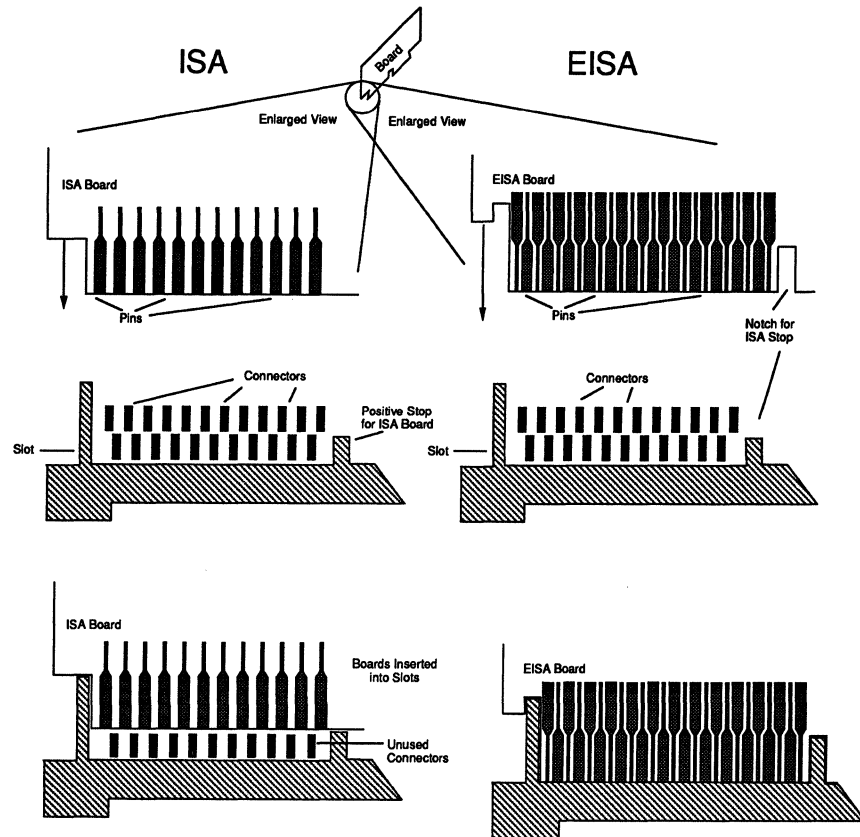
The source of most of this information is the EISA specification version 3.10. Contained in this massive 430-page document is an overview of the architecture, the EISA bus specification, a description of the system board I/O control functions, EISA system configuration, and over 100 figures. The specification provides complete timings, functional characteristics, electrical requirements, and mechanical designs for both the EISA bus and the AT bus (which is referred to as the Industry Standard Bus, or ISA, in the EISA specification).

There is some irony in that the ISA bus is more completely specified by the EISA documentation than by IBM, the originator of the bus. This was necessary since some old AT boards don't meet the ISA specifications and are not fully compatible with EISA systems. From our earlier discussion, this was not an unexpected problem, since some incompatibilities had always existed among AT boards and the various AT clones.

**Overview**

It is somewhat difficult to identify what actually constitutes EISA. Since EISA requires compatibility with AT expansion boards, and also requires a separate processor and memory bus, EISA (the architecture) actually consists of three buses, as illustrated in Figure 2. These three buses are the EISA bus, the processor-memory complex (vendor-specific), and the AT (or ISA) bus. There is, then, a distinction between EISA architecture and the EISA bus.

**Figure 3.**  
**The EISA Connector Design**  
**Accepts Either ISA or EISA**  
**Boards**



The three-bus structure of EISA systems is reminiscent of the split bus architecture that characterized earlier "32-bit AT" systems. Many of the vendors making up the original "gang of nine" actually made AT systems of this type. Their evolution into EISA architecture with its separate processor-memory complex was predictable.

The following is a summary of the major features of EISA:

- Compatibility with most PC and AT expansion boards (called ISA boards in the EISA specification);
- An expansion bus interface supporting both AT and EISA boards;
- 32-bit memory addressing for the CPU, memory, and EISA devices;
- 16- and 32-bit data transfers for EISA attached devices;
- An arbitration procedure that provides for bus ownership by up to 15 arbitrating devices, one of which must be the system CPU;
- Synchronous data transfers, including basic and burst modes;
- Automatic configuration of system and EISA expansion boards;
- A procedure for translation and transfer of information between EISA and AT (ISA) masters and slaves;
- Faster direct memory access procedures;
- Data transfer rates of up to 33 MB/sec;
- Shared interrupts on the EISA bus; and
- Edge-triggered and level-sensitive interrupt handling.

Sound familiar? There is a strong similarity between Micro Channel architecture and EISA. The major technical features are summarized in Table 1. The major differences lie in their different structural philosophies, the degree of processor independence, maximum data transfer rates, and in some missing features (for instance, address and data parity). And, of course, EISA systems can accept AT boards and Micro Channel systems cannot.

### Structural Philosophy

Let's return to the simplistic diagram in Figure 2. EISA architecture is based on a superior/subordinate structural philosophy as was the AT design. There is an assumption that the CPU is resident on the system board. It is unclear if future EISA designs will accommodate microprocessors other than Intel's i386 and i486 or future members of the X86 family, although it is feasible.<sup>3</sup>

The system board manages all the translations between the three buses and serves as the "switch" that controls movement of information among the three buses. The processor-memory bus can operate independently from the ISA and the EISA buses, but they cannot operate simultaneously since they share some common signals.

In EISA architecture, the CPU-memory complex, the AT expansion boards, and the EISA expansion boards have different views of the system. This is in contrast to the basic Micro Channel architecture in which there is a single system image, as illustrated in Figure 2. Although there are a number of Micro Channel systems that have separate processor/memory buses, each device attached to the Micro Channel bus has the same image of the system and follows the same architectural rules. The point is that EISA *requires* a more complex structure, whereas it is *optional* with Micro Channel architecture.

### EISA Connector Characteristics and AT Compatibility

Perhaps the strongest selling point for EISA systems is that they are architected to be compatible with most PC and AT boards. The connectors for EISA boards are a superset of the 16-bit AT connectors. Physically, the PC and AT boards can be installed in the same slots as the EISA boards. A section of this clever connector design is shown in Figure 3. The connector is a two-level affair with the 98 upper pins being compatible with the ISA boards and the lower pins keyed in such a way that only the newer EISA cards can make contact with both levels (188 total pins).

An additional compatibility consideration is that some of the older PC and AT cards were designed to work at speeds much slower than the EISA bus basic transfer rate. For example, the original PC boards were designed for the 4.77-MHz 8088 microprocessor. The current vintage of EISA implementations are based on a synchronous design of 8.25 or 8.33 MHz.<sup>4</sup> The conclusion is that some PC and AT boards may not work in EISA systems due to these timing considerations. This limitation may not be significant, however, since most of the slower boards are older and will probably not be migrated to EISA systems.

### Participants

Based on the EISA specifications, bus masters, memory slaves, I/O slaves, and DMA slaves are the participants that are allowed to use the bus. Theoretically, information can be transferred from any device on any bus to another device located on its own bus or another bus. However, there are some problems related to the fact that the ISA bus only allows one and two byte transfers, but the host bus and the EISA bus allow one, two, three, and four byte transfers. Because of this alignment problem, the system board is required to manage all transfers, adding some additional overhead. Peer-to-peer transfers between ISA and EISA devices are not enabled by current implementations, but are theoretically possible whenever the alignment problem is solved. Micro Channel architecture does not have this limitation, since its participants are self-aligning and do not require the intervention of the system board logic to manage transfers.

### EISA/ISA Signals

A summary of the EISA signals is shown in Table 2. There are a total of 188 ISA and EISA pins. EISA requires 64 additional logic signals and another 26 power and ground connections. There are no more than four signal pins between each pair of power and ground pins. This proximity of power and ground pins enhances electromagnetic compatibility and reduces crosstalk between pins. Micro Channel connectors have very similar characteristics in terms of the number of power and ground pins and their distribution.

**Table 2. EISA Signals**

Pin Description	ISA	EISA Added	Total
Ground	4	17	21
+5V	3	8	11
-5V	1		1
+12V	1	1	2
-12V	1		1
Address	27	23	50
Data	16	16	32
Data/Address Control	14	14	28
Additional Control	6	2	8
Interrupt Request	11		11
DMA Request	7		7
DMA Acknowledge	7		7
MFG Specific		4	4
Reserved		5	5
Total	98	90	188

Signal Type	Signal	Function
Address Lines	LA2 through LA16	Latchable address bits.
	-LA24 through -LA31	Address bits 24 through 31 expand the address range to 32 bits. These lines are driven by the CPU or bus master.
Data Lines	D16 through D31	Data lines 16 through 31 used to expand the data bus to 32 bits. Driven by the CPU, bus master, or slave.
Timing and Control	-BE0 through BE3	Indicates which of the four data byte are active. Driven by the CPU or bus master.
	-CMD	Used to indicate a data cycle on the EISA bus. Supplied by the system board.
	-START	Supplied by the system board and used to control the timing of the start of an EISA bus cycle.
	M/-IO	Indicates a memory cycle when high and an I/O cycle when low for EISA cards. Supplied by the CPU or bus master.

**Table 2. EISA Signals (Continued)**

Signal Type	Signal	Function
	W/-R	Indicates a write cycle when high and a read cycle when low for EISA cards. Supplied by the CPU or bus master.
	-EX16	Driven by an EISA slave to indicate that it can support 16-bit transfers.
Timing and Control	-EX32	Driven by an EISA slave to indicate that it can support 32-bit transfers.
	EXRDY	Driven by an EISA slave to indicate that it needs wait states.
	-SLBURST	Driven by an EISA slave to indicate that it can support burst cycles.
	-MSBURST	Driven by a bus master or the CPU to indicate to a slave that it can perform burst cycles.
	-LOCK	Used to prevent the DMA controller from re-arbitrating for the bus. Driven by the CPU or bus master.
	-MREQx	Used by the CPU or bus master to request use of the bus (slot-specific).
	-MAKx	Used to indicate to the EISA bus master in slot x that it may take control of the bus. Driven by the system board.

However, the older AT boards cannot access the additional power and ground pins of the EISA connectors. Because of this, some of the problems of electromagnetic interference in the AT design are transferred to EISA when AT boards are used in EISA systems. While the EISA connector is definitely an improvement over the older design, the electromagnetic interference design problems remain, and the potential for data integrity problems still exists.

The first 98 signals are identical to those of the AT bus. This layout provides for electrical compatibility of the ISA boards used in EISA systems and AT systems. The EISA boards use a number of these signals but not all of them. This is due mostly to the 32-bit characteristics of EISA, the faster timing characteristics of the EISA boards, and the use of unlatched address signals in EISA. The original PC provided latched address signals. The AT provided for unlatched versions for the bits 17-23. The EISA bus provides for all address signals to be unlatched, including those of the original PC and the eight new bits associated with the

32-bit bus. This is done to provide the address information earlier in the cycle, in order to enhance data rates.

In addition, there are some differences in the characteristics of ISA signals on the EISA bus compared to the way they were used on the AT bus. The most significant difference is in the system clock (BCLK) signal. In the AT, this signal had a fixed 6-MHz or 8-MHz frequency, with a fixed duty cycle. In EISA, the signal frequency is 8.333 MHz at a 50 percent duty cycle or 6 MHz at a 33 percent duty cycle. This is done to provide higher data rates on the EISA bus while retaining compatibility with most of the older boards.

The EISA unique signals are of several varieties. Many of these are required for 32-bit operations and are taken by additional address, data, and control signals. There are also lines used to signal the start and stop of EISA bus cycles, supplied by the system board. There is another line used to signal burst capabilities by masters and slaves. There are also signals for masters to request the bus and to acknowledge to the master that it may take control of the bus. These signals are used for the same purpose as is the arbitration bus in Micro Channel architecture (that is, to pass ownership from one master to another). There are some significant differences, however, in EISA and Micro Channel bus arbitration and prioritization of arbitrating devices. These differences are discussed later in this report.

There are no audio bus or audio signals defined by EISA. For systems used as servers, this is probably not an important feature, however. It is more appropriate for the desktop environment. But there is nothing in EISA that is comparable to the video extension signals of Micro Channel architecture. This is the feature that allows graphics boards to access the video capabilities on the system board and paves the way for advanced multimedia applications. Again, this is probably not an important consideration in the EISA server world since graphics capabilities are not as important there.

There are no address and parity signals defined in EISA architecture, a significant difference compared to Micro Channel architecture. This is surprising, since several of the EISA systems introduced to date have been in the high end server arena, where address and data parity checking has been "standard" for many years. There are also no signals defined for anything comparable to the streaming-data procedures of Micro Channel architecture.

There are four EISA signals that can be used for unique functions by a given manufacturer. In addition, there are five reserved pins. The function of these manufacturer-specific signals and reserved pins is open for speculation at this time. One guess is that the reserved pins may be used for implementation of something like the streaming-data procedures defined in Micro Channel architecture or for address and data parity. There do not appear to be enough reserved pins to implement both of these features, however.

### Timing Considerations and Data Transfers

The EISA bus is a synchronous bus. Bus cycles use a clock located on the system board as the reference for the transfer. The system board is the source of almost all activity, and has the job of adjusting the bus clock characteristics to achieve the best performance of the CPU and memory. Because of these adjustable clock characteristics, a variety of cycle types that cover a range of speeds and transfer types are provided. There are, of course, numerous types



of transfers and timings required for transferring information from one bus to the other and from the many versions of ISA cards to other devices.

At this time, all EISA systems use a set of Intel chips that control all these various transactions. The set of four chips is collectively called the 82350 chip set. Three of these chips are used on the system board and the fourth chip is designed as a bus master interface to be used on expansion boards. These three system board chips support Intel's i386 and i486 processors. There are different versions required for the 25-MHz and 33-MHz microprocessors. The heart of EISA implementations is the EISA BUS CONTROLLER (the EBC). This chip acts as a traffic cop for the movement of information within the system. It provides the interface to the host CPU and the other buses, translates the data bus width to and from the three buses, controls 8-, 16-, and 32-bit masters and slaves, and numerous other functions.

The Intel chip set is almost inseparable from EISA architecture. It is difficult to imagine an implementation of EISA that does not use a very similar design, if compatibility with ISA boards is a part of the architecture. Control of the timings and bus translations is exclusively in its province. The design of the chip set is also directly related to the microprocessor characteristics on the CPU-Memory bus, which is keyed to Intel CISC microprocessors.

Micro Channel architecture contrasts strongly with these fixed characteristics of EISA, if current systems are representative. There is no direct relationship between the type of microprocessor nor its timing characteristics and the timings of the bus cycles because of the processor independent and asynchronous nature of the Micro Channel bus. This is true in the older system as well as in newly designed systems. Existing systems can accommodate higher speed and different types of system masters. Future systems can also be designed with different masters on the system board.

### EISA DMA Operations

DMA operations are very similar to those of the AT bus. EISA provides for full 32-bit DMA transfers instead of the 16-bit AT capabilities. The transfers are parallel as opposed to the serial transfers of Micro Channel DMA operations. Thus, DMA operations for EISA are single step if both the source and the target have the same data path width and reside on the same bus. If the data paths are different, however, multiple steps are required. For example, transferring information from 32-bit devices to 8-bit devices requires four steps. The data transfers are conducted at the rate of the slowest participant. Micro Channel's serial DMA capabilities are conducted at the maximum data rates between the participants, which may be different for the serial operations.

EISA defines several different types of DMA operations, depending on the characteristics of the various devices. The slowest DMA speeds are designed for the slowest ISA cards. The fastest DMA speeds are used for transfers to and from EISA cards that are designed to support burst transfers. Table 3 summarizes the various DMA data rates supported by EISA and Micro Channel architecture. These are maximum data rates as opposed to sustained data rates and do not include any overhead or memory refresh cycles, which would, of course, reduce these numbers.

**Table 3. Micro Channel and EISA Data Transfer Rates**

Transfer Type System	Micro Channel	EISA System EISA Cards	EISA ISA Cards
DMA Basic (MB/sec)			
8-bit	2.5/5.0*	2.0	1.0/1.3/2.0
16-bit	5/10	4.0	2.0/2.6/4.0
32-bit	10/20	8.0	N/A
DMA Stream/Burst (MB/sec)			
8-bit	N/A	8.2	N/A
16-bit	10/20	16.5	N/A
32-bit	20/40	33.0	N/A
64-bit	40/80	N/A	N/A
Bus Master Basic (MB/sec)			
8-bit	5	N/A	2.6
16-bit	10	8.2	5.3/8.2
32-bit	20	16.5	N/A
Bus Master Stream/Burst (MB/sec)			
8-bit	N/A	N/A	N/A
16-bit	20	16.5	N/A
32-bit	40	33.0	N/A
64-bit	80	N/A	N/A

\*The lower number is for third-party DMA transfers (two steps). The higher number is for DMA operations using bus masters (one step).

The two architectures appear to have a different philosophy in terms of the role of the DMA controller in DMA transfers. Micro Channel architecture focuses on the bus master as the prime mover of information to and from memory. The role of the DMA controller is secondary and is there basically for compatibility.

The data rates involving bus master transfers for the two architectures (Table 3) show definite advantages for Micro Channel architecture. However, the faster data rate that has been implemented in Micro Channel systems to date is 40 MB/sec. In the case of EISA, the fastest rate is 33 MB/sec.

The chart shows that maximum data rates for Micro Channel systems greatly exceed what is currently defined by EISA, but, so far, actual implemented data rates are not greatly different. EISA has not defined any direction for increasing data rates beyond the current 33 MB/sec maximum. While it is possible to increase the clock speed for EISA systems, loss of compatibility with ISA boards may result. There are some other possibilities for increasing data rates in EISA systems, but the currently implemented data rates are probably more than adequate for today's environments.

### EISA vs. Micro Channel Arbitration

EISA systems implement a centralized arbitration method (Figure 4) as opposed to the distributed method used in Micro Channel systems. In the EISA method, the central arbitration logic actually determines which of the devices is granted control of the bus. This is in contrast to the distributed arbitration method used in Micro Channel systems, in which the arbitrating devices determine ownership.

EISA systems use a rotating arbitration method. Bus ownership priority is established by their position in the rotation rather than the order in which they assert requests. Compatibility with ISA cards imposes certain restrictions on the arbitration priorities for EISA. Because memory cards may reside on the ISA bus, memory refresh cycles must be made available for these cards on a timely basis. In addition, ISA cards that function as DMA slaves must also be served in a timely fashion. Both ISA memory cards and DMA devices were designed with highly specific timing characteristics and must be serviced first.

The EISA arbitration scheme is illustrated in Figure 5. EISA provides for a fixed DMA priority arbitration sequence and, alternately, a rotating sequence. In the first case, bus access is granted sequentially to the highest priority DMA channel, the refresh controller, and either the system master or a bus master in rotational sequence. In the second case, the only difference is that there is another rotation involving the DMA channels.

The DMA subsystem is given some preference in order to provide compatibility with existing ISA DMA devices, however. The EISA specification devotes several pages to the characterization of the latencies that may be experienced in the various arbitration situations. Some of these latencies can be quite long, requiring extensive buffering to be used on EISA bus masters.

A possible arbitration sequence is shown below:

1. MEMORY REFRESH
2. SYSTEM MASTER
3. DMA CHANNEL 0
4. BUS MASTER 1
5. DMA CHANNEL 1
6. BUS MASTER 2
7. DMA CHANNEL 0
8. MEMORY REFRESH
9. SYSTEM MASTER
10. DMA CHANNEL 4
11. ETC.

The main differences between the two arbitration procedures is the flexibility and programmability of Micro

Channel priorities compared to the relatively fixed assignments of the EISA architecture. Thus, priorities can be changed depending on the application environment. In the case of EISA, it is possible for a critical bus master to wait a very long time until it gets serviced, especially when there are many devices requesting the bus. It is also possible for masters on the ISA bus to dominate the bus since there doesn't appear to be a provision for EISA masters to preempt ISA masters.

Micro Channel architecture provides for 16 arbitration levels and EISA 15. The system master "owns" one of these levels in both architectures (DMA Channel 4 for EISA Systems). Micro Channel architecture provides the flexibility to assign all 15 of the remaining levels to be used for any combination of DMA channels and bus masters even though current implementations use eight of these levels for DMA channels. In order to increase EISA DMA channels beyond seven, a physical connector change is required.

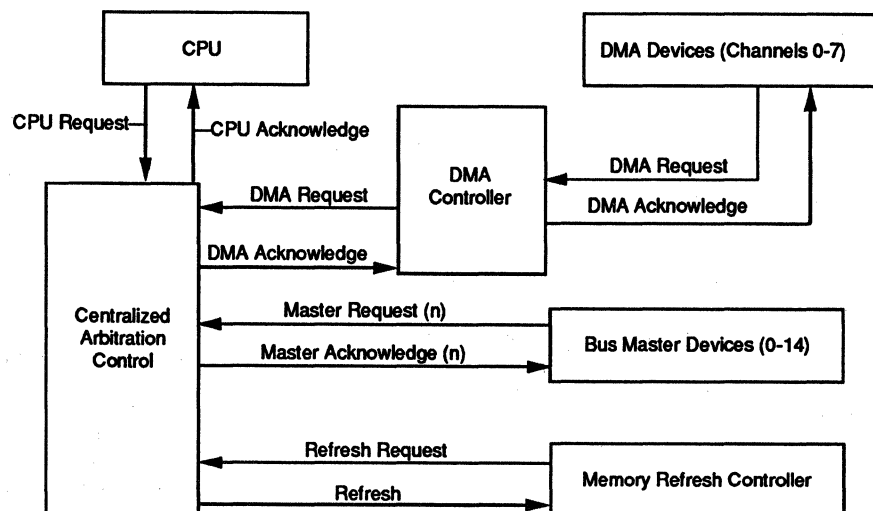
Overall, the EISA arbitration scheme is fundamentally sound. It represents a reasonable compromise for bus ownership between ISA and EISA devices. Even though some of the priorities may appear to be reversed in some cases, the difference will probably not be apparent in most EISA systems. The lack of flexibility will be most apparent to designers of EISA boards in the near term and will cause costs to be somewhat higher due to the added buffer requirements.

#### EISA Automatic System Configuration

EISA systems have a partial method of configuring systems automatically. Devices that exist on the system board (for instance, serial ports, parallel ports and VGA) can be fully configured automatically. EISA expansion boards are given the option of including configuration files with their products that can be used for automatic configuration provided the boards are designed with programmability in mind. EISA and ISA boards that have switches require manual switch setting. There is, of course, no way to make ISA boards fully programmable since they were not designed for this purpose.

EISA uses a slot-specific method for assigning I/O addresses. No two slots have the same I/O address range, so conflicts are avoided for EISA cards. This is similar to the approach used by NuBus described later in this report.

Figure 4.  
EISA Arbitration Is  
Centralized Using the  
Centralized Arbitration  
Control Logic



This geographic addressing, as it is called, provides a simple method for a device to be selected. Micro Channel boards have to be smart enough to determine if they are the selected device.

EISA also includes a product identification mechanism similar to that used for Micro Channel boards. The identifier of each product is selected by the manufacturer and does not need the approval of the other members of the consortium. The numbers are registered with BCPR Services, the legal firm that is responsible for the distribution of the EISA specification. This is in contrast to Micro Channel identification numbers, which are supplied by IBM. In the case of EISA boards, configuration problems are possible if the selection guidelines are not followed. For Micro Channel boards, each one has a unique identification code that cannot be used by another device. Both schemes are workable, however.

For those devices that are programmable for automatic configuration, the process for EISA boards is almost identical to that of Micro Channel boards. A summary of the steps is shown below:

1. The configuration files are read.
2. The system resources are allocated to create a conflict-free system.
3. The configuration information is saved on a diskette.
4. The configuration is then written into the system's nonvolatile memory.
5. The configuration information contained in the system's nonvolatile memory on powering up the system is read, and the expansion boards are "initialized" or are given their resource assignments.

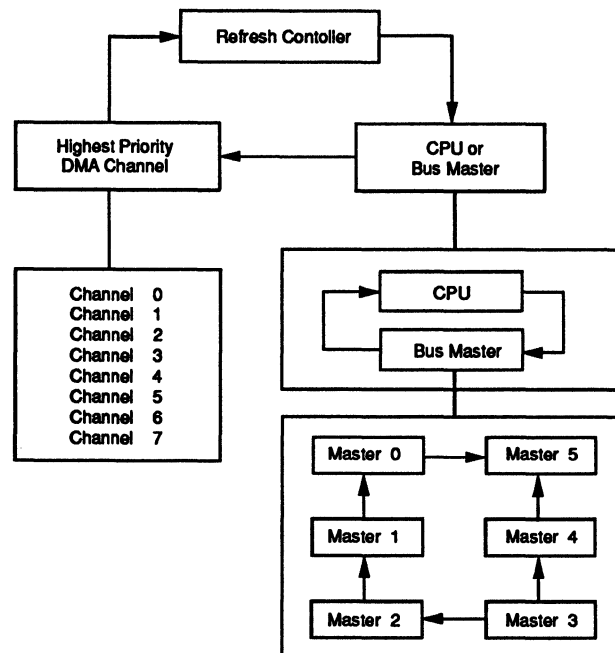
Overall, the most important difference between EISA's and Micro Channel's automatic configuration schemes is that Micro Channel provides a complete solution, allowing all devices to be configured. Only the EISA expansion boards and EISA system devices can participate in automatic setup. One still must manually set the switches on the old boards. In some cases, the allowable switch settings for ISA boards are not readily available, which makes setup difficult to say the least.

**Interrupt Handling Procedures**

EISA provides for 11 interrupts, which are assigned at setup time. ISA levels are set using switches on the boards and some are hard-wired (cannot be changed). In the case of EISA devices, interrupt handling uses a level-sensitive method similar to that used by Micro Channel systems. ISA devices require edge-triggered interrupt handling. Even though level-sensitive interrupt handling is preferable, there is no way to change the design of the older ISA boards. However, each interrupt line can be programmed for either level-sensitive or edge-triggered mode so that both can exist in the same systems. Because of this, EISA devices can share interrupt lines.

EISA defines a procedure for sharing interrupt request lines similar to that of Micro Channel architecture. Of course, any interrupt lines assigned to ISA cards cannot be shared since there was no provision for sharing for these boards. In addition, some ISA boards are required to use specific interrupt lines, and this could result in unresolvable conflicts.

Figure 5. EISA Arbitration Involves a Round-Robin Procedure



**Technical Summary: EISA vs. Micro Channel Architecture**

Let's summarize. The number one technical advantage that EISA architecture has over Micro Channel architecture is its compatibility with AT boards. If this is important, then EISA definitely has the advantage. However, many of these boards were built using older designs and slower logic. Because of the high-end applications that EISA systems are currently targeting, AT boards won't meet the needs, so compatibility may not be as important. This will be especially true after a reasonable number of EISA expansion boards are available.

The price of compatibility is high. The EISA approach is unnecessarily complex because of it, and EISA bus masters have low priorities as a result. The bus speed is not as fast as it could have been if it were not for the ISA compatibility requirement. Some of the EMC problems of ISA cards also carry over to the EISA environment. Automatic configuration and interrupt sharing are limited to EISA boards.

Obviously, both Micro Channel and EISA architectures have many common features and offer significant advantages over the AT architecture. In some cases, EISA offers a partial advantage due to its compatibility with AT expansion boards. Its flexibility and performance capabilities are likewise limited by the same factors.

Based on the available implementations today for systems in the 386/486 category, the apparent technical features in the two buses are not greatly different if only EISA boards are used, although Micro Channel has the edge. What is important is that there are a large number of advanced function Micro Channel boards available today, but very few EISA boards. Over time, this situation may change, but the very large numbers of Micro Channel systems represent a larger opportunity for board vendors, with less risk than the relatively low EISA volumes.

However, the real differences in the two architectures do not lie in today's 386/486 implementations, but in other processor families and future implementations. There is no doubt that Micro Channel is far more flexible, has higher performance capabilities, and is a much richer and broader architecture than EISA. It has superior data integrity features that make it much more appropriate for use in a multitasking or multiuser environment than EISA. These features have already been implemented in the RISC System/6000 family and are anticipated in other systems in the near future.

### Micro Channel Architecture vs. NuBus

NuBus is an industry standard bus that originated at M.I.T. in the late 1970s. The goals of NuBus were the creation of a simple 32-bit backplane bus that was independent of a particular CPU type and was also suitable for multiprocessor systems. This "bus-based" concept was first described by M.I.T. professors Steve Ward and Chris Terman in 1979 and was later specified and prototyped. Western Digital worked with the M.I.T. group to enhance the original concepts and to refine the engineering. The result of this joint effort was the essence of the present NuBus. In 1983, the work was moved to Texas Instruments, who further pursued its development. (Texas Instruments also owns the trademark and several related patents.) The bus was presented to an IEEE working group with the hope that it could become a "sister" bus to Futurebus, described later in this report. Unfortunately, this did not happen, and a separate work group for NuBus was established in 1984. The result of this work group activity is the ANSI/IEEE Std. 1196-1987, the current bus definition.

Today, NuBus is used in the Macintosh II and the Next system. Texas Instruments also used the bus in its System 1500, a multiprocessor system, and the Explorer LX, a system used for artificial intelligence development.

NuBus is an elegantly simple bus. It has only 96 signals, 45 of which are ground and power pins. This small pin count earns it top honors for delivering the most function with the fewest signals (Table 4). Part of this simplicity and low pin count is due to its multiplexed data and address bus. The address bus is also used to transfer data similar to Micro Channel's 64-bit streaming-data procedure. However, the multiplexing in NuBus yields 32-bit data transfers.

### Table 4. NuBus Signals

NuBus gets top honors for its advanced bus features with a minimum number of signals.

Signal Type	Signal	Function
Address/Data	AD31-AD0	Address and Data Lines (Multiplexed)
Timing and Control	Start	Start
	ACK	Acknowledge
	TM0	Transfer Mode 0
	TM1	Transfer Mode 1
	CLK	Clock
	NMRQ	Non-master request Provides an interrupt mechanism for slaves

### Table 4. NuBus Signals (Continued)

Signal Type	Signal	Function
Arbitration	ARB3-ARBO RQST	Used for arbitration Request
Parity	SP SPV	System Parity System Parity Valid
Slot ID	ID3-ID0	Slot identification
Electrical	+5VDC +12VDC -12VDC -5.2VDC GND	Power lines for various voltages
Miscellaneous	RESET PFW	Resets the system Power Fail Warning

NuBus has many other features in common with Micro Channel architecture. It is a 32-bit high data rate bus that supports multiple masters with an arbitration bus and fairness algorithm, has automatic configuration capabilities, block move capabilities, superior electromagnetic characteristics, parity features, and operates at 10 MHz. It is a synchronous bus architecture with some of the timing flexibility of an asynchronous bus. The number of clock periods is allowed to vary, providing a range of timings for attached devices.

The multiple master capability of NuBus allows a Mac to do a reasonable imitation of an IBM PC by supporting an expansion board with a 386 microprocessor. Conversely, installing an expansion board with a Motorola 680X0 microprocessor (the Mac II's "engine") in a Micro Channel system allows that system to run Motorola 680X0 applications with the appropriate software support.

Both NuBus and Micro Channel architectures are "democratic" in nature. That is, the expansion board devices are not subordinate to the devices on the system board. In both Micro Channel and NuBus systems (for example, the Mac II), the system microprocessor effectively occupies a "slot" that has the same rights and privileges as the other expansion boards.

A closer look reveals some significant differences, however. NuBus has no third-party DMA transfers using a DMA controller. Transfers are between bus masters and slaves. As you learned previously, these transfers are more efficient than third-party DMA transactions, anyway.

Another difference lies in the method of handling interrupts, where NuBus interrupt capabilities are limited in comparison to Micro Channel architecture. NuBus defines a single interrupt signal to be shared by all devices. The software must poll the slots to determine the source of the interrupt. Another interrupt procedure requires a device to be implemented with write transactions into an area of memory that is monitored by that processor. This can only be used by a master however. The Macintosh II uses an interrupt handling method that is not defined by NuBus. It adds signals for each slot to be used for interrupts signals to the 680X0.

The absence of DMA capabilities and the method of handling interrupts are some of the reasons why IBM did not adopt NuBus for its 32-bit bus standard. The lack of DMA and the different method of handling interrupts would have prohibited software compatibility with the PC and the AT.

Another significant difference in the two buses lies in the basic philosophy of resource sharing by multiprocessors. NuBus has chosen to manage the sharing of resources such as system memory in the hardware architecture. IBM has chosen to handle resource sharing utilizing software, the Subsystem Control Block Architecture. The use of control block architecture as a programming model in multiprocessor implementations such as the System/370 is a proven approach to resource sharing and one that IBM chose because of its flexibility. Since SCB Architecture can be tailored to fit the varied operating system environments of Micro Channel systems, it is the better choice.

### Micro Channel Architecture vs. Futurebus

Futurebus has several similarities to NuBus. It is a multiplexed bus that uses the same 32 bus lines for data and address on a single 96-pin connector. As mentioned earlier, the goal was compatibility for the two buses.

Futurebus also has numerous features in common with Micro Channel architecture. It is a multiple master processor-independent architecture featuring a separate arbitration bus, burst mode and streaming-data procedures, and high data rates.

Futurebus is an asynchronous bus, does not require a central clock, and should be able to accommodate faster technologies in the future, while retaining compatibility with slower devices. Sound familiar? Micro Channel's asynchronous characteristics also provide this capability.

Another important attribute of Futurebus is that there is no limitation on its data rate. Since there is no maximum defined, the nod goes to Futurebus on this one compared to Micro Channel's 160 MB/sec, the currently defined maximum. However, the same technology that will allow Futurebus to achieve higher data rates can probably be applied to Micro Channel architecture, with a connector change. The significance of data rates in excess of 100 MB/sec remains to be seen, however, since the bandwidth of either bus is not expected to inhibit performance for a long time.

Futurebus designers have paid very close attention to resource sharing among multiprocessors. One unique aspect of Futurebus is the ability to support three-party transactions involving a master, a slave, and a third-party cache or multiple slaves. This contrasts with Micro Channel architecture, in which there are only two-party transactions, involving a master and a slave. The problem of maintaining valid information in multiprocessor caches residing on the bus is not resolved in Micro Channel architecture. It is not clear how IBM and other Micro Channel manufacturers plan to proceed in this area, but they may choose to extend Subsystem Control Block architecture. However, the consequence of not implementing three-party bus procedures are additional cycles to ensure that all caches have the latest information. In very fast buses, this problem would probably not result in performance losses, however, since the number of bus cycles would not be the limiting factor.

Futurebus also has a number of features that are designed specifically for fault-tolerance. A power cable and procedures are defined that allow the live insertion and removal of a board in an operating system. Another important fault-tolerant feature in Futurebus is that it makes provision for redundant buses. These features are a level above what is currently a part of Micro Channel architecture. The use of redundant boards and dynamic reconfiguration is, however, a key part of Micro Channel architec-

ture. The difference is that there is no provision for the live insertion and removal of cards.

To summarize, Futurebus and the Micro Channel bus have a great deal in common. However, in terms of implementation, Micro Channel systems are far ahead. Most of the advantages of Futurebus truly lie in the future and are not well-defined in terms of implementation. Futurebus is viewed as being a more advanced bus architecture than Micro Channel by many, but there are many unanswered implementation questions. Perhaps Futurebus is an appropriate name!

### Conclusions, Technically Speaking

The conclusions of almost every technical analyst would give Micro Channel architecture the edge in terms of technical merits over the other contenders, with the possible exception of Futurebus. This author's technical ranking of the architectures would look like this:

Micro Channel  
NuBus  
EISA  
AT

(I have chosen not to include Futurebus in this ranking because of the lack of implementation in systems. While Futurebus has many desirable features, ranking it would be purely rhetorical.)

The above ranking of bus architectures may not be so obvious in today's systems and computing environments, however. This is due to a variety of reasons. Some of these features may not be incorporated in systems built today, for example. Or, the bus data rate does not translate into a performance benefit because the speed of the I/O devices is so slow. Adapter cards may not be available that demonstrate these benefits. Some of the advantages depend on the development of software support in the future. In terms of error rates, the end user may not be aware of the true error rates of these systems because the operating system covers them up.

The framework of this conclusion is strategic, however. In the long term, these differences will become more and more apparent. The pervasiveness of multitasking and 32-bit operating systems and expansion boards in the near future will differentiate these bus architectures. As I/O devices become faster and faster, the data rate differences will become more important. The processor-independent nature of Micro Channel architecture and its asynchronous characteristics will provide technical longevity.

### References

<sup>1</sup>These systems are those with 32-bit microprocessors (such as the 386SX, i386, and i486) and 16-bit I/O buses.

<sup>2</sup>The AT specification also defined a capability for the microprocessor to operate at a higher frequency than the bus. This capability was never used by IBM, but was incorporated in some AT-compatibles made by other manufacturers. This feature was used in high-speed 286 and 386 systems.

<sup>3</sup>The EISA bus architecture has a very close relationship to the Intel 80X86 family of microprocessors as did the AT bus. However, it would be possible to accommodate a system CPU with a different instruction set, using fairly complex bus interface logic in order to "translate" the signals. This is analogous to the IBM RT with its RISC processor and AT I/O bus.

<sup>4</sup>The bus speed is a factor of the microprocessor frequency. For example, systems with 25-MHz and 33-MHz microprocessors have bus speeds of 8.33 MHz and 8.25 MHz, respectively. ■



# Extended Industry Standard Architecture (EISA)

## In this report:

A Backward Glance.....	2
The Bus Stops Here.....	3
Knowing Your ABCs .....	5
Fringe Benefits.....	5
Micro Channel vs. EISA .....	5
EISA Takes Its Stand.....	6
Make New Friends but Keep the Old .....	7

## This report will help you to:

- Learn the differences between the EISA and Micro Channel buses.
- Trace the development of the EISA standard from the ISA standard.
- Learn the features and benefits of the EISA bus.

Supercomputing has a definite ring to it. It even has a magical quality. It makes you think of the HAL 9000 or of buildings full of giant computers that make the earth spin or change gravity. But when you analyze the concept of supercomputing, you find a combination of elements, each of which contributes to the whole.

The goal of any supercomputing architecture is to obtain increased performance through increased system throughput. One of the more important ways to increase throughput is to free up your processor (or processors) from such mundane tasks as system I/O.

Supercomputers use either a separate I/O processor or a separate computer

(often a minicomputer) to handle I/O to and from main memory.

Until the advent of sophisticated buses such as the Micro Channel architecture from IBM, Extended Industry Standard Architecture from a collaborative effort nicknamed the "Gang of Nine" (AST, Compaq, Epson, Hewlett-Packard, NEC, Olivetti, Tandy, Wyse, and Zenith), and Nu-Bus from Apple, personal computers handled I/O the old-fashioned way, by moving every byte through the processor when performing I/O with main memory. Now, however, the new buses perform I/O without tying up the CPU. The EISA bus master is the latest example of using such a scheme to improve I/O for microcomputers.

EISA, as the name suggests, is an extension of the original Industry Standard Architecture bus (the PC AT bus). Since EISA evolved from ISA, it kept all the preceding technology. Thus, you can use your current

---

This Datapro report is a reprint of "Join the EISA Evolution" by Min-Hur Whang and Joe Kua, pp. 241-247, from *Byte*, Vol. 15, No. 5, May 1990. Copyright © 1990 by McGraw-Hill Inc. Reprinted with permission.

IBM-compatible expansion cards on an EISA machine. The Micro Channel architecture doesn't offer this evolutionary approach, so simple items like network cards, SCSI adapter cards, and modems have to be specially designed to support its I/O bus. EISA, however, is already supported.

## A Backward Glance

The system bus for the original IBM PC was an 8-bit extension of the Intel 8088 microprocessor. The 8088 has only 20 address lines, restricting the amount of addressable memory on the PC to about

1 megabyte. Since today's memory-hungry operating systems (e.g., OS/2, Unix, and Novell NetWare) require more memory than that, the 8-bit system has become obsolete. In addition, the original 8-bit system bus offered only an eight-level edge-triggered interrupt and four DMA channels whose transfer speeds ranged from about 250,000 bps to 500,000 bps.

The next generation of bus was the 16-bit PC AT bus, which became the ISA bus. Improvements in the microprocessor (with Intel's 286) and in the number of addressable memory locations (with 24 address lines) brought the total available memory to 16 MB. The edge-triggered interrupts increased from eight to 15 levels, and the DMA channels from four to seven.

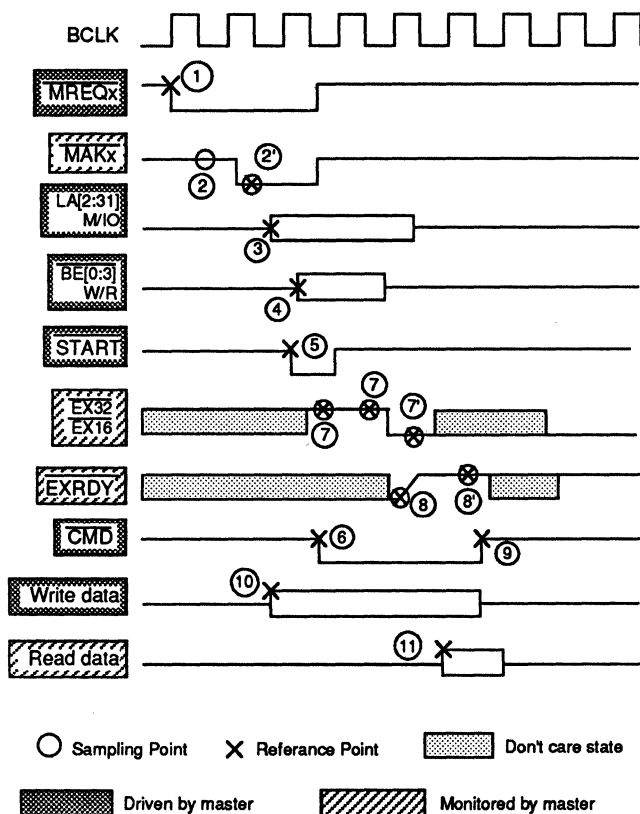
The speed of the 286 supports 8-bit and 16-bit modes with DMA data transfer rates of 100,000 bps to 400,000 bps, and 400,000 bps to 800,000 bps, respectively. Note that the new 8-bit DMA mode is slower than the original PC bus. Intel had to sacrifice some speed to maintain downward compatibility to 8-bit devices. One benefit of the 16-bit bus is that you can continue to use any 8-bit cards you have, because ISA is a superset of the 8-bit PC bus.

From a technological point of view, the problem of I/O bus design became obvious with the arrival of the Intel 386 and i486 processors. ISA just doesn't cut it. An I/O bus for these applications workhorses needs the following additional capabilities:

- more memory capacity;
- multiple bus-master devices with high-speed burst-transfer rates;
- a 32-bit data transfer rate for the CPU, DMA, and bus-master devices;
- enhanced DMA arbitration and transfer rates;
- level-triggered (shareable) interrupts; and
- automatic configuration of system and expansion boards.

To promote these features, two completely new bus architectures, IBM's Micro Channel and EISA, were introduced. A generation gap exists between EISA and the current reigning ISA buses.

Figure A.  
EISA Standard Transfer Cycle



This timing diagram of a standard data transfer cycle shows how the master and slave use signals to synchronize their activity. When writing, the master dumps the data onto the bus when it starts the transfer. The slave reads the data and indicates when it is through using the EXRDY line. When reading, the master waits until the slave indicates it is done before sampling the data bus.



## Requesting a Transfer

An EISA bus master supports two types of data transfer cycles for moving data between itself and main memory or an I/O slave. The cycles are either standard or burst. A third cycle type, the compressed cycle, is available only to the CPU. EISA cycle types are summarized in Table A.

In a standard cycle, an EISA bus master is completely synchronous with the bus clock (BCLK), which provides synchronization with the main system clock. The bus clock usually operates at frequencies of between 8.33 MHz and 6 MHz, although its period is sometimes extended to synchronize with the host CPU or other motherboard devices.

An EISA bus master drives nine signal lines on the bus and monitors five others (see Table B). Using these signals, you can examine how a bus master transfers data to memory or to I/O slaves within a standard cycle. To follow the events in a standard transfer cycle, the numbers in the discussion below refer to the circled numbers in the timing diagram in Figure A.

### Let the Games Begin

When a bus master wants to initiate a transfer, it requests control of the bus by asserting the MREQx signal (1). It then monitors the MAKx line (2) until the EISA arbitration system grants its request (2').

Next, the bus master places the target address on the bus and indicates whether it is initiating a transfer with system memory or with an I/O slave (3). Then, it indicates whether it is reading from or writing to the indicated address, and which bytes of the 32-bit-wide data path it wants to transfer (4).

At the same time, the master drives the START signal to initiate the transfer cycle (5). At the end of the START signal, which lasts one BCLK cycle, the bus master negates the CMD line (6). Following the confirmation of START, the bus master samples EX32 and EX16 (7) to determine whether the slave is a 32-bit or a 16-bit device. Once it determines the type of slave (7'), it checks for EXRDY from the slave. When the slave is ready to terminate a transfer cycle, it drives the EXRDY signal

(8). When the master confirms EXRDY, it negates CMD to finish the cycle (9).

When writing, the data from the bus master is valid once it drives the START signal (10). When reading, the master samples the data lines on the first rising edge of BCLK after the slave asserts EXRDY (11).

If the slave does not reply immediately to the bus master, it will wait another BCLK cycle. In the example, EX16 or EX32 missed the call from the bus master, thus incurring two wait states (7), and the master missed EXRDY, adding another wait state (8'). The result is a three-wait-state cycle. A standard cycle with not wait states lasts only two BCLK cycles.

### Burst of Power

In a burst cycle, a burst-capable bus master uses all the signals defined for the standard bus master, in addition to driving the MSBURST signal and monitoring the SLBURST signal from the slave device. The numbers in the discussion below refer to the circled numbers in Figure B.

A burst cycle follows the same sequence as a standard cycle until after the master asserts START. At this point, the master begins sampling SLBURST,

as well as EX32 and EX16. As before, once the slave affirms EX32 or EX16, the master starts sampling EXRDY. However, if the slave is burst-capable, it asserts SLBURST (1) in addition to EX32 or EX16. When this happens, the master negates MSBURST (2) and begins the burst cycle.

As before, data from the master is valid when START is asserted (3). Data from the slave is valid (4) until CMD ends the cycle.

After the cycle begins, the bus master provides the next LA[2:31], M/IO, BE[0:3], and W/R when it samples EXRDY. It also samples MSBURST at every rising edge (5) unless EXRDY is not accessible (6). For the burst cycle, each wait state must have one BCLK cycle.

The master ends the cycle by disasserting MSBURST (7) and inverting CMD (8). As with the standard cycle, CMD is first negated at the end of the START signal. In a zero-wait-state system, a burst-mode cycle will take only one BCLK cycle with the exception of the initial cycle, which requires two BCLK cycles.

### The Bus Stops Here

When the specifications for implementing EISA were released, the bus standard attracted a lot of attention. The intent of EISA was to develop a 32-bit architecture that, unlike the Micro Channel,

would be downward-compatible with the existing AT bus. Systems using EISA would still be able to use 8-bit and 16-bit boards. The EISA standard

**Table A. Cycle Types and Times**

*EISA supports three types of bus transfers, of which only two—standard and burst—are available to bus masters. This table lists the average number of bus clock cycles it takes to complete a data transfer cycle. (N/A = not applicable.)*

Cycle	Speed	Standard Cycle	Compressed Cycle	Burst Cycle
EISA	32-bit	2	1.5	1
One-wait-state ISA	16-bit	3	N/A	N/A
	8-bit	6		
Zero-wait-state ISA	16-bit	2	N/A	N/A
	8-bit	3		

adds many functional and performance enhancements, such as improvements in DMA handling, but it retains a subset of the features available in the ISA bus.

EISA's data path can be 8, 16, or 32 bits wide. The number of address lines is increased to 32, which theoretically allows access to up to 4 gigabytes of memory, depending on the system design. This larger amount of addressable memory better meets the needs of today's memory-hungry applications, such as multiuser systems, network-server systems, AI CAD/CAM simulations, and even computer hardware design.

Not only can the CPU fully implement EISA's entire 32-bit address space, but so can the 32-bit bus-master and DMA devices on the EISA bus. EISA's 32-bit DMA functions allow today's 16-bit ISA DMA products to use the 32-bit address space by adding a simple software device driver. In other words, you can keep your system's old equipment and improve its performance.

The EISA bus master includes a dedicated I/O processor and local memory. This specialized processor drives the address, data, and control signals for intelligent peripherals, which become slave devices, during a bus cycle. Bus masters improve system performance by taking on simple tasks that would otherwise fall to the host processor. Thus, they reduce the main processor's work load (the beginning of multiprocessing, perhaps?). To understand the benefits of the bus master, it is important to grasp the mechanics of transferring data between the bus masters and their slave devices (see the sidebar "Requesting a Transfer").

The capabilities of the EISA bus master are aimed at sophisticated I/O peripherals that require

**Table B. EISA Bus-Mastering Signals**

*The EISA bus features special signals that enable a master to take control of the bus. If both the master and slave are burst-capable, the signals also enable burst transfers.*

**Controlled by a Master**

MREQx	Asserted by a bus master when requesting use of the bus. The x is the master's slot number.
MAKx	Indicates that the master in slot x has control of the bus.
LA[2:31]	Address lines.
M/IO	Indicates whether the transfer involves main memory or a slave device.
BE[0:3]	Indicates which bytes of the 32-bit-wide bus are used in the current bus cycle.
W/R	Indicates whether the master is writing or reading data.
START	Indicates the start of a bus-mastering cycle.
CMD	Asserted by a master during a transfer cycle.
MSBURST	Indicates that a master is capable of controlling burst transfers.
<b>Monitored by a Master</b>	
D[0:31]	Data lines.
EX32	Indicates that the slave handles 32-bit transfers.
EX16	Indicates that the slave handles 16-bit transfers.
EXRDY	Indicates that the slave has completed its function and is ready to terminate the cycle.
SLBURST	Indicates that a slave is capable of burst transfers.

optimum performance or advanced memory-access functions (e.g., non-ordered scatter/gather data operations). As a result, EISA is primarily intended for 32-bit devices, which typically contain dedicated I/O processors and require the fastest data transfer rate available through the bus. For this purpose, a fast burst-transfer mode is available to bus masters, in addition to the more typical timing of I/O and memory cycles.

**Table 1. DMA Cycle Types**

*Most ISA-compatible DMA devices can transfer data faster if the EISA controller is set up to use Type A and Type B transfers instead of ISA-compatible timing. Enhanced arbitration shortens the time between the DMA device's request and grant events.*

DMA Cycle Type	Size of Transfer	Transfer Rate (MBps)	Compatibility
Compatible	8-bit	1.0	All ISA
	16-bit	2.0	All ISA
Type A	8-bit	1.3	Mostly ISA
	16-bit	2.6	Mostly ISA
	32-bit	5.3	EISA only
Type B	8-bit	2.0	Some ISA
	16-bit	4.0	Some ISA
	32-bit	8.0	EISA only
Burst DMA (Type C)	8-bit	8.2	EISA only
	16-bit	16.5	EISA only
	32-bit	33.0	EISA only

In one scenario, a system could have several bus-master peripherals running on a network file server, each dedicated to a special application, such as distributed database processing. EISA can support up to 15 bus masters with an arbitration method that determines the latency of each device. Using this latency, the response time for requests from expansion-bus devices can be determined.

### Knowing Your ABCs

In EISA systems, DMA devices have seven channels, just like they do in ISA systems, but the EISA transfer rate is much faster. The DMA controllers support 8-, 16-, and 32-bit data transfer sizes. They have four cycle-control sequences for transferring data between the DMA device and memory. These four cycles are

- the ISA-compatible cycles, which execute one transfer cycle in eight bus cycles;
- Type A cycles, which execute one transfer cycle in six bus cycles;
- Type B cycles, which execute one transfer cycle in four bus cycles; and
- Type C (burst DMA) cycles, which execute one transfer cycle in one bus cycle.

Types A, B, and C support 8-, 16-, and 32-bit DMA devices and perform automatic data-size translation for transfers to mismatched memory.

Table 1 indicates peak data transfer rates and compatible DMA devices for each DMA cycle type. Moreover, most ISA-compatible DMA devices can transfer data 130 percent to 200 percent faster by programming the EISA controller to use Type A and B transfers instead of ISA-compatible timing.

How is this increased performance and efficiency provided? By enhanced arbitration, which shortens the time between the DMA device's request and grant events. This enhancement does not imply a decrease in compatibility. Existing hardware and software can take advantage of it without modification, so it actually improves compatibility with older systems.

### Fringe Benefits

The original PC and ISA buses used edge-triggered interrupts, which are easy to implement but susceptible to false triggering and cannot be shared with other peripherals. In addition to supporting these edge-triggered interrupts to maintain compatibility, EISA also provides level-triggered interrupts, which are less susceptible to noise and allow multiple peripherals to share the same interrupt level. Theoretically, level-triggered interrupts can have an infinite number of levels.

Another practical benefit that EISA provides is automatic configuration of system resources and expansion boards. That means an end to DIP switches, jumpers, and installing configuration files. A system that supports plug-and-play peripherals improves efficiency (switch configurations for ISA or EISA products are still allowed, if you need them).

EISA also includes a product-identification mechanism for systems and expansion-board products. As you turn the system on, the computer automatically runs a power-up sequence of self-tests. When it does, it interrogates each device for the product identifier, compares that with those stored in RAM, and configures the board according to the setup data stored in ROM.

### Micro Channel vs. EISA

The Micro Channel's design is similar to that of EISA, and it includes support for the 386 and i486 microprocessors. It also supports up to 15 bus masters and can transfer data in burst mode. The Micro Channel also replaces DIP switches and

jumpers with a self-configuring system. Unlike the ISA bus, the Micro Channel's level-triggered interrupts and sophisticated DMA-arbitration scheme allow multimaster operation. The major disadvantage of the Micro Channel, however, is that it is not downward-compatible with the ISA bus.

Supporters of the Micro Channel architecture argue that you won't want to migrate your original expansion cards to new machines. However, even if you don't, there are tens of thousands of application programs and thousands of expansion boards on the market today that EISA supports that the Micro Channel does not. In addition, most Micro Channel-based systems have basic features built into the machine, such as graphics adapters, mouse/serial/parallel ports, and floppy and hard disk controllers. These built-in features limit your future choice of products.

Micro Channel systems transfer data between the CPU, memory, and 32-bit peripherals at a maximum rate of 20 megabytes per second. That's less than two thirds the speed of an EISA system, which normally performs at about 33 MBps. The Micro Channel offers a maximum of eight expansion slots, compared to EISA's 15.

Physically, the form factor of a Micro Channel board is about half the size of an EISA board. This makes Micro Channel product design more difficult and expensive. Since the board is smaller, the circuitry must also be smaller. In some cases, this means expensive application specific ICs or surface-mounted ICs are necessary.

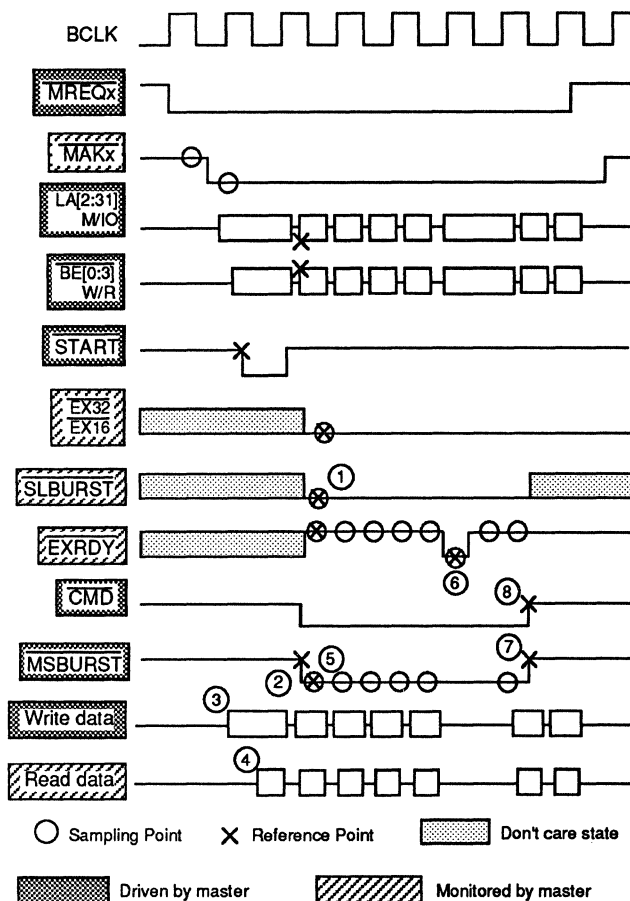
In addition, the Micro Channel's card has less than half the power of an EISA card available. This makes peripherals, like I/O boards with large amounts of memory, much more complex and expensive to design and build for Micro Channel systems.

## EISA Takes Its Stand

Intel has already introduced EISA chip sets, including three chips for the motherboard and one chip for an add-in card. The first motherboard chip has been designated the integrated system peripheral and provides 32-bit DMA control, timer/ counter control, interrupt control, bus arbitration, and DRAM refresh functions on a single chip.

The EISA bus controller, the second chip in the set, interfaces with the 8- and 16-bit ISA bus, the 32-bit EISA bus, and the CPU by performing

Figure B.  
EISA Burst Mode



Whereas standard mode requires that the master and slave synchronize for every 16- or 32-bit transfer, burst mode requires that they synchronize just once, before the first transfer. The average cycles per transfer in burst mode is thus very close to one.

the timing and control functions. The last motherboard chip, which is called the EISA bus buffer, contains buffering logic for any one of the data, address, and parity-control modes.

The intelligent add-on cards chip, which is called the bus-mastering interface controller, serves as the interface between a plug-in card and the 32-bit EISA bus and provides interface logic, drivers, and a DMA controller that is capable of becoming a bus master. EISA chip sets work with the 386 and i486 CPUs at speeds of up to 33 MBps.

The members of the original Gang of Nine are starting to introduce EISA systems, but not all

of them are ready yet. Only a few currently have EISA systems. Hewlett-Packard was the first to produce an EISA microcomputer system, the Vectra 486PC. It transfers data at a rate of 20 MBps with a 16-bit ISA ESDI controller.

Right before Fall Comdex 1989, Compaq revealed its EISA system, the Systempro. The Systempro can decrease the time required to transfer a file to or from the disk by distributing a single file among multiple bus-master controllers, called an Intelligent Drive Array (IDA). Compaq also claims that its system will transfer data at 33 MBps.

Zenith Data Systems' 386/33E is the latest EISA PC system. The bus-master controller, which uses 1 MB of memory and is expandable to 4 MB, will support up to four ESDI devices, seven SCSI devices, and two floppy disk drives. The Zenith 386/33E is also supposed to be able to support a data transfer rate of about 33 MBps.

Some companies outside the original Gang of Nine are also working on EISA products. By the end of June, Arche Technologies will expand its current 486 AT system (up to 64 MB of memory, 256K bytes of external cache, and enhancements for write posting, page memory cycle, and prefetch cycles) to support the EISA standard. The Arche 486 EISA system will be able to use 32-bit VGA, 32-bit Ethernet, and 32-bit SCSI host adapters.

Although EISA systems are downward-compatible with ISA expansion boards, a number of EISA-specific products are expected to appear by the end of this year. These include

- Adaptec's 32-bit SCSI controller,
- Distributed Processing Technology's 32-bit hard disk controller,
- Proteon's 32-bit EISA version of a token-ring adapter,
- Standard Microsystem's 32-bit EISA version of an Ethernet board,
- SunRiver's 32-bit fiber optic board,
- 3Com's network card,
- Arcnet's 16-bit multiuser board, and
- Western Digital's 16-bit ESDI controller.

EISA systems will thrive on the availability of EISA-specific products and the already existing stock of ISA products from current PC machines.

---

### **Make New Friends but Keep the Old**

EISA will save you money, time, and paperwork because it will allow you to make better use of the hardware and software resources you already have. Like its predecessor, ISA, EISA is an evolutionary approach to the need to improve system performance. It offloads I/O processing from the microprocessor's list of things to do.

Unlike the Micro Channel, however, EISA does not require new peripheral cards to support its I/O bus. It uses standard AT cards. While adapting to the current AT bus architecture and using the more powerful 32-bit processors, EISA preserves compatibility with the existing world of AT-compatible devices. EISA has it all. ■



# Overview of Data Communications Standards

## In this report:

Trends and Issues .....	4
Standards Organizations .....	5

## Datapro Summary

Standards are important in data communications. Without them, two devices that must share information may not have any protocols in common. The goal of standards is not necessarily to stamp out proprietary specifications for protocols or interfaces, but to ensure that there is a set of specifications, with sufficient support, to ensure free communications between two devices, regardless of their type or vendor. This report provides information on the principal data processing standards and on the major standards organizations.

Data communications is the exchange of computer-coded data over a system-to-system transmission medium. Such communications require the parties to agree on the format of the information to be exchanged and any special control strategies to be used to establish a connection, detect and correct errors, and regulate flow. Such an agreement is called a "protocol." Similar agreements are required for physical connections between devices, governing the electrical connector types, voltages, etc. These specifications are called "communications standards."

## An Overview of Standards: The OSI Model

There are many different types of transmission media used for data communications from cable, intended to span distances measured in feet, to satellite paths capable of spanning the globe. There are also many different types of computing devices, ranging from supercomputers to simple ASCII terminals, and many applications that join them.

This diversity creates a broad range of communications features that a user-to-user connection employs. What supercomputers running on local cable might need to support high-resolution, direct visualization displays of atomic structure is very different from what terminals might need to access a worldwide bulletin board service for casual electronic "chats" between users.

To facilitate the development of specifications that match media, requirements, and devices properly, designers of protocols developed a model of how protocol functions could be separated into "layers," each having a specific set of functions, providing successive layers a specific set of services. Layers could be matched to related requirements, independently of the way other layers were selected, and still support communications. This concept was formalized in the late 1970s by the International Organization for Standardization as the Open Systems Interconnection Basic Reference Model, or the OSI Model.

Figure 1 shows the structure of the OSI model and the names given to each of the OSI layers. Literature on the protocols reference layers either by these names, or by the layer numbers shown, with Level 1 being the lowest layer and Level 7 the highest.

The OSI layers are divided into two groups:

- *Local procedures* (Levels 1, 2, and 3). Standards for these layers, also called

—By Thomas Nolle  
President,  
CIMI Corp

“low level” standards, govern the communications across any communications line in a network and include the physical interface (Physical Layer, Level 1), error detection and correction (Data Link Layer, Level 2), and routing (Network Layer, Level 3). Functionality for these layers is included in each user system and each “routing node” that makes up a network. Data networks are implemented using protocols of these layers.

- *End-to-end procedures* (Levels 4 through 7). These layers provide the means of linking communicating applications to networks. They reside only in devices that serve users, supporting user-to-user connections.

The OSI model is not a protocol in itself. International standards bodies have defined specific protocols at each of the OSI layers, and vendors have also structured their proprietary protocols to match the structure of the model. There are, therefore, many choices, with various degrees of standardization, at each of the OSI layers. A set of protocols designed to support a specific range of applications is called a “protocol suite.” IBM’s Systems Network Architecture (SNA) is an example of a protocol suite made up of proprietary protocols. The Government OSI Profile (GOSIP) is a protocol suite made up of international standard protocols.

Table 1 provides a listing of the principal data processing standards. A list of standards organizations and their areas of attention can be found at the end of this report.

### Lower Level Standards (Levels 1 through 3)

The lower level standards of the OSI model are familiar to users because they define the specifications for physical attachment of devices and the protocols used for interfacing to network services.

OSI Level 1, the Physical Layer, includes standards that define the electrical interface specifications to be met in attaching devices to networks. These standards will specify a connector type, the meaning of pins or leads on the connector, the voltage levels used, etc. The most common sources of interface standards are the Electronic Industries

Association (EIA), which published the popular RS-232 and RS-449 specifications, and the CCITT, the source of V.24 and V.35.

A recent development that may fall between Levels 1 and 2 of the OSI model is frame relay. The ANSI Frame Relay specification (T16acDSS1) defines a protocol to interface with a frame relay service network. The protocol is expected to “encapsulate” a normal Level 2 protocol so that error recovery in a frame relay network would be across the entire network, making the network appear to the user as a single circuit.

There are no pervasive proprietary standards at Level 1. The fact that this level specifies physical interfaces between devices makes it risky for a vendor to promote such a standard. Support for it among other vendors might be limited, and this would make it impossible for users of a device supporting such an interface to connect with anything at all. This conformance to formal standards is lacking at Level 2, the Data Link layer of the OSI model. Level 2 specifies the protocol to be used in supporting error-free transmission across a single span of physical media. It includes procedures for establishing the identities of the parties on a connection, the exchange of data, retransmission of errors, and recovery from temporary loss of connection.

Early data link protocols like IBM’s Binary Synchronous Communications (BSC) were either based on “polling” or “contention.” Polled protocols define a master/slave relationship that assigns one station on the line responsibility for periodically asking others if they have traffic. “Contention” systems allow either station to “bid” the line. Virtually every computer vendor has a proprietary data link protocol.

A need for formal standardization of data link protocols created the ISO standard, High Level Data Link Control (HDLC). This protocol defines both a “balanced” and an “unbalanced” mode of operation, with the latter allowing one station to become the master. The unbalanced mode is based on IBM’s Synchronous Data Link Control (SDLC), and the balanced mode is the basis for the Link Access Procedure, Balanced (LAPB) used in the CCITT X.25 packet network standard. It is also the basis of the Link Access Procedure for the D channel, LAPD, used in the Integrated Services Digital Network (ISDN) and specified in CCITT Q.921.

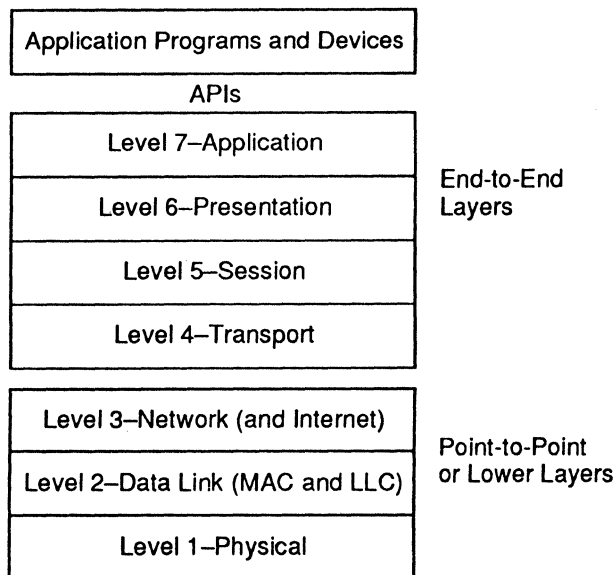
With the advent of local area networks, it became necessary to split OSI Level 2 into “sublayers.” The lower sublayer, called media access control (MAC), handled the interaction among stations associated with gaining permission to send data and resolving any conflicts or contention encountered in the process.

There is a MAC standard defined by the Institute of Electrical and Electronic Engineers (IEEE) for both the popular LAN access control strategies. Carrier Sense Multiple Access with Collision Detection (CSMA/CD), used in Ethernet LANs, is defined by IEEE 802.3. Token Passing, used in IBM’s Token-Ring LAN, is defined by IEEE 802.5.

The higher LAN Level 2 sublayer is logical link control, or LLC. LLC provides the more traditional Level 2 services of error detection and correction. But because many LANs were implemented without explicit Level 3 functions (for reasons discussed later), the LLC sublayer also provides a very limited level of support for datagram and virtual connection services, normally Level 3 functions. The IEEE standard for LLC is IEEE 802.2.

These IEEE LAN standards have also been adopted by the ISO, where the equivalent number can be developed by adding an “8” to the front, and replacing the decimal with

Figure 1.  
The OSI Model





a dash. Thus, IEEE 802.2 is ISO 8802-2. The Fiber Distributed Data Interface (FDDI) has not been standardized by the IEEE; its ISO standard is ISO 9314.

Level 3 is the Network Layer of the OSI model and is the layer where information routing takes place. Level 3 also provides some data segmentation control, limiting the size of messages that can be sent at Level 2.

Level 3 offers two major service types to the higher layers:

- *Connectionless services*—each element of information is sent without regard for its context with respect to other elements. Consecutive messages sent this way can be delivered out of order, because they take different routes through a network; some may not be delivered at all. This type of service is often called dataion can be detected by the network. Such a route appears to the communications user as a “circuit”; the service is also referred to as a virtual circuit service.
- *Connection-oriented services*—a specific route is established for information flowing between parties, and the data always follows that route. Because of this, information cannot pass prior messages, and lost information can be detected by the network. Such a route appears to the communications user as a “circuit”; the service.

After the initial specification of the OSI model, standards bodies realized that special consideration at level 3 would be required if a “network” were to be formed by linking together a series of independent “sublayer” of the model, called the Internet layer. Internet protocols allow information to be routed through “gateways” between networks and over one network on the way to another.

IBM’s Path Control in SNA was the first commercially used protocol that could be called a Network Layer protocol. The “IP” portion of TCP/IP is also a Network Layer protocol. The first international standard protocol at Level 3 was X.25’s Packet Level Protocol (PLP). The International Organization for Standardization has since added an ISO Connectionless Protocol (ISO 8473), which some vendors (notably AT&T and NCR) have made optional in their LAN products.

A family of protocols related to the Internet protocol sublayer are the “routing control protocols.” These are designed to enable internetwork gateways (generally routers) to communicate with one another in order to learn network structure and to coordinate routing. TCP/IP’s Open Shortest Path First (OSPF) procedure is such a protocol, as is the ISO End System to Intermediate System (ESIS) Routing Information Exchange Protocol and the Intermediate System to Intermediate System (ISIS) Routing Information Exchange Protocol.

Another special class of standards at the lower OSI layers supports intelligent network interfaces used in establishing connections, but not necessarily for the transfer of data. CCITT X.21 is the first example of this type of standard. Specified as the Physical Layer of X.25, X.21 is rarely used in the U.S., although it is more common abroad. X.21 uses character-coded dialing and message formats to replace many of the control signals used on earlier interfaces, simplifying the connector.

The most advanced intelligent network interface standard currently in use is the call control interface of the Integrated Services Digital Network (ISDN). The ISDN

signalling, or “D,” channel uses a Level 2 procedure called LAPD, whose relationship to HDLC has already been noted. For Level 3, ISDN specifies both a special call control protocol, defined in standard Q.931, and X.25’s PLP. This allows users to set up calls on the ISDN “B,” or user, channels or transmit packet data on the D channel. Given the low costs of VLSI and microprocessors, future network interfaces are likely to be intelligent as well, and large connectors with many pins will be things of the past.

### End-to-End Protocol Standards (Levels 4 through 7)

International standardization of protocols at Level 3 and below is facilitated by the fact that, according to the OSI model, public data network interfaces would take place at Level 3. The CCITT, an organization of carriers, was therefore instrumental in promoting standards at the lower levels of the model. Above Level 3, all protocol exchanges are end to end across the network and appear to the network elements as user data. This may explain why these protocols have lagged the lower layers in definition.

Proprietary protocols, such as IBM’s SNA, generally conform to the OSI model structure in the lower layers, though the protocols at each layer may not be standardized. At Level 4 and above, there is much less conformance because all of the handling of the protocols is presumed to take place within the vendors’ own systems.

Level 4 of the OSI model is the Transport Layer, which provides for reliable end-to-end connection. In a complex data network with many paths and many nodes, a message might be sent by the user to a node, received correctly, and acknowledged—only to be lost later within the network. Level 4 is responsible for recovering from such failures. Level 4 of SNA is Transmission Control, but this SNA layer also performs many of the functions mandated for the OSI Session Layer (Level 5). The Transmission Control Protocol (TCP) of ARPANET is a Level 4 protocol implemented on virtually every UNIX-based system.

There is only one formal standard for Level 4 defined by the ISO, but it has five classifications. The OSI Transport Protocol (TP) Class 0 is designed for use where the network on which it is operating is essentially error free. This is most likely to be a valid assumption on LANs with a robust data link and network/internet protocol. Class 1 provides basic end-to-end recovery from internal network failures but assumes error-free data delivery. Classes 2 and 3 are similar to Classes 0 and 1, but they support multiple conversations over a single network connection (multiplexing). Class 4 is the most robust of all transport protocols and is suitable for use over a network whose information delivery is not reliable. The CCITT has adopted the same transport standard as X.224.

The Transport Layer of the model also provides end-to-end flow control, pacing the information so that one device cannot overrun the other and so that the network is not burdened with “holding” data to match device speeds.

Level 5, the Session Layer of the OSI model, is responsible for creating application-to-application connections (sessions) that support a period of information exchange. Session services also allow for the synchronization of processes across a network, so that “checkpoints” can be created to permit recovery of application integrity in the event of a network failure.

SNA functions at Level 5 and higher are provided by a single logical layer called function management, though IBM does describe a "Data Flow Control" layer as its Level 5. The popular ARPANET TCP/IP environment has no session protocol at all. ISO standards at this level have been developed for both Connectionless (ISO 9548) and Connection Oriented (ISO 8326) protocols. The connection mode standard was also adopted as CCITT X.225.

Level 6 of the OSI model is the Presentation Layer. The purpose of this layer is to accommodate differences in the data representations of the systems communicating: code set, binary number structure, etc. There are both a Connectionless (ISO 9576) and Connection Oriented (ISO 8823/CCITT X.226) standard here. The Presentation Layer tends to be "null" in many real communications protocol suites. SNA defines Network Addressable Unit services at this layer and does provide specific formatted datastreams for individual devices, but it does not provide translation of formats, so no real presentation function is offered. Most computer systems today conform to the ASCII or EBCDIC standards for character coding, and most use compatible binary number representations.

CCITT X.400, the electronic mail standard, was formerly based on a specialized presentation protocol, X.409. This has been displaced by the ISO presentation protocol standard. Level 7 of the OSI model, the Application Layer, is where the greatest current development activity is focused. Level 7 defines network services that may be accessed by users and provides protocol specifications to control that access. The term "services," as used here, is very broad, ranging from the simple, reliable transfer of information (ISO 9066/CCITT X.218) to the complex directory services (X.500), electronic mail (X.400), and Common Management Information Service and Protocol (CMIS and CMIP, ISO 9595/CCITT X.710 and ISO 9596/CCITT X.711, respectively).

SNA Network Services functions are IBM's equivalent to Level 7; SNA does not have the layer explicitly. The most common example of these services is the IBM LU6.2 Advanced Program-to-Program Communications (APPC) service, a protocol similar to several ISO standards (Reliable Transfer, Commitment, Concurrency, Recovery, and Online Transaction Processing). Another example is SNA Distribution Services (SNADS), the service used by IBM's office system products to deliver documents.

TCP/IP networks and OSI networks have two service classes in common. Both provide for file transfer between systems: TCP/IP through FTP and OSI through FTAM (ISO 8571). Electronic mail is also supported in both: TCP/IP's Simple Mail Transfer Protocol (SMTP) and ISO and CCITT's X.400. The TCP/IP Simple Network Management Protocol (SNMP) and OSI's Common Management Information Protocol (CMIP) are competitive standards in the current period of integrated management expansion in the market.

There are already over 20 Level 7 standards completed or in review by the ISO, and several new ones are added each year. Unlike the lower levels of the model, in which each end system is expected to implement in an "open systems interconnection" environment, users would implement Level 7 standards only if they required the services.

### Application Programming Interfaces

Standards for data communications are a confusing issue at best, and they have recently become more so with the

definition of so-called application architectures. Application architectures provide the specifications for application development so that portability of code across multiple systems platforms is possible. IBM's Systems Application Architecture (SAA) is the best-known example.

Application architectures include the specifications for the connection between an application program running on a computer and a remote user or computer system via a network. This type of interface is often called an Application Programming Interface (API). While network users often recognize computers and devices by their protocols, computer users recognize them by their APIs.

The most popular APIs are the High Level Language Application Programming Interface (HLLAPI) used by IBM 3270 emulation applications on PCs, Berkeley UNIX "sockets," UNIX System V's "streams" and "Transport Layer Interface" (TLI), and IBM's Common Programming Interface for Communications (SAA CPIC).

There are some efforts to standardize APIs in the same way as protocols are standardized. The IEEE POSIX committee is considering standards for communications interfaces to applications, as is the international UNIX consortium X/Open. If such standards are developed quickly and accepted, they would simplify communications application development.

### Trends and Issues

The OSI standards for data communications have been an industry success story. Since their introduction, vendors have curtailed development of their own proprietary standards and have gradually improved their support of standard protocols.

This is not to say that standards are displacing proprietary protocols completely. Standards face three major problems:

1. The rate of development is slow compared to the pace of the marketplace. When a standard is needed and not available, an ad hoc vendor proprietary solution is likely to arise. Once this happens, it is difficult for the standard to displace it.
2. Generality of standards leads to inefficiency, both in terms of execution speed and memory requirements. A full OSI network management implementation would require nearly as much memory as a personal computer has available, leaving little or nothing to run programs.
3. Conformance to standards is hard to prove, since there are no domestic organizations with a statutory right to enforce them. Most testing in the U.S. is done by trade groups, private corporations, or carriers.

The problem of the pace of standards has been reduced significantly in recent years by the use of informal consortiums as a starting point for rapidly evolving concepts to take form. The OSI/Network Management Forum is an example of such a group. Once the vendors involved have agreed on a structure, they sponsor it jointly to the appropriate standards body, helping ensure its acceptance.

Where older de facto standards are already in place, the acceptance of newer standards is being encouraged through the use of transitional strategies. There are already "gateways" and conversion programs that allow TCP/IP's FTP and ISO FTAM to interoperate, an example of a Level 7 conversion. Application Layer standards from OSI

are also being overlaid on non-OSI protocols where such protocols are already implemented on computer systems and cannot be easily displaced. OSI's Common Management Information Protocol (CMIP), running over TCP/IP, is sometimes called "CMOT" for "CMIP Over TCP." The problem of efficiency is more difficult to solve, but one path to a solution may be the use of "short stacks," or protocol suites, that implement higher layer functions without full support of the intermediate layers.

IBM and 3Com jointly developed a management protocol called "CMOL," which stands for CMIP over LLC, and allows the ISO CMIP protocol to be used over LANs by sending messages directly via the LLC sublayer of Level 2. By eliminating the intermediary layers, the memory requirements for implementing CMIP on PCs are reduced enough to make the concept practical.

Conformance testing of standards is a problem not only because of the potential for false claims, but also because of the multiplicity of standards that may exist at each OSI level. The first logical step in testing is the establishment of logical sets of standards for a given application, the "protocol suite." The Manufacturing Automation Protocol (MAP) and Technical and Office Protocol (TOP) are examples of suites developed by a trade group (the MAP/TOP Users Group). Another suite, the Government OSI Profile (GOSIP), was developed by the federal government.

Once suites are defined, organizations can be mandated to test them. Low-level standards for network interfacing can be conformance tested by the carriers. Full-layer suites can be tested by user groups, trade groups, or independent corporations, such as the Corporation for Open Systems (COS).

The greatest problem to be faced by "standard" protocols is not a subject of protocol standards at all; it is the need for a common API. Without a consistent way to access OSI services over a wide variety of computers and operating systems, OSI services cannot be easily incorporated into application programs developed by third-party suppliers. This would limit OSI usage to companies whose programming organizations were skilled enough to develop communications applications.

In fact, an OSI-standard API would not be a full solution. The use of standards-based protocols would be best promoted by the integration of these protocols into vendor application architectures, such as IBM's SAA or Digital's NAS, as alternatives to the vendor's own proprietary network architectures. If this were done, standard protocols could be used wherever justified without affecting the applications themselves.

Such a situation cannot be expected in the near future, but with the growth in popularity of UNIX, portability of programs from system to system is gaining in interest. Vendors that continue to offer proprietary operating systems may find it necessary to adhere to UNIX API standards. (POSIX is a generic definition that both UNIX-based and other operating system software-based products could adopt.) They will almost certainly be motivated to support standard protocols more fully within their own application architectures, preserving developer interests.

## Standards Organizations

The major domestic and international standards organizations are listed below. Table 1 provides a cross-reference of data communications standards developed by seven of the listed organizations.

## American National Standards Institute (ANSI)

1430 Broadway  
New York, NY 10018  
(212) 354-3300

ANSI is the principal standards-forming body in the United States. Originally formed in 1918, it is a nonprofit, nongovernmental organization. It is the USA's representative to the International Organization for Standardization (ISO). ANSI standards result either from the work of its 300 Standards Committees or from associated groups, such as the EIA.

Technical Committees and Task Groups form the lower level of the ANSI hierarchy. They consist of technically qualified individuals rather than organizational representatives, and membership is open to anyone. The next level, the Standards Committees, consist of three categories of members: consumers, producers, and general interest. None of the member categories have a majority of the total vote, providing a balance of influence and ensuring a consensus.

In 1960, Standards Committee X3 was established to investigate all standards related to the computer industry; the Computer and Business Equipment Manufacturers Association (CBEMA) sponsors this committee. X3 is made up of 25 Technical Committees, each dealing with an assigned technical area. The Data Communications Technical Committee X3S3, established in 1961, has seven Task Groups: X3S31-Planning, X3S32-Glossary, X3S33-Transmission Formats, X3S34-Control Procedures, X3S35-System Performance, X3S36-Signalling Speeds, and X3S37-Public Data Networks.

## Electronic Industries Association (EIA)

2001 I Street NW  
Washington, DC 20006  
(202) 457-4966

The EIA is a trade organization representing a large number of U.S. electronics manufacturers. It was founded in 1924 as the Radio Manufacturers Association. Through the efforts of over 4,000 government and industry representatives in over 200 Technical Committees, the EIA Engineering Department has produced over 400 standards and publications.

Technical Committee TR-30, Data Transmission, established in 1962, deals with data communications. TR-30 is divided into three subcommittees: TR-30.1, Signal Quality; TR-30.2, Digital Interfaces; and TR-30.3, Telecommunications Network Interfaces. TR-30 develops and maintains DTE/DCE interface standards, working in liaison with both CCITT Study Group XVII and the ANSI Technical Committee X3S3. The EIA's work is hardware oriented, while the ANSI committee's work is more procedure oriented. TR-30 developed the RS-232 interface standard and the RS-449 interface standard.

## European Computer Manufacturers Association (ECMA)

114, rue de Rhone  
CH-1204 Geneva, Switzerland  
41 22 35-36-34  
Telex 222 88

ECMA, formed in 1961, develops data processing standards. It is not a trade organization. There are two classes

of membership within ECMA: 14 Ordinary Members, European companies that develop, manufacture, and market data processing equipment; and Association Members, whose ranks include organizations with interests and experience in the European area in matters of concern to the Technical Committees.

ECMA Technical Committees (TCs) develop standards; 28 TCs have been formed; 17 are currently active. These include TC 9, Data Communications, which, although currently inactive, has in the past worked closely with CCITT Study Groups VII and XVII; TC 23, Open Systems Interconnection; TC 24, Communications Protocols, the first group to define the OSI Transport Layer Protocol (ECMA-72); and TC 25, Data Networks. TC 23, TC 24, and TC 25 work in close cooperation with the CCITT Study Groups and the ISO Subcommittees.

### Federal Information Processing Standards (FIPS)

U.S. Department of Commerce  
National Technical Information Service  
5285 Port Royal Road  
Springfield, VA 22161  
(703) 487-4650: ordering  
(703) 487-4600: general information

FIPS is the identifier applied to standards developed under the federal government's computer standardization program. The FIPS program is the result of Public Law 89-306, which calls for the Secretary of Commerce to make recommendations to the President concerning uniform data processing standards. The National Institute of Standards and Technology (NIST)—formerly the National Bureau of Standards (NBS)—drafts FIPS specifications.

Over 80 FIPS have been adopted. The NIST works closely with the Federal Telecommunications Standards Committee (FTSC) to develop FIPS data communications standards. The thrust is toward commonality with other standards whenever possible. The NIST, therefore, maintains close cooperation with national and international standards activities.

### Federal Telecommunications Standards Committee (FTSC)

General Service Administration  
Specification Distribution Branch  
Building 197 (Washington Navy Yard)  
Washington, DC 20407

The FTSC is a federal government interagency advisory body established in 1973. Its objectives are to achieve interoperability among functionally similar telecommunications networks, to work with the NIST to establish data communications interface standards, and to ensure the federal government's participation in programs for national and international standardization.

The FTSC works closely with the CCITT, ANSI, ISO, and EIA and adapts or applies the standards developed by these organizations for federal use. In general, the FTSC avoids developing its own standards unless a clear need exists. Three areas where the FTSC has been active, however, are the development of standards for modems (compatible with CCITT Recommendations), defining system performance evaluation procedures, and developing standards for implementing the Data Encryption Standard (DES) algorithm.

### Institute of Electrical and Electronics Engineers (IEEE)

IEEE Computer Society  
111 19th Street NW, Suite 608  
Washington, DC 20036  
(202) 785-0017

The IEEE, a U.S.-based organization, was established in 1884. It actively establishes standards for the data communications industry. Its best-known effort is Project 802, which attempted to define local area network (LAN) standards. The market, however, accepted local area networks faster than the IEEE's standards efforts could define the possible configurations. IEEE 802, first begun in early 1980, has received an unusually high degree of support; over 125 companies and universities are actively involved. The IEEE 802 recommendation consists of a set of standards that deals with the Physical and Data Link Layers of the ISO Reference Model for Open Systems Interconnection (OSI). The six sections of 802 are 802.1, which describes the relationship among the 802 standards and their relationship to the OSI Reference Model; 802.2, which specifies the functions and features of the logical link protocol; 802.3, which describes a bus topology using CSMA/CD as the access method; 802.4, which describes a bus topology using token passing as the access method; 802.5, which describes a ring topology using token passing as the access method; and 802.6, which describes a metropolitan area network standard using CATV or other media. The IEEE 802 committee's charter is to seek a LAN model and to recommend interface and protocol specifications for logical link control, access methods, encoding techniques, and physical media.

### International Organization for Standardization (ISO)

Central Secretariat  
1, rue de Varembe  
CH-1211 Geneva, Switzerland  
41 22 34-12-40

Copies of ISO standards can be ordered from:

American National Standards Institute  
1430 Broadway  
New York, NY 10018  
(212) 354-3471

The ISO, currently comprising 90 member nations, is a nontreaty organization founded in 1947. Each nation assigns its principal standardization body to the ISO; ANSI represents the U.S. Nations may be active contributors (Participating, or "P," members) or observers of the standardization process (Correspondent, or "O," members). Each of the "P" members manages one or more technical committees or subcommittees.

Technical Committee 97 (TC 97), established in 1961, handles computer-related issues. In 1981, the ISO expanded this committee's scope to include office equipment (previously the concern of TC 95, now dissolved). TC 97 consists of 18 subcommittees (SCs). SC 6 addresses data communications standards, and SC 16 addresses Open Systems Interconnection. SC 6, made up of 18 "P" and 12 "O" member countries, works in close cooperation with the CCITT. Formed in 1978, SC 16, which has 13 "P" and 8 "O" members, is responsible for the draft proposal

ISO 7498, the OSI Reference Model. Most of SC 6's interface and protocol work conforms to the requirements of ISO 7498.

### OSI Management Standards

The OSI Management Framework, DIS 7498-4, is an ISO standards document that establishes guidelines for coordinating the development of existing OSI Management standards. It serves as a reference document for other OSI Management standards. The OSI Systems Management concepts provide mechanisms for monitoring, controlling, and coordinating all managed objects with open systems.

OSI Management standards are documents written by OSI committees that define information structures, services, and protocols required for OSI Management. These standards define the software tools necessary to monitor, control, operate, and administer network components in OSI environments.

OSI Management standards are important because they offer the only realistic and workable method for enabling multivendor networks to exchange management information on a worldwide basis. As public and private networks become more closely integrated in the future, the exchange of administrative and operational data will become increasingly crucial. OSI Management standards will provide each network and network component with an identical set of tools to support information exchange. Without these standards, every network would require proprietary gateways to every other network in order to exchange management information. This is not acceptable on either a cost or an efficiency basis.

### International Telegraph and Telephone Consultative Committee (CCITT)

General Secretariat  
International Telecommunications Union  
Place de Nations  
1211 Geneva 20, Switzerland

For ordering copies of CCITT standards in the U.S., contact:

United States Department of Commerce  
National Technical Information Service  
5285 Port Royal Road  
Springfield, VA 22161  
(703) 487-4650

The Union Telegraphique was established in 1865 when nations realized the need for standards to create international telecommunications services. A treaty between

participating nations formed the original organization, which was renamed the International Telecommunications Union (ITU) in 1947.

Over 160 countries belong to the ITU. Participants are placed in one of five categories. Private or government-controlled organizations providing public telecommunications services are Recognized Private Operating Agencies (RPOAs). Member countries' various governmental telecommunications administrations comprise the Administration category. Scientific and Industrial Organizations encompass commercial organizations with an interest in solving telecommunications problems and manufacturers of telecommunications equipment. International Organizations consist of private and commercial organizations with an interest in international telecommunications. Recognized treaty organizations in a telecommunications-related field are considered Agencies.

Either of two groups within the ITU handles standards recommendations: the International Radio Consultative Committee (CCIR) or the International Telegraph and Telephone Consultative Committee (CCITT). The CCITT, which handles data communications standards, consists of 15 Study Groups, each with a specific responsibility. A few of the more significant groups are Study Groups VII and XVII—data communications interfaces, services, and transmission work; XVIII—digital networks; XI—telephony; and III—tariffs.

CCITT operations are divided into four-year blocks during which the Study Groups act upon a series of technical questions assigned by the Plenary Assembly of the CCITT. After each four-year period, each study group presents the results of its work to the Plenary Assembly for approval as recommended standards. Although not intended to be mandatory, the CCITT Recommendations have the effect of law in many European countries.

The group with responsibility for public data network standards, Study Group VII, was formed in 1972. Now called the Data Communications Networks Study Group, it is the most active committee in the CCITT. Study Group VII is now working on a set of electronic mail standards, called Series X.400. Series X.400 defines the network architecture, protocol, implementations, message transfer, and messaging services for the interconnection of electronic mail networks. All the X Series recommendations are the result of Study Group VII's work. Study Group XVII, formed in 1960, is responsible for the V Series of recommendations. These standards are achieving widespread implementation worldwide.

**Table 1. Data Communications Standards Cross-Reference Table**

Organization/Standard	Title/Description	Related Standard
CCITT—		
I.200	Series includes I.210, I.211, and I.212; describes ISDN's service capabilities (bearer services and teleservices). Bearer services provide information transfer between users at layers one through three. Teleservices provide full communication between users, including capabilities at layers four through seven.	—
I.300	Series contains five recommendations for ISDN's network aspects. Included in series are I.320, the ISDN Protocol Reference model; I.330 on ISDN numbering and addressing principles; and I.331 (also known as E.164) on the ISDN numbering plan and principles.	—
I.400	Series addresses layers one through three of the user-network interfaces and provides procedures for adapting existing non-ISDN terminals for operation on an ISDN. As specified in the ISDN principles, a limited set of standardized interfaces will integrate services. These standards allow for terminal portability; for example, from the home to the office or from one network to another.	—
I.410	General aspects and principles relating to Recommendations on ISDN user-network interfaces	—
I.411	ISDN user-network interfaces—reference configurations	—
I.412	ISDN user-network interfaces—interface structures and access capabilities	—
I.420	Basic user-network interface	—
I.421	Primary rate user-network interface	—
I.430	Basic user-network interface—layer 1 specifications	—
I.431	Primary rate user-network—layer 1 specifications	—
I.440	Also known as Q.920, ISDN user-network interface data link layer—general aspects	—
I.441	(Q.921) ISDN user-network interface data link layer specifications	—
I.450	(Q.930) ISDN user-network interface layer 3—general aspects	—
I.451	(Q931) ISDN user-network interface layer 3 specification	—
I.460	Multiplexing, rate adaption, and support of existing interfaces	—
I.461	(X.30) Support of X.21 and X.21 bis-based DTEs by an ISDN	—
I.462	(X.31) Support of packet-mode terminal equipment by an ISDN	—
I.463	(V.110) Support of DTEs with V Series-type interfaces by an ISDN	—
I.464	Multiplexing, rate adaption, and support of existing interfaces for restricted 64K bps transfer capability	—

**Table 1. Data Communications Standards Cross-Reference Table (Continued)**

Organization/Standard	Title/Description	Related Standard
I.500	Series includes recommendations for ISDN network interfaces	—
I.600	Series includes recommendations for ISDN maintenance	—
V.3	International Alphabet No. 5	ISO 646; ANSI X3.4; FIPS 1-1
V.4	General structure of signals of International Alphabet No. 5 code for data transmission over public telephone networks	CCITT X.4; ISO 1155, 1177; ANSI X3.15, X3.16; FED-STD 1010, 1011; FIPS 16-1, 17-1
V.5	Standardization of data signaling rates for synchronous data transmission in the general switched telephone network	ANSI X3.1; EIA-RS-269-B; FED-STD 1013; FIPS 22-1
V.6	Standardization of data signaling rates for synchronous data transmission of leased telephone-type circuits	ANSI X3.1; EIA RS-269-B; FED-STD 1013; FIPS 22-1
V.10	Electrical characteristics for unbalanced double-current interchange circuits for general use with integrated circuit equipment in the field of data communications	CCITT X.26; EIA RS-423-A; FED-STD 1030A
V.11	Electrical characteristics for balanced double-current interchange circuits for general use with integrated circuit equipment in the field of data communications	CCITT X.27; EIA RS-422-A; FED-STD 1020A
V.21	300 bps duplex modem standardized for use on the general switched network	—
V.22	1200 bps duplex modem standardized for use on the general switched telephone network and on leased circuits	FED-STD 1008
V.22 bis	2400 bps duplex modem using the frequency division technique standardized for use on the general switched telephone network and on point-to-point, two-wire leased, telephone-type circuits	FED-STD 1008
V.23	600/1200 bps modem standardized for use on the general switched network	—
V.24	List of definitions for interchange circuits between data terminal equipment and data circuit-terminating equipment	EIA RS-232-C, RS-449, RS-449.1, RS-266-A
V.25	Automatic calling and/or answering equipment on the general switched telephone network, including disabling of echo suppressors on manually established calls	EIA RS-366-A
V.26	2400 bps modem standardized for use on four-wire leased circuits	—
V.26 bis	2400/1200 bps modem standardized for use in the general switched telephone network	FED-STD 1005
V.27	4800 bps modem with manual equalizer standardized for use on leased telephone-type circuits	—
V.27 bis	4800/2400 bps modem with automatic equalizer standardized for use on leased telephone-type circuits	FED-STD 1006
V.27 ter	4800/2400 bps modem standardized for use in the generalized switched telephone network	FED-STD 1006
V.28	Electrical characteristics for unbalanced, double-current interchange circuits	EIA RS-232-C

**Table 1. Data Communications Standards Cross-Reference Table (Continued)**

Organization/Standard	Title/Description	Related Standard
V.29	9600 bps modem standardized for use on point-to-point, leased, telephone-type circuits	FED-STD 1007
V.31	Electrical characteristics for single-current interchange circuits controlled by contact closure	—
V.32	A family of two-wire duplex modems operating at data signaling rates of up to 9600 bps for use on the general switched telephone network and on leased telephone-type circuits	—
V.35	Data transmission at 48,000 bps using 60-180kHz group band circuits	—
V.36	Modems for synchronous data transmission using 60-180kHz group band circuits	—
V.42	Specifies data compression technique for modems using LAPM error-control protocol	—
V.54	Loop test devices for modems	EIA RS-449
X.1	International user classes of service in public data networks	ANSI X3.1, X3.36; EIA RS-269-B; FED-STD 1001, 1013; FIPS 22-1, 37
X.2	International user services and facilities in public data networks	INT-FED-STD 001041
X.3	Packet assembly/disassembly (PAD) facility in a public data network	—
X.10	Categories of access for data terminal equipment (DTE) to public data transmission services provided by PDNs and/or ISDN through on-terminal adapters	—
X.21	Interface between DTE and DCE equipment for synchronous operation on public data networks	—
X.21 bis	Use on public data networks of data terminal equipment (DTE) which is designed for interfacing to synchronous V Series modems	—
X.24	List of definitions for interchange circuits between DTE/DCE on public data networks	—
X.25	Interface between data terminal equipment (DTE) and data circuit-terminating equipment (DCE) for terminals operating in the packet mode and connected to public data networks by dedicated circuit	—
X.28	DTE/DCE interface for start-stop mode data terminal equipment accessing the PAD in a public data network situated in the same country	—
X.29	Procedures for the exchange of control information and user data between a PAD and a packet mode DTE or another PAD	—
X.121	International numbering plan for public data networks	—
X.200	Reference Model of Open Systems Interconnection for CCITT Applications	—
X.400	Message Handling Systems—System model—service elements	—



**Table 1. Data Communications Standards Cross-Reference Table (Continued)**

Organization/Standard	Title/Description	Related Standard
X.401	Message Handling Systems—Basic service elements and optional user facilities	—
X.408	Message Handling Systems—Encoded information-type conversion rules	—
X.409	Message Handling Systems—Presentation transfer syntax and notation	—
X.410	Message Handling Systems—Remote operations and reliable transfer server	—
X.411	Message Handling Systems—Message Transfer Layer	—
X.420	Message Handling Systems—Interpersonal messaging user agent layer	—
X.430	Message Handling Systems—Access protocol for teletex terminals	—
X.500	Directory Services—searches database for address of person or machine to match a given name	—
<b>ANSI—</b>		
X3.44	Determination of the performance of data communications systems	—
X3.57	Structure for formatting message headings for information exchange using ASCII	—
<b>IEEE—</b>		
IEEE 488	Parallel Interface	—
IEEE 802.2	Functions and features for use in a multi-station, multiaccess environment	—
IEEE 802.3	Broadband and baseband bus using carrier-sense multiple access with collision detection (CSMA/CD) and physical interface specifications	—
IEEE 802.4	Broadband and baseband bus using token passing as the access method, and physical interface specifications	—
IEEE 802.5	Token-passing ring access method using token passing as the access method	—
<b>ISO—</b>		
ISO 2593	Connector pin allocations for use with high-speed data terminal equipment	—
ISO 4903	Data communication—15-pin DTE/DCE interface connector and pin assignments	—
ISO 7498 (Draft)	Open Systems Interconnection—Basic Reference Model	—
<b>EIA—</b>		
RS-232-C	Interface between data terminal equipment and data communications equipment employing serial binary data interchange	—
EIA 232-D	Interface between data terminal equipment and data communications equipment employing serial binary data interchange	—

**Table 1. Data Communications Standards Cross-Reference Table (Continued)**

Organization/Standard	Title/Description	Related Standard
RS-363	Standard for specifying signal quality for transmitting and receiving data processing terminal equipment using serial data transmission at the interface with nonsynchronous data communications equipment	—
RS-366-A	Interface between data terminal equipment and automatic calling equipment for data communications	—
RS-404	Standard for start-stop signal quality between data terminal equipment and nonsynchronous data communications equipment	—
RS-410	Standard for the electrical characteristics of Class A closure interchange circuits	—
RS-422-A	Electrical characteristics of balanced voltage digital interface circuits	—
RS-423-A	Electrical characteristics of unbalanced voltage digital interface circuits	—
RS-449	General-purpose 37-position and 9-position interface for data terminal equipment and data circuit-terminating equipment employing serial binary data interchange	—
EIA-530	High-speed, 25-position interface for data terminal equipment and data circuit-terminating equipment employing serial binary data interchange; developed to serve as a complement to EIA-232-D for data rates above 20K bps	—
FTSC—		
FED-STD-1002	Time and frequency reference information in telecommunications systems	—
INTERIM FED STD-001003	Telecommunications: digital communications performance parameters	—
FED-STD-1037	Glossary of telecommunications terms	—
ECMA—		
ECMA-72	Transport protocol	—

This report was prepared exclusively for Datapro by Thomas Nolle, president, CIMI Corp. Located in Voorhees, NJ, CIMI Corp. is a technology firm that specializes in strategic planning and market development.

Tom's views on new communications products and services are regularly quoted in major trade publications. ■

# IEEE 802 Standards for Local Area Networking

## In this report:

Overview of the IEEE 802 Standards .....	2
IEEE 802.3 (CSMA/CD) .....	4
IEEE 802.4 (Token Bus) .....	7
IEEE 802.5 (Token-Ring) .....	8
Local Area Network Interconnection .....	10

## Datapro Summary

The local area network market is one of the fastest-growing segments of the computer and communications industry. Developing standards have helped to fuel the growth of this market. The IEEE 802 standards body is chartered to define the standards for local area networking, and this report provides an overview of the IEEE 802 LAN standards, some of which are not yet final. The Institute of Electrical and Electronics Engineers (IEEE) began Project 802 in February 1980 in an attempt to establish standards in advance of the local area network (LAN) market. The IEEE 802 Committee has defined interface and protocol specifications for logical link control and access methods for various LAN topologies. The project has maintained an open-door policy, and from 20 to 300 people have participated in any one working group. Most of the participants work for a computer or network components vendor, and many have a communications or marketing background. Since the project's commencement, it has been under intense scrutiny by the computer industry, because the resulting set of standards has—and will continue to have—a significant impact on the growing LAN market.

## Introduction (A Little History)

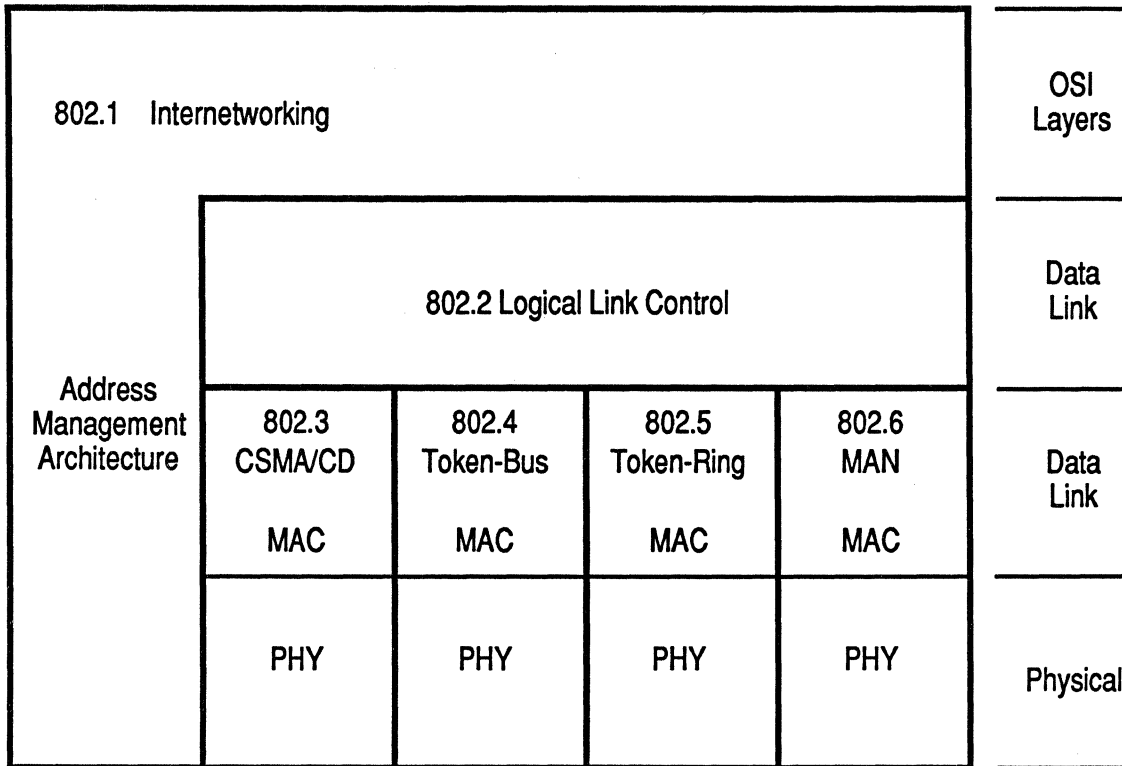
During the mid-to-late 1970s, a small company in Texas developed a capability to provide access to shared direct access storage devices (DASDs) from microprocessor workstations located within a reasonable distance from the DASD. The company was Datapoint Corp. and the technology was Arcnet—Attached Resource Computer Network. At about the same time, Xerox Corp. was developing its experimental Ethernet.

These were the first local area networks (LANs) to be offered as commercial products.

In the 1980s, with the near demise of Datapoint, and Xerox' alliance with Digital and Intel, the IEEE had little choice when it adopted an "Ethernet-like" approach for its local network standard. Other organizations such as General Motors and IBM had their own ideas about what the "ideal" LAN should look like. The result was a family of LAN standards to be known as the IEEE Project 802 LAN standards.

Despite these developments, Arcnet did not disappear. Vendors that had acquired the licenses continued to make Arcnet interfaces, but

Figure 1.  
IEEE Project 802 Working Groups



The IEEE 802 standards address the two lower layers (Physical, Data Link) of the OSI Reference Model.

now focused on the PC as the workstation. Arcnet continues to enjoy a useful life as a mature and highly functional proprietary LAN implementation.

With the ever-increasing popularity of LANs, new requirements have developed for higher speed networks of greater geographic range. In response, vendors have developed alternatives to the standard implementations, and standards bodies such as ANSI have produced the Fiber Distributed Data Interface (FDDI), among others.

The bottom line is that while there are currently a fair number of standard LAN implementations, there are also proprietary implementations from leading vendors, which, while not "standard," may suit one's needs very well. As long as technology continues to improve, entrepreneurs will continue to come up with "better ways." Some of these better ways will fade quickly after initial flurries of excitement, while others will pave the way for new and better standards.

Standards are not static, and the network architect must realize that there will always be a better solution tomorrow. Of course, if one continues to wait for the better solution, no solution will ever be implemented.

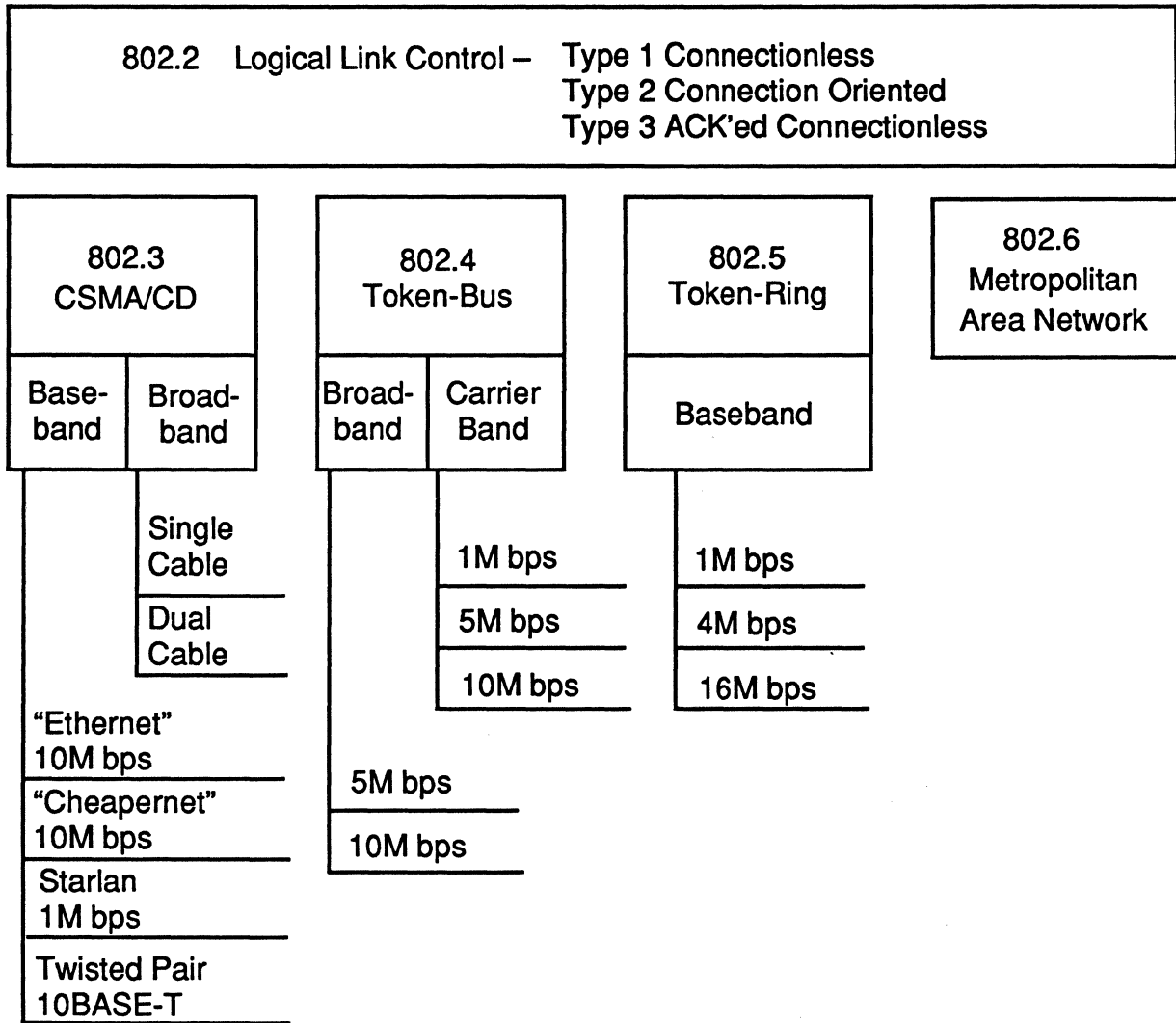
It is also essential to realize that a variety of forces competes in the standards development world—each with its own agenda (whether hidden or unhidden). The result is often a less than perfect compromise. It is often said that "the two things you really don't want to watch being made [if you're going to be involved with them] are sausages and computer network standards."

### Overview of the IEEE 802 Standards

The IEEE 802 standards essentially address only the two lower layers of the Open Systems Interconnection (OSI) Reference Model (see Figure 1).

The Physical Layer corresponds to the OSI Physical Layer, while the OSI Data Link Layer is divided into two sublayers: Medium Access Control (MAC) and Logical Link Control (LLC). The Medium Access Control sublayer addresses the

Figure 2.  
IEEE Standards Variations



There are variations within each of the IEEE 802 standards.

specific procedural issues associated with distributed arbitration of access to the channel. The Logical Link Control sublayer provides a mechanism accommodating those functions of wide area network Data Link protocols that pertain to LAN link management. Unlike the wide area Data Link protocols such as High-level Data Link Control (HDLC), which addresses specific nodes, LLC frames contain only service access points or internal memory addresses of software entities. Physical node addresses are handled by the MAC sublayer.

There are four basic access methods with published standards, as well as a subset of higher

Data Link Layer functions. In addition, there are several working groups whose activities are focused on specific technologies which are applicable across a broad range of the access methods.

The carrier sense multiple access with collision detection (CSMA/CD) method was the first to be developed by the IEEE and was modeled after the Digital/Intel/Xerox (DIX) Ethernet. Although there are differences between the Ethernet and 802.3, manufacturers now typically produce hardware that can support both, so that effectively the

**Table 1. Ethernet/IEEE 802.3 Differences**

Feature	Ethernet Version 2	IEEE 802.3
<b>Specification</b>	1982 Blue Book	1985, 1989
<b>Transceiver cable</b>	4 Pairs AWG 20	4 Pairs AWG 20
<b>Grounding at host</b>	Inner/Outer shield common at backshell & pin 1	Inner shield to pin 4; outer backshell
<b>Signal Quality Error (SQE)</b>	Yes, Heartbeat	Yes, Heartbeat
<b>Repeater specification</b>	None	Multiple collision protection
<b>Jabber control</b>	Yes	Yes
<b>Type/length field</b>	Type (>1500)	Length (<1500)
<b>Coaxial cable</b>	50-ohm Double shielded	50-ohm Double shielded

two are compatible. Differences in the packet format are resolved in firmware for a particular implementation. We will continue to use the terms Ethernet and IEEE 802.3 interchangeably. Table 1 defines the differences between Ethernet and IEEE 802.3 implementations.

The 802.4 specifications were developed primarily in response to requirements for the deterministic performance of token passing, coupled with the facility of bus-oriented cabling. The use of broadband technology provided the additional benefits of increased bandwidth, geographic coverage, and numbers of terminations.

The 802.5 token-ring specification was developed under the "guidance" of IBM and reflected the emerging "blue" perspective on local area networking. While the initial versions of the network provided less capacity than Ethernet, the expected improvements due to deterministic performance and priority mechanisms yielded other benefits.

With time, however, we have seen a wide variety of implementations emerge—each reflecting a specific application arena (see Figure 2). Some of these have been standardized, while others will likely become standards in the near future.

Work began recently in several new technology areas including integrated voice and data (IEEE 802.9—IVD), security standards for interoperable LANs (IEEE 802.10—SILS), and wireless LANs (IEEE 802.11—WLAN). Preliminary work continues on the use of fiber optics by the Fiber Optic Technical Advisory Group (IEEE 802.8—FOTAG).

With this backdrop, we will explore the specific 802 LAN standards.

### IEEE 802.3 (CSMA/CD)

IEEE 802.3 standards are characterized by a shorthand notation which facilitates their description in as few words as possible. The notation (e.g., 10BASE5) is composed of three elements:

- 10—megabits per second
- BASE—baseband (or BROAD for broadband)
- 5—meters per segment divided by 106

With standards adopted more recently, such as 10BASE-T, IEEE has tried to be more descriptive with its notation. For example, the "T" in the 10BASE-T standard is short for "twisted-pair wiring."

### 10BASE5

Using the formula, 10BASE5 means 10M bps, baseband, 500-meter segments. This was the first version of the specification to be developed, and it most closely resembled the earlier Ethernet Versions 1 and 2 (1980 and 1982, respectively). The 10BASE5 LAN employed the "thick Ethernet" 50-ohm coaxial cable. While this cable is difficult and relatively expensive to install, it provides significant advantages over other implementations in terms of distance and the number of terminations permitted for each segment.

The workstation contains an adapter board, called the "bus controller" in Ethernet parlance (see Figure 3). Attached to the bus controller is a multiconductor cable known as the Attachment Unit Interface (AUI) cable. This, in turn, is connected to a transceiver/tap assembly called the Medium Attachment Unit (MAU), which is connected to the Ethernet trunk cable employing a "vampire" tap.

When Ethernet products were first developed, this assemblage of components normally cost \$1,500 to \$2,000. Since LAN implementations are very sensitive to workstation termination costs, less expensive alternatives were required. This problem was resolved in two ways. First, vendors developed less expensive implementations (the old “better way” trick), which we will explore in a moment; and second, the natural momentum in declining semiconductor costs reduced these implementations to a fraction of their former costs.

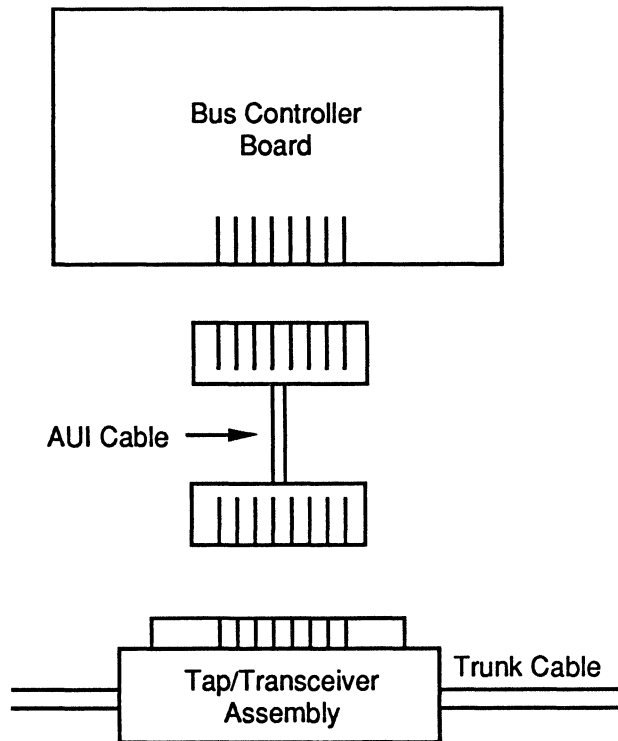
Due to the sensitive timing issues associated with the performance of the CSMA/CD protocols, limits were imposed upon the overall length of a multisegment LAN, as well as the maximum signaling rate. A typical large-scale CSMA/CD LAN is limited to a distance of 3,000 meters between any two communicating stations. This is often implemented by using three 500-meter segments, two 500-meter link segments, and up to ten 50-meter AUI cables. An important distinction between a link and a segment should be noted. Segments can have workstations attached, while links are simply media used to extend the overall distance of the LAN.

Figure 4 illustrates a 10BASE5 LAN with the maximum distance between two workstations. Other constraints associated with 10BASE5 LANs concern the number of devices that can be terminated on the trunk cable. Up to 100 devices can be placed on a 500-meter segment, with a maximum of 1,024 devices on the entire network. This limitation can be circumvented through the use of bridges, which partition a LAN into several connected, but independent LANs—thus yielding the maximum length and number of workstations for each.

### 10BASE2

10BASE2 (also known as “thin Ethernet” or “Cheapernet”) employs a thin flexible coaxial cable (RG-58) that connects to the bus controller board in the workstation by means of a BNC “T” connector. In earlier implementations, the transceiver functions were onboard, but in the interests of using the bus controller for either implementation, MAUs and bus controllers have been developed which provide options for both 10BASE5 “vampire” taps and 10BASE2 BNC connectors. More recently, board manufacturers commonly provide boards with built-in transceivers that can

Figure 3.  
10BASE5 Termination Hardware



The 10BASE5 version of IEEE 802.3 uses thick Ethernet coaxial cable and various termination hardware.

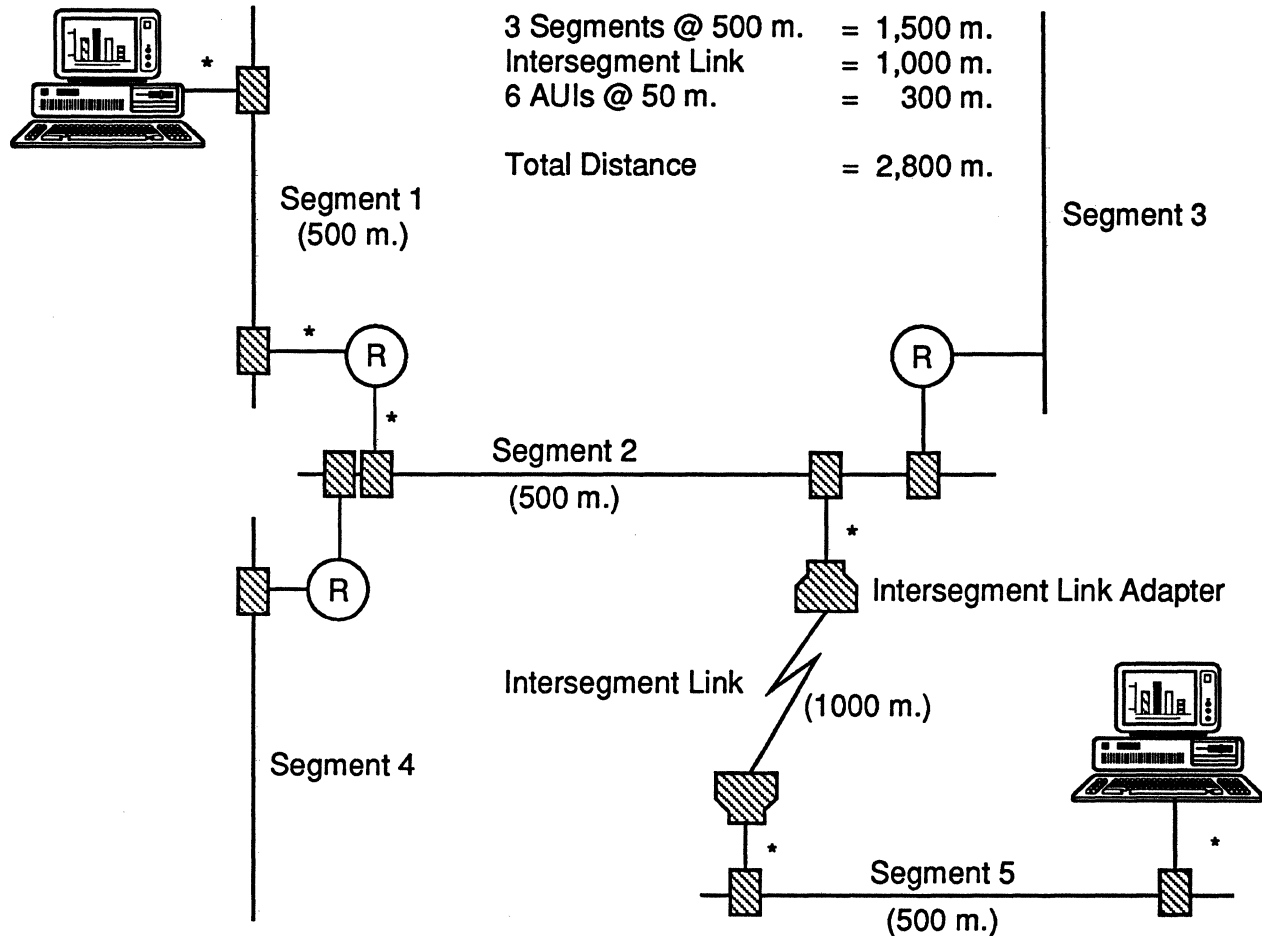
be switched on or off by the component manufacturer for a particular application.

The standard 10BASE2 LAN can support only 30 terminations on each coaxial cable segment of 185 meters. While this may seem like a major constraint, it is often adequate for most work area environments. Where a requirement exists for interconnecting multiple work areas, or work areas with multiple 10BASE2 segments, a backbone 10BASE5 segment can be employed to provide intersegment connectivity. Figure 5 illustrates this type of configuration. Table 2 shows the differences between 10BASE5 and 10BASE2.

### 1BASE5

This standard approach was contributed by AT&T to accommodate its earlier Starlan products. It operates at 1M bps, and as such is often most useful for small work areas or low traffic environments. 1BASE5 also employs inexpensive twisted-pair wire interconnected through a hierarchical system

Figure 4.  
Multisegment CSMA/CD LAN



\*50 m. each AUI.

For a multisegment 802.3 LAN, the maximum distances between segments can vary; however, the maximum distance between any two communicating stations is limited to 3,000 meters.

of concentrator hubs. The hubs emulate a bus configuration by broadcasting all data and collision information on all ports.

### 10BASE-T

One of the most exciting developments on the local network scene has been the development of the 10M bps unshielded twisted-pair (UTP) Ethernet. This implementation has now received final approval from the IEEE. One of the best-known products to claim compliance with this standard is SynOptics' LattisNet. There are now several major manufacturers producing products meeting this standard (including AT&T, HP, Digital, and 3Com); in fact, virtually every vendor active in the Ethernet market now offers 10BASE-T products.

It is important to note that these implementations will be limited to 100-meter segments due to

the greater attenuation and signaling difficulties of twisted pair. This should not present any unusual problems since these networks' connections usually only have to reach to the "communications closet." From there, FOIRL and coax can be used to concatenate and interconnect LANs with standards such as 10BASE2.

It is imperative, however, that organizations planning these networks have their existing twisted-pair wire certified for both attenuation and capacitance before making any assumptions on its salvageability.

Like the AT&T Starlan, this system uses a hub concentrator to interconnect multiple stations and emulate the bus operation.



### 10BROAD36

The 10BROAD36 implementation uses much of the same hardware as the baseband implementations. The essential difference is the substitution of a broadband electronics unit and a passive broadband tap for the baseband MAU. This enables an organization to use its existing bus controller boards in the workstations for connection to either a baseband or broadband system. In recent years, this standard is being used less frequently.

The primary functions of the broadband electronics unit are to create the frequency-derived channels of 14MHz for data and 4MHz for collision consensus. It also converts the signals from the baseband-coded signal of the AUI to the analog signal necessary on the broadband channel (see Figure 6).

Workstations can be placed up to 1,800 meters from the "head-end" of the broadband cable plant. By placing the head-end in the center of the configuration, workstations can be installed up to 3,600 meters from each other.

### IEEE 802.3 Standards Status

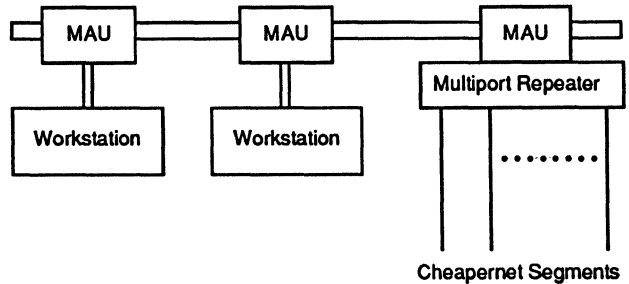
Within the IEEE 802.3 group, the following standards have been completed as of this date:

- CSMA/CD Medium Access Control Layer
- 10BASE5 Medium
- 10BASE2 Medium
- 10BROAD36 Medium
- Repeater Specifications
- Fiber Optic Inter Repeater Link (supports distances up to one kilometer)
- Layer Management
- 10BASE-T
- ATS for AUI Conformance Testing

Several projects remain open, with adoption expected imminently on some:

- Conformance Testing
- Maintenance
- 10BASE-F (Fiber Optics Task Force)
- Hub Management

Figure 5.  
10BASE5/10BASE2 Interconnectivity



*In environments with multiple work areas, or work areas with multiple 10BASE2 LAN segments, a backbone 10BASE5 segment can be used to provide intersegment connectivity.*

### IEEE 802.4 (Token Bus)

The 802.4 Token Bus working group wrestled with the issues of coordinating both IEEE and ISO standards development activities. Although the initial broadband implementations of the token bus appeared to be highly flexible and desirable in terms of the generic manufacturing requirements, a number of difficulties arose.

First, the industry found that migration from the early versions of the Manufacturing Automation Protocol (MAP) suite (Version 2.1) to current specifications (Version 3.0) is less than facile. It has become a manager's nightmare for a number of reasons.

For instance, fewer and fewer people are interested in broadband implementations primarily due to the difficulty in design, installation, and maintenance. Additionally, the apparent benefits of broadband networks, in terms of the number of terminations, geographic range, and bandwidth, have been overtaken and negated by the introduction of Medium Access Control bridges that provide even greater capabilities for baseband networks—nearly transparently. These bridges enable an organization to increase the traffic loading by simply partitioning the network and eliminating the concern. Couple these high-risk implementation issues with the scarcity of products, difficulty in migration from MAP 2.1 to 3.0, and soft industry support, and one will find that the token bus presents a quagmire of implementation risks that most managers would rather avoid.

There is some hope on the horizon for the medium access specification. Other broadband physical medium specifications are being developed for optical fiber. Some difficulties lie ahead

**Table 2. 802.3 10BASE5/10BASE2 Differences**

Feature	10BASE5	10BASE2
Name	802.3 "Ethernet"	Cheapernet, THIN Ethernet, THINWIRE Ethernet, etc.
Type of cable	50Ω Thick dual shield	50Ω RG-58
Maximum segment length	500 m.	185 m.
Spacing of devices on cable	2.5 m. minimum	0.5 m. minimum
Maximum number of taps for a segment	100	30
Maximum number of full repeaters in a path between two stations	2	2
Type of taps	Vampire or coax	BNC "T" connector for "daisy chaining"

here since the dominant fiber specification in the U.S. is the 62.5μm fiber specified by ANSI for the Fiber Distributed Data Interface. In Japan and Europe, 50μm fiber is a more common implementation. In the final versions of this standard, both options are permitted—62.5μm is the standard and 50μm is allowed.

The 802.4 Token Bus architecture has matured despite the uncertainties presented by the MAP protocol suite. Standards for medium access control, broadband media, carrier-band media, and optical fiber have been completed. Open projects include conformance testing.

### IEEE 802.5 (Token-Ring)

The token-ring implementation, which has received so much attention since it was first approved in 1985, has undergone a variety of modifications, and completion of essential specifications.

### Media Issues

The initial version of the ring was a 4M bps implementation which ran on shielded twisted-pair wire. The issues surrounding shielded twisted pair have always been controversial. Telephony carriers avoid shielded wire to the extent possible, since the shielding introduces capacitance changes and ultimately increases attenuation, thus requiring more frequent repeater placement. The LAN proponents, such as IBM, feel differently. They contend that the shielding protects the media from unwanted EMI/RFI and that the distance between repeaters is not an issue since each station is its own repeater. The company using shielded wiring must decide if lower attenuation is worth the extra

cost associated with a thicker (i.e., harder to install) and more expensive wire.

Considering the context of their respective positions, both contenders are correct. In the case of LANs, however, the shielding does buy some value. One thing we can be sure of is that where there is a requirement, someone will stand up to fill the niche. Thus, when IBM introduced the 16M bps token-ring network, running only on shielded wire, it was not surprising that other vendors immediately introduced unshielded wire 16M bps implementations. It is reasonably certain that in due time, the IEEE 802.5 working group will introduce a specification for unshielded twisted-pair wire. Considering the work that the Electronic Industries Association (EIA) has done concerning intrabuilding wiring (PN-1907), it is likely that the EIA specifications for unshielded wire will be candidates for the 16M bps ring.

Other media-related issues being explored by the 802.5 group are the use of Optical Fiber Station Attachment equipment and redundant media for backup (reconfiguring dual rings).

### Token-Passing and Multi-Ring Protocol Issues

Recently, IBM introduced a new version of the token-passing protocol called "Early Token Release." This new protocol is intended to make more efficient use of the available bandwidth on physically large rings operating with particularly small packets. In earlier versions of the token-passing protocol, a new free token could not be released by the sending station until it recognized the address in its own packet coming back around the ring to itself. If the packet was small, and the ring was large, there was a great deal of wasted time on the medium.

Using Early Token Release, a sending station can release the free token immediately upon completing its transmission. The empty time slots on the ring can now be used by other parties. When coupled with the 16M bps ring, this new protocol appears to have significant advantages in terms of performance.

Another area of interest in the token-passing world is the controversy on Medium Access Control Bridges. While Ethernet proponents prefer a minimum spanning-tree approach, many token-ring developers prefer source routing bridges. The 802 spanning tree bridge is an approved standard.

A discussion of Medium Access Control Bridges can be found in the Data Link Layer Repeaters section.

### 802.5 Standards Status

Presently, the following completed standards are available from the 802.5 working group:

- ANSI/IEEE 802.5 Token-Passing Ring (1985)
- 802.5A Station Management Functions Revision
- 802.5E Management Entity Specification
- 802.5F 16M bps Operation
- 802.5H Acknowledged Connectionless Logical Link Control
- 802.5I Early Token Release
- ANSI/IEEE 802.5 (1989)
- 802.5B Unshielded Twisted-Pair (being published)
- 802.5C Reconfiguring Dual Ring Specifications (being published) (redundant media)

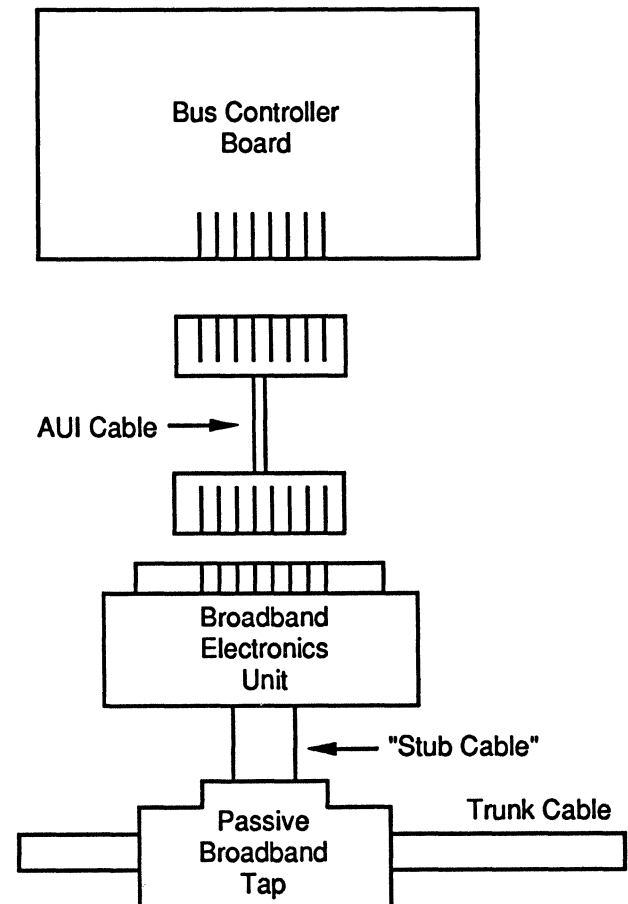
The list of ongoing open projects includes:

- 802.5D Multi-Ring Configurations
- 802.5G Conformance Testing
- 802.5J Optical Fiber Station Attachment
- UTP 4/16 megabits per second

### IEEE 802.6 (Metropolitan Area Network)

The IEEE 802.6 Metropolitan Area Network is a fourth MAC alternative that has been defined by the IEEE. Early plans for this moderate geographic area service focused on CATV-type networks, while later proposals revolved around a slotted ring concept. Current specifications call for a Queued

Figure 6.  
10BROAD36 Model



*A 10BROAD36 broadband 802.3 implementation uses much of the same hardware as base-band 802.3 LANs.*

Packet Synchronous Switch, which is a hybrid approach. It has been developed under the auspices of the Australian Postal, Telephone, and Telegraph administration and appears to be gaining general acceptance. The standard is approved.

### IEEE 802.9 (Integrated Voice & Data LAN)

Topics under consideration by this working group include MAC frame delimiting, TDM frame formats, 20M bps PMD, and Layer Management. Both medium (4M bps) and higher speed Physical Layer standards are being investigated. Voting on a relatively "mature" specification is expected in the near future.

**IEEE 802.10 (Standard for Interoperable LAN Security)**

This group is making progress on defining an architectural model for implementing interoperable LAN security. Licensing terms for the use of patented public key technology are being studied. Group 802.10B is working on secure data exchange. Group 802.10C is studying Key Management. The group is not predicting a standards ballot in the immediate future.

**IEEE 802.11 (Wireless Local Area Network)**

Interest in this standard comes from all over the world including Japan, Canada, and Europe. The group has started on specifications for MAC and the Physical Layer, though a ballotable draft standard is still at least a year away. The work done by this group will be applicable to other MAC standards including 802.3, 802.4, and 802.5. Interest includes radio frequency and the infrared spectrum. The group hopes to have a standard in place by the end of 1992. The group is keeping other standards bodies, including T1P1, ETSI, and ECMA, informed of its progress.

---

**Local Area Network Interconnection**

As LANs proliferate, it is becoming more important that standard techniques for interconnection be adopted.

**IEEE 802.1 Higher Layer Interface****Data Link Layer Repeaters**

Interconnection of similar but separate LANs has resulted in the need for specifications on Medium Access Control bridges. MAC bridges are hardware/software implementations that are limited to resolving the MAC sublayer differences between two or more interconnected LANs. No further higher layer protocol translation is required, and they are often transparent to the user in terms of delay and performance.

MAC bridge specifications have been addressed by the IEEE 802 working groups. The 802.10, 802.1, and 802.5 teams have developed significantly different approaches, but even these are beginning to converge. It is likely that within the next year, we will see more mature guidance in this area.

The current approaches are the Minimum Spanning Tree for bus implementations and the Source Routing bridge for interconnected rings.

The essential difference is that in the bus environment, only one path between any two devices exists. The bridges learn the LAN segment and node addresses and filter packets accordingly as required. Provision for multiple alternative paths is provided in the interconnected ring environment, which, in turn, yields a requirement for a routing protocol. This routing protocol is facilitated by adding "routing information" (RI) fields to the packet header. The RI field contains all of the source node routing information necessary for the bridge to determine which path is to be adopted for a specific packet.

There are certainly advantages and disadvantages to both of these approaches, but the common goals are to provide global, transparent interconnection. Global in the sense that any device on any LAN can share resources with any device on any other LAN; transparent in the sense that performance must be adequate to ensure that access to remote resources is provided rapidly and accurately. This guarantees that users do not perceive a difference between local and global objects.

**IEEE 802.3 Physical Layer Repeaters**

In the case of the IEEE 802.3 CSMA/CD LANs, intra-LAN segment connection standards are well developed and mature. These physical layer relays are implemented in the form of repeaters that regenerate the signals from one segment for retransmission to the next. The unique aspect of these repeaters is that they must be capable of retransmitting collisions as well as data frames. Unlike Data Link Layer relays (or MAC bridges), these repeaters are not addressable. Since all segments are part of a unified LAN, the nature of the shared channel must be preserved by broadcasting all information to all terminated devices.

The latest specifications for repeaters are contained in the IEEE 802.3C supplement (1989). Unlike the earlier version of this supplement (1988), this specification provides rich detail on coaxial cable, AUI, and optical fiber repeater interfaces. Repeater specifications now pertain to all 10BASE implementations.

In addition to the functions described above, repeaters as specified in the 802.3C supplement can provide "partitioning" between segments.

**Table 3. Logical Link Control Alternatives**

Service	Type 1	Type 2	Type 3
<b>Basic Service</b>	Connectionless	Connection	ACK'ed connectionless
<b>Acknowledgments</b>	No	Yes	Yes
<b>Error Recovery</b>	No	Yes	Yes
<b>Flow Control</b>	No	Yes	No

Thus, if conditions on a given segment are causing the extensive proliferation of collisions, the rest of the LAN can be protected from this anomaly. The repeater will count the number of collisions from the source segment and interrupt these from transmission to the next segment. This function is designed to address an abnormal situation such as a cable break or network card failure.

#### IEEE 802.4 Physical Layer Repeaters

The issues of signal attenuation in a broadband LAN are normally resolved in two ways. First, the maximum placement of a device from the head-end provides a maximum bound on signal loss in the context of attenuation. Second, since many stations can be connected to the bus, each resulting in a specific "insertion loss," amplifiers are often required to ensure that the total loss does not exceed specifications.

The IEEE 802.4 broadband bus specifications define a Regenerative Repeater Machine (RRM) as an optional component that is present only in special repeater stations such as the head-end. Since broadband systems are analog, amplifiers are usually used to boost signal strength. Regenerative repeaters actually re-create a new signal in accordance with amplitude and time specifications.

A regenerative repeater is also defined for the single-channel carrier-band system. Since the latter is not a multichannel broadband bus (a medium supporting multiple frequency-derived channels such as a Community Antenna Television [CATV] system), a head-end is not required to facilitate this function. Physical placement of these devices is a function of the number and placement of user devices on the network. There are no explicit maximum terminations defined in the specification, but the standard suggests that 30 may be an appropriate user limitation.

#### IEEE 802.5 Physical Layer Repeaters

The nature of a token-passing ring obviates the necessity for repeaters, since each station's ring interface performs repeater functions. The maximum attenuation of a signal is thus guaranteed by limiting the distance between any two devices in the ring. As with 802.3, the issues of overall length of the ring impact protocol performance as opposed to signal attenuation.

#### IEEE 802.2 Logical Link Control

The IEEE 802.2 Logical Link Control (LLC) specifications include those Data Link Layer functions that are common to all 802 LAN MAC sublayer alternatives. Three basic service types are provided.

##### Type 1 (Connectionless)

This service provides a best-effort delivery mechanism between origin and destination nodes. No call or logical circuit establishment procedures are invoked. Each packet is treated as an independent entity by the network. There are no flow control mechanisms or acknowledgments. If the packet arrives at the destination—all well and good. If not, it is the responsibility of the higher layers to resolve the problem through time-outs and retransmission.

##### Type 2 (Connection Oriented)

Like many wide area network protocols, this service requires that a logical circuit or call be established for the duration of the exchange between the origin and destination nodes. Packets usually travel in sequence and are not routed as independent entities. Positive acknowledgments and flow control mechanisms are an integral part of this service.

**Type 3 (ACK'ed Connectionless)**

No circuit is established in this service variation, but acknowledgments are required from the destination node. This type of service adds additional reliability to Type 1, but without the potentially excessive overhead of Type 2.

Specific LAN types lend themselves to different types of service. Table 3 illustrates the LLC variations as they apply to the different MAC implementations.

---

**Summary**

The IEEE 802 local area network standards have evolved and matured significantly since their development in the early 1980s. It is essential that we

not view this maturation process as an end. The standards will continue to evolve, and as new technologies and requirements develop, new standards will follow. The ideal utopian environment would be for standards development to lead product development, but it is unrealistic to believe, in an environment as volatile as local area networking, that vendors will wait patiently while users clamor for more and better products.

The IEEE will be faced with a continuing challenge to ensure that as new requirements and products evolve, the standards also evolve. This challenge will also be coupled with a requirement that migration from prior implementations is as painless as possible—both in terms of development risk and cost. ■

# ANSI Fiber Distributed Data Interface (FDDI) Standards

## In this report:

Applications.....	2
OSI Reference Model.....	3
FDDI Specifications.....	5
FDDI-II and Other Enhancement Efforts .....	13
FDDI Information Sources .....	14

## Datapro Summary

The FDDI standard describes an optical fiber-based local area network operating at 100M bps. The standard covers the first two layers of the OSI Reference Model. The LAN is configured as dual redundant fiber rings to protect against disruptions caused by station failures. The standard is essentially complete. The last of the four FDDI components, Station Management (SMT), was forwarded to the ANSI X3T9 committee to be prepared for a letter ballot. The draft standard is technically stable. Since the initial standard was proposed, ANSI has begun work on enhancements to FDDI. FDDI II, which is considered a superset of FDDI, will add the capability of carrying voice and video signals over FDDI-based networks. FDDI follow-on, or FFOL, will allow interconnection of multiple FDDI networks, connection to wideband facilities, and Synchronous Optical Network (SONET) compatibility.

## Technology Overview

FDDI addresses the bottom two layers of the OSI model. FDDI's designers expect users will use ISO protocols for the other layers, where possible. The optical-based FDDI LAN was designed to provide the same type of serial interconnection provided by LANs while providing the high bandwidth, inherent noise immunity, and security offered by fiber. At FDDI's inception in 1982, fiber was used mostly for point-to-point applications, and not for the many configurations supported by LANs. In this sense FDDI was a breakthrough. Although it is possible to achieve considerably higher data rates over fiber (up to 3.7G bps with current point-to-point technology, and 500M bps on rings), higher rates result in significantly increased costs and shorter transmission distances between repeaters. Its designers intend FDDI to provide relatively inexpensive connectivity and therefore focused on the 100M bps rate. FDDI can be configured to support a sustained

transfer rate of approximately 80M bps. The remaining bandwidth is reserved for various overhead functions.

FDDI is a token-passing, dual-ring network accommodating synchronous and asynchronous data transmission as well as isochronous channels for realtime digitized voice and compressed video. Unlike existing open standards for LANs, where fiber optic variants have followed copper implementations, FDDI has been designed from the start as a fiber optic network. This has involved standardization issues in such areas as duplex optical connectors, fiber characteristics, optical bandwidth, bypass relays, and cable assemblies. The FDDI ring is designed for an overall bit error rate of less than  $10^{-9}$ . The network can tolerate up to 2 km. of fiber between stations and can support a total cable distance of 100 km. around the ring with 500 attachments (1,000 physical connections and a total fiber path of 200 kilometers). FDDI topology

—By *Tim McElgunn*  
Assistant Editor/Analyst

is a counterrotating, token-passing ring (note the arrows in Figure 1). FDDI, however, is not part of the well-established IEEE 802 family of LAN standards.

FDDI operates at 1300 nanometers (nm.). Current transmitter/receiver fiber technology operates at 850 nm., 1,300 nm., or 1,550 nm. While performance increases with wavelength, so does cost. For local data communications, both in LANs and in point-to-point applications employing fiber optic modems, 850-nm. light sources are typically employed; however, this technology becomes unfeasible for 100M bps beyond a couple of miles. At the other end of the range (1,550 nm.), the system becomes expensive and may provide unnecessary bandwidth. The committee designing FDDI also investigated short-wavelength implementations. It became evident that to meet all the requirements, particularly the two-kilometer station-to-station spacing, the system would require 1,300-nm. wavelength. Using 1,300-nm. technology, less expensive light-emitting diodes (LEDs) provide distance and data rates within the range desired for LANs and LAN backbones. The issue of fiber size was settled after the wavelength decision.

The FDDI standard directly addresses the need for reliability. A backbone system transports a large number of user sessions, and its loss would be serious. FDDI incorporates three reliability-enhancing features. First, a failed or unpowered station can be bypassed by an optional automatic optical bypass switch; second, wiring concentrators are used in a star wiring strategy to facilitate fault isolation and correction; and third, two rings interconnect stations so that failure of a repeater or cable link results in the automatic reconfiguration of the network (loopback).

FDDI consists of four adopted or draft proposals. The Media Access Control (MAC) is specified in X3.139-1987, approved on November 5, 1986. The Physical layer standard (PHY) is contained in X3.148-1988. The Physical Medium Dependent (PMD) is specified in X3.166-1988. Station Management (SMT) is described in draft proposal X3T9.5/84-49.

MAC was the first FDDI standard completed. MAC specifies access to the medium, addressing, data checking, and frame generation/reception. PMD specifies the optical fiber link and related optical components. PHY specifies encode/decode, clocking, and data framing. A Station Management standard specifies the control required for proper operation of stations on the ring. Services include station management, configuration management, fault isolation and recovery, and scheduling procedures.

The token-passing protocol used in FDDI is based on the IEEE 802.5 standard for a 4M or 16M bps twisted-pair ring. When no station is transmitting, the token (a small control packet) circulates around the ring. When a station needs to transmit data, it waits until it detects the token. The station then removes the token from the ring and transmits the data in packetized form. The data packet will circulate around the ring until it reaches its intended recipient. That station copies the packet and returns it to the ring. The packet continues to the transmission station, which removes it from the network and places the token back on the ring for the next station to use.

The choice of FDDI's ring topology is based on optical communications characteristics. Bus and passive star topologies require the optical transmission to be detected at several sources simultaneously. Although practical fiber optic taps are available, attenuation is still such that the number of nodes is relatively limited. Since fiber optic

transmission is best suited to a point-to-point configuration, FDDI included this aspect in its definition. Two topologies are possible with point-to-point links: the active star and the ring. The active star has a single point of failure. A defective hub incapacitates the entire network. A similar problem affects a single ring. Consequently, the FDDI specification calls for a dual ring.

## FDDI Station Types

Two FDDI station types are allowed: Class A and Class B. Class A stations are dual-attachment stations. They connect to both the primary and secondary rings of the network. Data flows in opposite directions on the two rings. A Class A station can act as a wiring concentrator to interconnect several Class B (single attachment) stations. Wiring concentrators give the network administrator a common maintenance point for a number of stations (see Figure 1). Ring reconfiguration following a failure is shown in Figure 2. Here, the link between two Class A devices is broken. The two stations detect the failure and patch the network by looping the data path onto the secondary ring, thus creating a single ring.

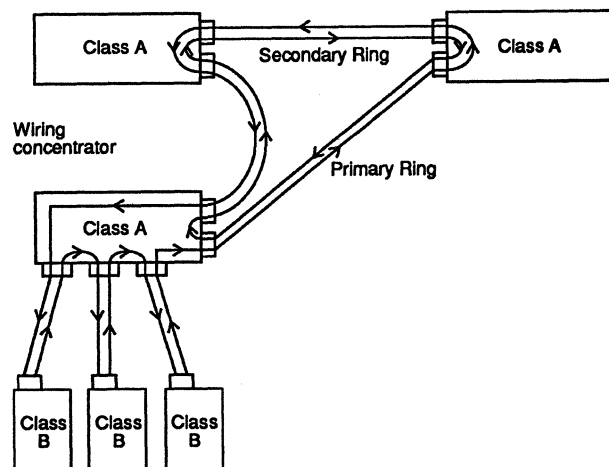
Figure 3 illustrates a break in the cable between a Class A and a Class B device. Here, communication continues over the primary link, as the Class A device detects the failure and makes appropriate internal modifications. The Class B device, however, remains detached. Class B devices trade lower cost against the fault tolerance of the more sophisticated Class A device.

The secondary ring can also be used to carry traffic. This gives the fully configured FDDI system 200M bps of effective throughput. When the two rings merge to support a backup configuration, the network data rate drops to 100M bps.

## Applications

Although optical fiber is widely deployed in the telecommunications environment (long-hauls, interoffice, feeder plant, and so on), it still has not exploded in the LAN environment. (This is not true in Japan, where fiber optic LANs dominate.) Three reasons can be identified: 1) increased technical complexity compared to passive copper and coaxial cable; 2) cost considerations—especially the

Figure 1.  
The FDDI Environment





cost of writing off huge existing cable investments; and 3) lack of a workable standard. FDDI solves the third problem, and in the process it also begins to resolve the second.

FDDI allows designers to 1) build larger capacity LANs or LAN backbones to serve new data needs (file transfer, graphics, and so on) and some voice needs; or 2) interconnect LANs in metropolitan area networks (MANs). Thus, FDDI can be used directly as a LAN or as a backbone to interconnect slower speed LANs into a single network over relatively large geographies.

The initial application of FDDI as a "backend" interconnect for high-powered computing devices and peripherals required a high degree of fault tolerance and data integrity. As developers proceeded, it became obvious that FDDI could also serve high-speed "front-end" applications. Front-end applications include terminal-to-terminal and terminal-to-server communications typical of a LAN. In large networks (3,000 to 10,000 terminals, particularly when workstations are involved), the aggregate demand for network resources can overwhelm a 4M/16M bps token-ring or 10M bps Ethernet LAN. At that point, FDDI's bandwidth becomes important.

A high-speed FDDI ring is ideal as a backbone for other "departmental" LANs, typically operating at lower speeds. High capacity on LANs can be achieved in two ways: by using multiple channels at low speeds or by using one channel at a relatively high speed. The multiple-channel approach can be used with a broadband bus LAN. A drawback of this approach is that bridges must be provided between channels, and the architecture must be designed to avoid high rates of interchannel traffic and bottlenecks at the bridges.

The one area where FDDI is not well suited is broadband LANs carrying full analog video (6MHz). A coaxial-based broadband LAN can easily carry multiple channels of video, as well as data, in analog form. Digitizing a 6MHz TV channel, however, results in a 45M to 90M bps data rate (45M bps is a slightly compressed version), which can easily swamp an FDDI backbone. FDDI-II, discussed later in this report, has been designed to handle video efficiently.

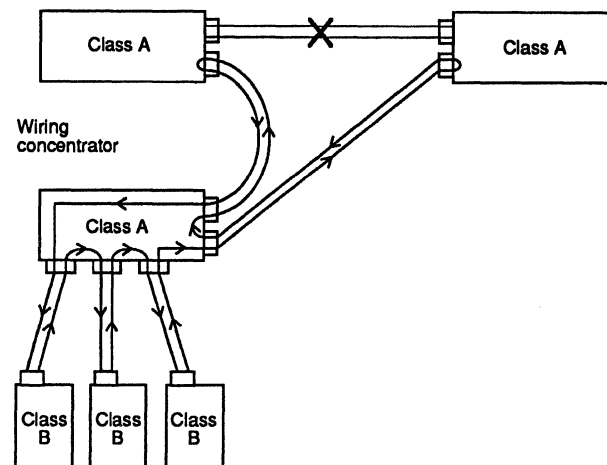
FDDI is not the end, however. Even larger bandwidths are envisioned in the next five years under the thrust of Broadband ISDN (BISDN). BISDN for video needs, particularly in a (High Definition TV (HDTV)) environment, will require approximately 150M bps of dedicated bandwidth per channel. Planners are now discussing delivery of up to three channels per domicile, reaching into the 600M bps range. FDDI is a shared-medium technology; uncompressed video applications and/or local loop applications may not be capable of using networks built on the original FDDI standards.

## OSI Reference Model

To fully understand the FDDI standard, some knowledge of the OSI model and how it applies to the LAN environment is necessary. A brief discussion of the OSI model is followed by a more detailed description of FDDI specifications and emerging vendor products. Readers intimately familiar with this material can skip to the section entitled FDDI Specifications.

The OSI Reference Model imposes order and structure on data communications. To achieve orderly communications, many functions must be performed. It is natural to group these functions into layers that share task affinity and logical proximity. OSI layers are hierarchical in the

Figure 2.  
Failure Recovery



Ring rearrangement under failure.

sense that a given layer calls for the services of the layer immediately beneath it. The given layer cannot ask for the services of a layer at a higher level, nor can it skip the layer beneath it to directly reach a lower layer, or even jump into the middle of another layer. The model includes precise definitions of the services provided by each layer to its next higher layer and request procedures.

Services defined for a given layer are, in turn, employed by the layer immediately above it. Each transmitting layer passes down to the lower layer (and up to the next layer, at the receiving end) blocks of data requiring processing for transmission, manipulation, or service. These layers normally attach a characteristic header that contains appropriate information (such as the real network address, block number, and so on). The headers are physically nested, with lower layer headers being outermost and higher layer headers being innermost. It is through the use of these well-defined headers that the protocols between the remote open systems are accomplished. To effectuate the OSI model, the International Organization for Standardization has formulated standards—specifications for how information is coded and passed between communicating partners. Only the protocols need to be implemented by a prospective vendor. Users should note that, when working at a terminal connected to a host or a PC connected to a LAN, all seven layers of the architecture must be employed.

The Reference Model and the service definitions are only structures for discussing the tasks involved in communicating between open systems. The OSI Reference Model is described by document ISO 7498, which was adopted as a standard in 1984.

## LAN Protocol Suites

In a LAN environment, one typically defines Layer 1 and Layer 2 standards specific to local area networks. Layers 1 and 2 are defined by the IEEE 802 standards. With some "internetworking" protocols defined at Layer 3 (such as ISO 8473), and typically some connection-oriented Transport protocols (for instance, FTAM and MHS), sessions can then employ the normal protocol suite up to Layer 7 described above. Prior to the establishment of ISO-based

## Glossary of FDDI Terms

The definitions given here apply to the ANSI FDDI standards. As the final section of the standard (Station Management) is completed, additional definitions may be added to this list.

**Asynchronous**—A class of data transmission service whereby all requests for service contend for a pool of dynamically allocated ring bandwidth and response time.

**Attachment**—A Port or pair of Ports, optionally including an associated optical bypass, that are managed as a functional unit. A dual attachment includes two ports: a Port A and a Port B. A single attachment includes a Port S.

**Bypass**—The capability of a node to optically isolate itself from the FDDI network while maintaining the continuity of the cable plant.

**Capture**—The act of removing a Token from the ring for the purpose of Frame transmission.

**Claim Token**—A process whereby one or more stations bid for the right to initialize the ring

**Code Bit**—The smallest signaling element used by the Physical Layer for transmission on the medium.

**Code Group**—The specific sequence of five code bits representing a DDL symbol.

**Concentrator**—An FDDI node that has additional ports beyond those required for its own attachment to an FDDI network. These additional ports (type M) are for attaching other FDDI nodes (including other concentrators) in a tree topology.

**Connection Management (CMT)**—That portion of the Station Management (SMT) function that controls network insertion, removal, and connection of PHY and MAC entities within a station.

**Counterrotating**—An arrangement whereby two signal paths, one in each direction, exist in a ring topology.

### Dual Attachment

**Concentrator**—A concentrator that offers a dual attachment to the FDDI network and is capable of accommodating a dual (counterrotating) ring.

**Dual Ring (FDDI dual ring)**—A pair of counterrotating logical rings.

**Entity**—An active service or management element within an Open Systems Interconnection (OSI) layer or sub-layer.

**Fiber Optic Cable**—A cable containing one or more optical fibers.

**Frame**—A PDU transmitted between cooperating MAC entities on a ring, consisting of a variable number of octets and control symbols.

**Jitter, Random**—The probabilistic offsets of pulse transition edges from the expected time. Includes both Duty Cycle Distortion and Data Dependent Jitter.

**Jitter, Systematic**—The deterministic offsets of pulse transition edges from the expected time. Some sources of systematic jitter are differences in rise and fall times and propagation delays.

**Logical Ring**—The set of MACs serially connected to

form a single ring. A fault-free FDDI network provides two logical rings.

**Media Access Control (MAC)**—The Data Link Layer responsible for scheduling and routing data transmissions on a shared-medium local area network (e.g., an FDDI ring).

**Media Interface Connector (MIC)**—A mated connector pair that provides an attachment between an FDDI node and a cable plant. The MIC consists of two parts: an MIC plug and an MIC receptacle.

**MIC Plug**—The male part of the MIC which terminates a fiber optic cable.

**MIC Receptacle**—The female part of the MIC which is contained in an FDDI node.

**Network (FDDI Network)**—A collection of FDDI nodes interconnected to form a trunk, tree, or a trunk with multiple trees. This topology is sometimes called a dual ring of trees.

**Node**—A generic term applying to an active element in an FDDI network (station or concentrator).

**NRZ**—Non Return to Zero, a technique where a polarity level (+ or -) represents a logical "1" (one) or "0" (zero).

standards, however, the use of Transmission Control Protocol/Internet Protocol (TCP/IP) was common for the upper layers. Below, one finds descriptions of the standards at Layer 1 and Layer 2, which are specific to LANs.

### Logical Link Control

IEEE Standard 802.2-1985 (ISO 8802/2) describes the peer-to-peer protocol procedures for the transfer of information and control between any pair of Data Link Layer Service Access Points (SAPs) on a local area network.

### Medium Access Control

#### Carrier Sense Multiple Access with Collision Detection (CSMA/CD)

IEEE Standard 802.3-1985 (ISO 8802/3) provides a medium access method by which two or more stations share a common-bus transmission medium. The standard applies to several media types and provides the necessary specifications for a 10M bps baseband local area network.

### Token-Passing Bus Access Method

IEEE Standard 802.4-1985 (ISO 8802/4) deals with all elements of the token-passing bus access method and its associated physical signaling and media technologies. The goal is to achieve compatible interconnection of stations in a LAN. The access method coordinates the use of the shared medium among the attached stations. It specifies the electrical and physical characteristics of the transmission medium, the electrical signaling used, the frame formats of the transmitted data, the actions of a station upon receipt of a data frame, and the services provided at the conceptual interface between the Medium Access Control sublayer and the Logical Link Control sublayer above it.

### Token-Passing Ring Access Method

IEEE Standard 802.5-1985 (ISO 8802/5) specifies the formats and protocols used by the token-passing ring medium at the control sublayer and physical layer. It also specifies

**NRZ1**—Non Return to Zero Invert on Ones, a technique where a polarity transition represents a logical "1" (one). The absence of a polarity transition denotes a logical "0" (zero).

**Physical Connection**—The full-duplex physical layer association between adjacent PHY entities (in concentrators or stations) in an FDDI network, i.e., a pair of Physical Links.

**Physical Layer (PHY)**—The Physical Layer responsible for delivering a symbol stream produced by an upstream MAC Transmitter to the logically adjacent downstream MAC Receiver in an FDDI ring.

**Physical Link**—The simplex path (via PMD and attached medium) from the transmit function of one PHY entity to the receive function of an adjacent PHY entity (in concentrators or stations) in an FDDI network.

**Port**—A PHY entity and a PMD entity in a node, together creating a PHY/PMD pair, that may connect to the fiber media and provide one end of a physical connection with another node.

**Primitive**—An element of the services provided by one entity to another.

**Protocol Data Unit (PDU)**—Information delivered as a unit between peer entities which may contain control information, address information, and data (e.g., a Service Data Unit from a higher layer).

**Receive**—The action in a station in accepting a Token, frame, or other signal sequence from the incoming medium.

**Receiver (optical)**—An electro-optical circuit that converts an optical signal to an electrical logic signal.

**Repeat**—The action of a station in receiving a Token or Frame from the adjacent upstream station and simultaneously sending to the adjacent downstream station. The FDDI MAC may repeat received PDUs (Tokens and Frames) but does not repeat the received signal stream between PDUs. While repeating a Frame, MAC may copy the data contents and modify the control indicators as appropriate.

**Repeater**—A Physical Layer relay in an FDDI network.

**Ring**—A set of stations wherein information is passed sequentially between stations, each station in turn examining or copying the information, finally returning it to the originating station. In

FDDI usage, the term "ring" or "FDDI ring" refers to a dual (countertrotating) ring.

**Service Data Unit (SDU)**—The unit of data transfer between a service user and a service provider. The MAC SDU is the data contents of a Frame. The PHY SDU is a symbol.

**Services**—The services provided by one entity to another. Data services are provided by a higher level entity; management services are provided to a management entity in the same or another level.

**Single Attachment Concentrator**—A concentrator that offers a single attachment to the FDDI network.

**Single Attachment Station**—A station that offers a single attachment to the FDDI network.

**Station**—An addressable logical and physical node on a ring capable of transmitting, repeating, and receiving information. A station has exactly one SMT, at least one MAC, at least one PHY, and at least one PMD entity.

**Station Management (SMT)**—The supervisory entity within an FDDI station that monitors and controls the various FDDI entities including PMD, MAC, and PHY.

**Symbol**—The smallest signaling element use by MAC, i.e., the PHY SDU. The symbol set consists of 16 data symbols and 8 control symbols. Each maps to a specific sequence of five code bits as transmitted by the Physical Layer.

**Synchronous**—A class of data transmission service whereby each requestor is preallocated a maximum bandwidth and guaranteed a response time not to exceed a specific delay.

**Token**—An explicit indication of the right to transmit on a shared medium. On a token-ring, the Token circulates sequentially through the stations on the ring. At any time it may be held by zero or one stations. FDDI uses two classes of Tokens: restricted and unrestricted.

**Transmit**—The action of a station in generating a Token, Frame, or other symbol sequence and placing it on the outgoing medium.

**Transmitter (optical)**—An opto-electrical circuit that converts an electrical logic signal into an optical signal.

the means of attachment for the Token-Passing Ring Access Method. The protocol defines the frame format including delimiters, addressing, and frame check sequence and includes timers, frame counts, and priority stacks. It also defines the medium access control protocol and provides finite-state machines and state tables supplemented with descriptions of the algorithms. It identifies the services provided by the Medium Access Control sublayer to the Logical Link Control sublayer and the services provided by the physical layer to the Medium Access Control sublayer. These services are defined in terms of service primitives and associated parameters. Also defined are the physical layer functions of symbol encoding and decoding, symbol timing and latency buffering, and the 1M bps and 4M bps, shielded twisted-pair attachments of the station to the medium.

Figure 5, modeled after documentation of the German Commission for Computer-Integrated Manufacturing, depicts typical OSI protocol suites for LANs, particularly in a MAP/TOP environment. Figure 6 depicts some additional details at the lower layers, including the positioning of

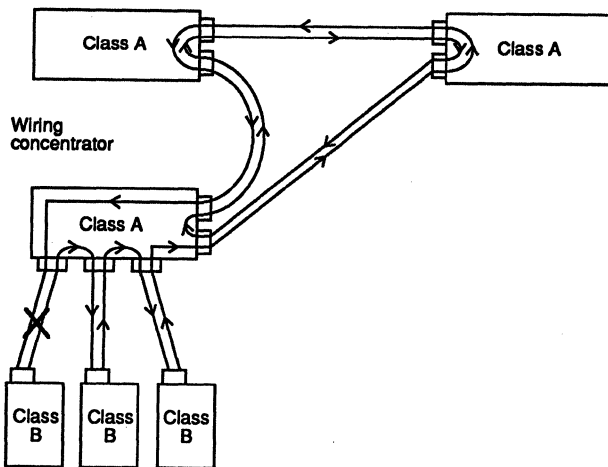
FDDI as it relates to other LAN standards. Once again, note that when a user works at a PC connected to a LAN or FDDI ring, all seven layers of the protocol architecture must be employed, as shown in Figure 5. Thus, while the FDDI standard concentrates on Layers 1 and 2, the upper layers are required to accomplish end-to-end communications.

## FDDI Specifications

FDDI is defined according to the OSI Reference Model and LAN protocol architecture. Layer 1 (physical layer) is specified in two documents: the FDDI Physical Medium Dependent (PMD) and the FDDI Physical Sublayer (PHY). (See Figure 7.) The Physical Layer provides the medium, connectors, optical bypassing, and driver/receiver requirements. It also defines encode/decode and clock requirements for framing data for transmission on the medium or to the higher layers of FDDI.

When the FDDI committee realized there would be considerable discussion on fibers, connectors, and other

Figure 3.  
Class B Failure



Failure of Class B stations.

hardware, they decided to break the standardization of the OSI Physical Layer into two pieces. In this way, the relatively noncontroversial issues—like coding and other matters that IC chip manufacturers need to know to begin design—could be put in a formal document and approved independently of other items pertaining to the standard.

The Data Link Layer is also divided into two sublayers:

1. A Media Access Control portion that provides fair and deterministic access to the medium, address recognition, and generation and verification of frame check sequences. Its primary function is the delivery of frames, including frame insertion, repetition, and removal.
2. A Logical Link Control portion that provides a common protocol for data assurance services between the Media Access Control and the Network Layer.

## Physical Medium Dependent (PMD) Specification

PMD defines the optical interconnecting components used to form a link. It describes the wavelengths for optical transmission, the fiber optic connector, the functions of the optical receiver, and (as an option) the bypass switch that can be incorporated into the station. It specifies the optical channel at the bulkhead of a station. The source is defined to radiate in the 1,300-nm. wavelength. PMD also describes the peak optimal power, optical rise and fall times, and jitter constraints. The minimum rise/fall time is 0.5 nanosecond. The standard includes the following specifications:

1. Services: PMD to PHY Services; PMD to SMT Services
2. Media attachment
3. Media Signal Interface
4. Interface signals
5. Cable Plant Interface Specification.

Multimode fibers are employed (at least initially) up to a distance of two km. Optical fiber dimensions are specified

in terms of its core diameter and the outer diameter of the cladding layer. Fiber specifications are 62.5/125 micron (core diameter/cladding diameter) and 85/125 micron. The nominal numerical aperture is around 0.26. Applicable standards for the fiber itself are EIA-455-48 (core), EIA-455-27 or EIA-455-48 (cladding), and EIA-455-57 (aperture).

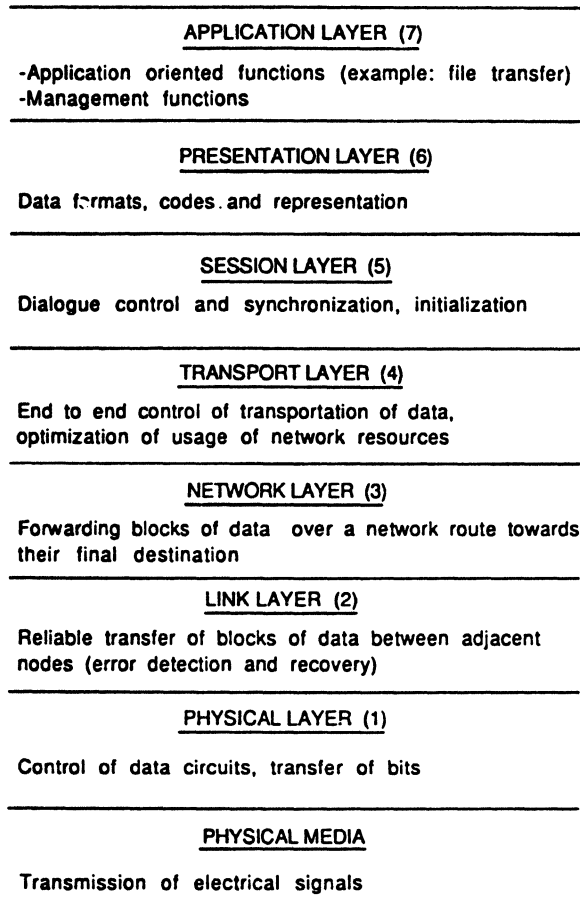
Listed in the appendix of the draft standard are two other fibers: a 50-/125-micron fiber and a 100-/140-micron fiber. These two fibers have not been extensively studied; the maximum achievable distance of two km. specified in the standard may not be possible. Thus, these two fibers are not officially part of the standard but are listed as "alternatives." Smaller diameters offer higher bandwidths but also more expensive, higher loss connectors. The 50/125 fiber has been used primarily for military applications and is not likely to be as widely available as the other fibers. The 100/140 fiber has been added primarily because it is used in IBM's cabling system and a number of customers have already installed it. On the other hand, the 62.5/125 fiber has been in production for some time and has become common for local applications, such as AT&T's Premises Distribution System (PDS). Component costs for this type of fiber are dropping most rapidly.

Some observers believe it is unfortunate that the FDDI standard could not specify a single fiber type, since this might have lowered costs and made it easier for the customer to start small and expand later. The FDDI committee settled on a 62.5-micrometer core, with advisory information about 50-, 85-, and 100-micrometer fiber sizes. While a debate remains about these other sizes, the issue is not really a critical one: as long as the fiber can meet the optical power, channel bandwidth, and distance requirements, it can conform to the FDDI standard at the interface between the FDDI box and the network, independent of the fiber type (the charts and data presentations in the PMD have been written so that they can be applied to various fiber sizes).

Like all Layer 1 specifications, PMD defines the duplex connector used for FDDI access. The primary and secondary ring connections to each Class A station are attached simultaneously using the duplex connector and a dual-fiber cable. The connectors can be used for both Class A-to-Class A as well as Class B-to-Class A (wiring concentrator) links. The bypass relay connects the optical inputs (at the primary and secondary rings) directly to the optical output in case of a station or link failure, allowing the ring to maintain continuity.

In the 1,300-nm. region, the dispersion due to multimode interference is at a minimum. The combination of physical parameters selected ensure the desired  $10^{-9}$  bit error rate. LEDs (either surface emitting or edge emitting) are implicitly assumed in PMD; however, PMD does not specify the emitter must be an LED. It could also be a laser, as long as the optical interface parameters at the optical port are met. At some future point, some manufacturers may include lower cost local-loop-grade laser emitters, or even long-haul-grade LDs in an FDDI package, by adjusting the optical output at the optical port to conform with the standard. For the foreseeable future, however, all manufacturers pursuing FDDI products will use LEDs. At least one vendor, Codenoll, has elected to use an 850-nm. LED. While this wavelength does not meet the standard, Codenoll claims that the only effect of the change is to shorten the maximum allowable distance between stations from 2,000 meters to 1,000 meters. Because Codenoll's products are microcomputer adapter boards, the distance limitation

Figure 4.  
OSI Functions



*Functions performed by the layers of the OSI Reference Model.*

is less significant than it is in internetworking or main-frame attachment environments. Such nonstandard implementations, however, cannot communicate with devices using standard 1,300-nm. components, requiring users to install a nonstandard device at both ends of the link.

### Physical Sublayer Specification

PHY represents the upper sublayer within OSI Layer 1. It defines the encoding scheme used to represent data and control symbols. It also describes the method for retiming transmission within the node. The standard includes the following specifications:

1. Services
  - PHY to MAC Services
  - PHY to PMD Services
  - PHY to SMT Services
2. Facilities
  - Coding
  - Symbol Set
  - Line States

Digital data must be encoded in some form for proper transmission. The type of encoding depends on the nature of the transmission medium, the data rate, and other factors such as noise, reliability, and cost. Intensity modulation is the normal method of representing digital data for transmission over fiber: a binary 1 is represented by a pulse of light and a binary 0 by absence of optical power. The disadvantage of using this method, in its simplest form, is its lack of synchronization. Long strings of ones or zeroes create a situation where the receiver is unable to synchronize its clock to that of the transmitter. The solution is to first encode the binary data in such a way as to guarantee the presence of signal transitions, even if there are no transitions in the incoming digital signal; after this encoding is performed, the signal can be presented to the optical source for transmission using intensity modulation. A typical encoding scheme is Manchester encoding (see Figure 8).

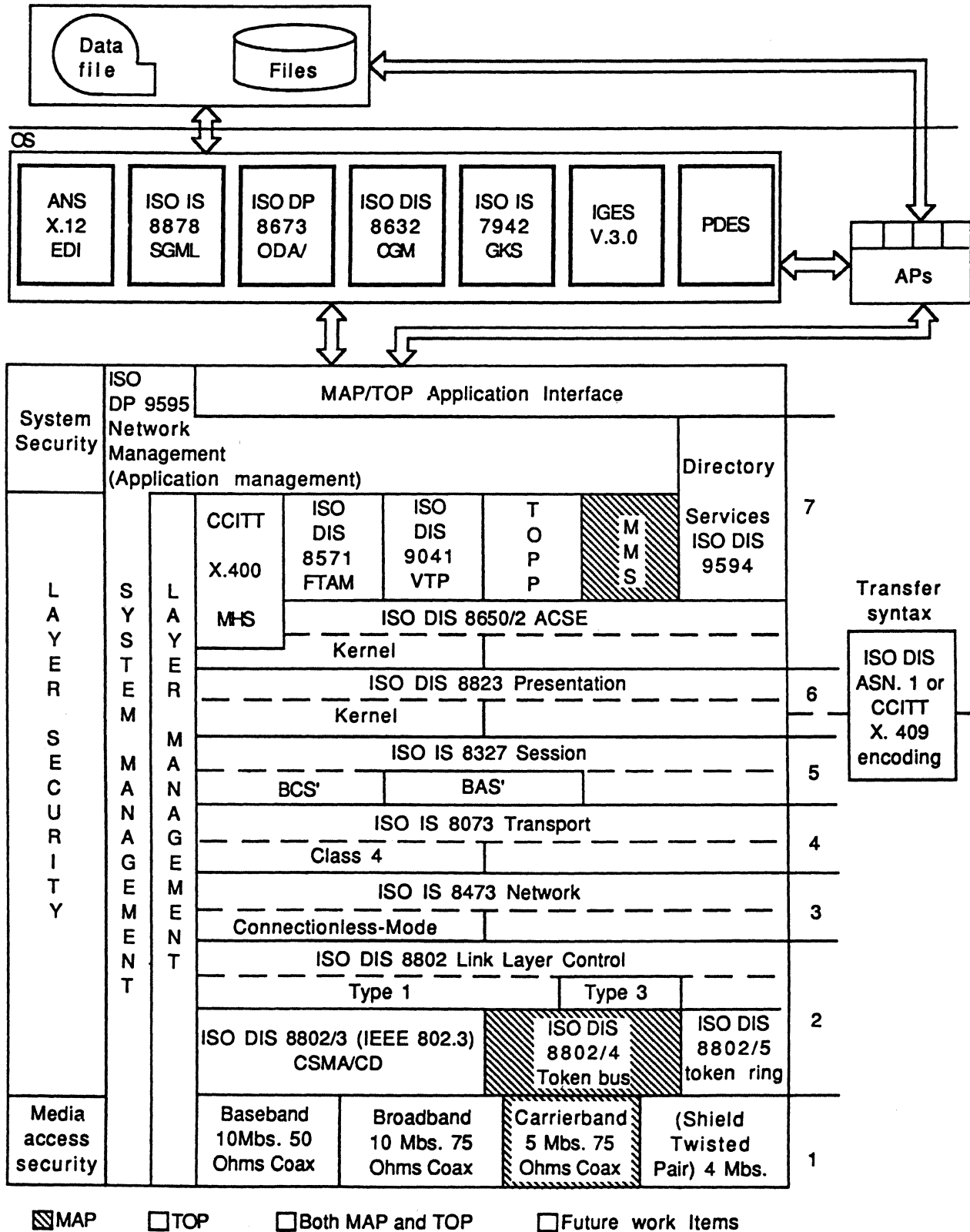
Differential Manchester is only 50% efficient since each data bit is represented by transitions in signal. Two transitions allow a degree of robustness in the presence of noise, as would be the case in coaxial cable. Since fiber is less susceptible to noise, two transitions are not required to identify a bit with a good degree of confidence. To avoid having to use a 200MHz signal, FDDI specifies a code referred to as 4B/5B group encoding. The result is that the 100M bps throughput is achieved in FDDI with a 125MHz rate, rather than the 200MHz rate needed in differential Manchester. This helps keep down the cost and complexity of equipment. One drawback of the group encoding pertains to clock recovery. Since differential Manchester has more pulses in its stream, it is easy to extract the clock in that scheme. One of the key responsibilities of this FDDI sublayer is to decode the 4B/5B nonreturn to zero inverted (NRZI) signal from the network into symbols that can be recognized by the station, and vice versa.

The synchronization clock is derived from the incoming signal. The data is then retimed to an internal clock through an elasticity buffer. In this scheme, four bits of data are translated into a five-baud value transmitted over the network, giving an 80% efficiency factor. This group-encoding scheme employed in FDDI is a departure from differential Manchester codes normally specified in LAN standards.

To understand how the FDDI scheme achieves synchronization, one must realize that there are two stages of encoding. In 4B/5B, the encoding is performed four bits at a time. Each four bits of data are encoded into a symbol with five cells such that each cell contains a single signal element (presence or absence of light). In effect, each set of four bits is encoded as five bits. Then, each element of the 4B/5B stream is treated as a binary value and encoded using NRZI. In this code, a binary 1 is represented with a transition at the beginning of the bit interval; there are no other transitions. The advantage of NRZI is that it employs differential encoding: the signal is decoded by comparing the polarity of adjacent signal elements rather than the absolute value of a signal element. This scheme is relatively robust in detecting transitions in the presence of noise or other distortions; therefore, the NRZI encoding will improve reception reliability. Table 1 shows the symbol encoding used in FDDI.

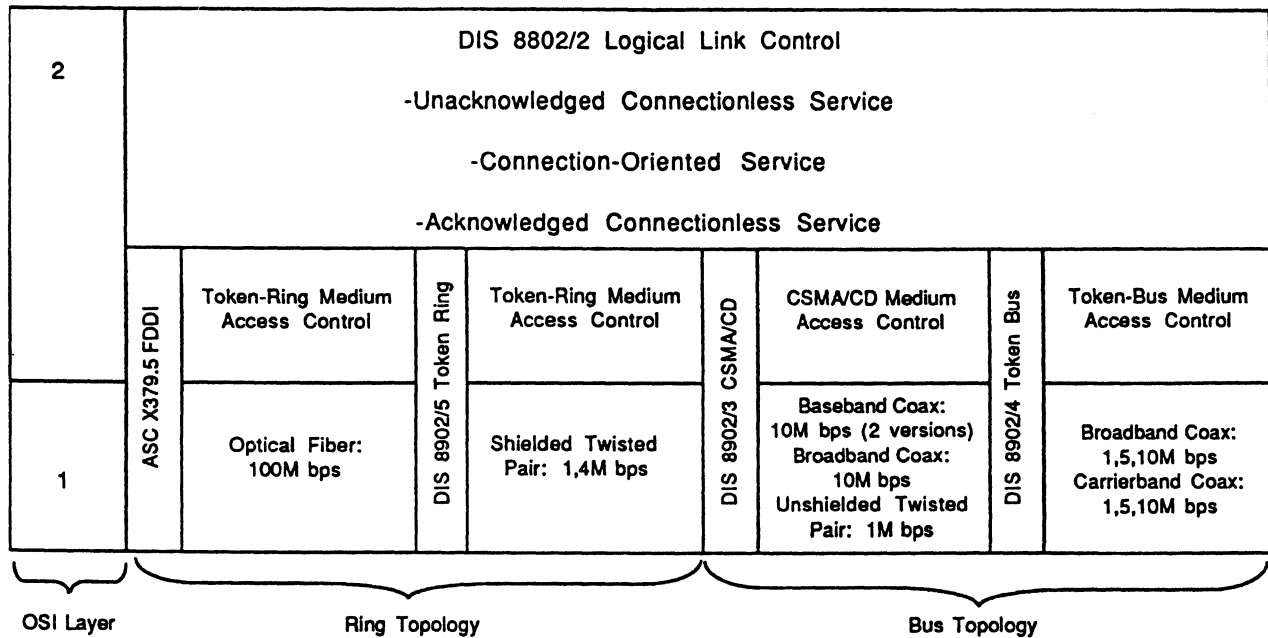
Since this scheme is encoding 4 bits (16 combinations) with 5 bit patterns (32 combinations), there will be patterns that are not needed. The codes selected to represent the sixteen four-bit patterns are such that a transition is present at least twice for each five-bit code. Given an

Figure 5.  
ISO Suites



ISO suites in a LAN environment.

Figure 6.  
ISO Lower Layers



Lower layers in a LAN environment.

NRZI format, no more than three zeroes in a row can be allowed, since the absence of a transition indicates a zero. The remaining symbols are either declared invalid or are assigned special meaning as control symbols as shown in Table 1.

PHY also provides line states for establishing the station's links with its neighbors (upstream and downstream) and to detect the integrity of the station's links to these neighbors. These line states are used to exchange a handshake with a neighbor. A node receiving a line state on its primary input can respond by sending the proper line state on the secondary output. The line states are composed of a repetition of one or more "I" symbols.

Another item that must be resolved in this sublayer is the issue of timing jitter. Jitter is the deviation of clock recovery that can occur when the receiver attempts to recover clocking as well as data from the received signal. The clock recovery will deviate in a random fashion from the transitions of the received signal. If no countermeasures are used, the jitter will accumulate around the ring. In LANs, the IEEE 802 standard specifies that only one master clock will be used on the ring and that the station with the clock will be responsible for eliminating jitter using an elastic buffer. If the ring as a whole runs ahead or behind the master clock, the elastic buffer expands or contracts accordingly. This centralized clocking method, however, is not practical for a 100M bps ring. At this speed, the bit time is only 10 ns, compared to a bit time of 250 ns. at 4M bps, making the effect of distortion more severe. Consequently, FDDI specifies the distributed clocking scheme.

In this environment, each station uses its own autonomous clock to transmit or repeat information onto the ring. Each station has an elastic buffer where data is clocked in at the clock rate recovered from the incoming stream, but it is clocked out at the station's own clock rate. This distributed system is considered stronger than the

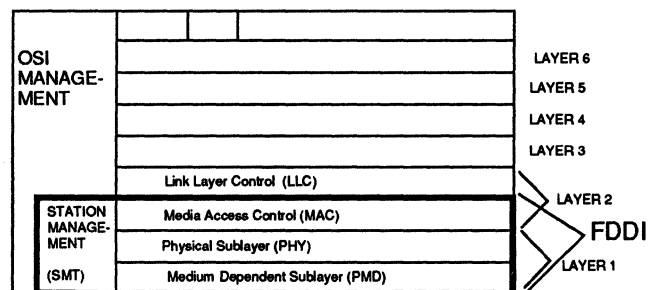
centralized method and minimizes jitter. As a consequence of reclocking at each station, jitter does not limit the number of repeaters in the ring, as is the case in LANs where a master clock is used.

**Media Access Control Specification**

Layer 2 (Data Link Layer) of the OSI Reference Model is traditionally divided into two sublayers in a LAN context: Link Layer Control (LLC) and Media Access Control (MAC). FDDI only defines MAC, which controls data flow over the ring. The token-passing protocol incorporated in FDDI controls transmission over the network. MAC defines packet formation (headers, trailers, etc.), addressing, and cyclic redundancy checking (CRC). It also defines the recovery mechanisms. This standard defines the following specifications:

1. Services:
  - MAC to LLC Services
  - PHY to MAC Services
  - MAC to SMT Services

Figure 7.  
FDDI Protocols



## 2. Facilities:

- Symbol Set
- Protocol Data Units
- Fields
- Timers
- Frame Counts
- Frame Check Sum

The FDDI packet format is shown in Figure 9. Packets are preceded by a minimum of 16 IDLE control symbols. The packet itself is characterized by a Start Delimiter composed of the J and K control symbols. This is followed by a Frame Control field that identifies the type of packet. The Destination Address, which follows, identifies the frame recipient. The Source Address is also included to identify which station originated the packet. The address field can be 26 or 48 bits in length. The variable information field follows, along with a Frame Check Sequence field of 32 bits. The check sequence covers the Frame Control Field, the two addresses, and the information field. An End Delimiter, which consists of the T symbol, is transmitted. The maximum packet length is limited by the size of the elastic buffer in the Physical Sublayer and by the worst case frequency difference between two nodes, the upper bound in 9,000 octets. Figure 9 also shows the format of the token.

Flow control is the other major function of the MAC. In an idle condition, MAC connects to an internal source of IDLE control symbols to be transmitted over the ring. When a Start Delimiter is detected from the ring, MAC switches to a repeat path; the packet is monitored and copied if it is meant for this destination. The packet is simultaneously repeated onto the ring for relaying. The MAC can also inject its own packet or issue a token. Packets are removed only by the originating station. The MAC repeats the packet only until the Sender Address field is detected. If the destination recognizes the Sender Address field as its own station, it will insert IDLE control symbols back onto

the ring (the fragmented packet is ignored and removed by any station holding a token for transmission). Stations that wish to transmit must first obtain a token (this is the unique six-symbol packet shown in Figure 9).

The procedures for obtaining the token and the amount of time allowed for data transmission (to retain fairness) are specified in the Timed Token Protocol (TTP). A station obtains the token by performing the stripping function on the incoming token. Only the Start Delimiter field is repeated onto the ring; the station will inject its own information at this juncture. When the packet is sent, the station immediately issues a new token. TTP guarantees a maximum token rotation time. TTP allows two types of transmission: synchronous and asynchronous. In the synchronous mode, stations obtain a predefined amount of transmission bandwidth on each token rotation. The balance of the bandwidth is shared among stations using the asynchronous service. These stations can send data when the token arrives earlier than expected. Any unused capacity left over from synchronous capacity is available to asynchronous traffic, which may be subdivided into up to eight levels of priority. The amount of time allowed for asynchronous transmission is bounded by the difference of the token's actual arrival time and the expected arrival time. In essence, each station keeps track of how long it has been since it last saw the token. When it next sees the token, it can send synchronous traffic and/or any asynchronous traffic for which time remains available.

**Station Management (SMT)**

The FDDI Station Management (SMT) specification describes software-based, low-level data link management and integrated network control functions of all stations attached to an FDDI LAN and of the LAN itself. Each FDDI station contains only one SMT entity. SMT initializes the network, monitors error rates and fault conditions in each network segment, and automatically reconfigures the network to isolate problem links. SMT components are Connection Management (CMT), which includes Entity Coordination Management, Physical Connection Management, (PCM), and Ring Management (RMT). SMT is intended to operate regardless of equipment type, vendor, protocols, or applications. Figure 10 presents the SMT architectural model.

SMT types (managed objects) have specific attributes indicating state, capabilities, and operation.

SMT managed objects are:

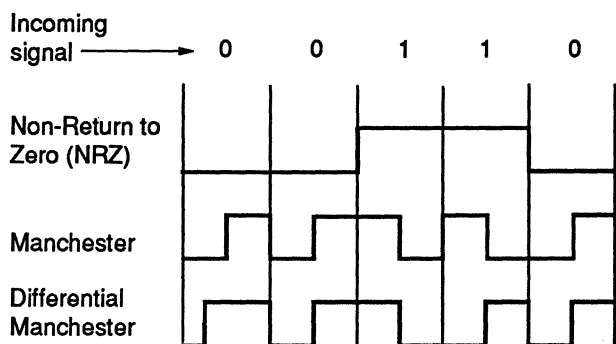
- Station or concentrator (SMT)
- MAC object(s)
- Path object(s)
- PHY object(s)
- PMD object(s)
- Attachment(s)

Attributes are:

- Attribute Identification (ID)
- Configuration
- Operational (Status, Counters, etc.)

Each attribute is defined in terms of Access Rights and whether it is Mandatory or Optional. Each attribute also carries an FDDI specification reference or a specific definition if it is not defined in the FDDI specification.

Figure 8.  
Comparison Coding Schemes



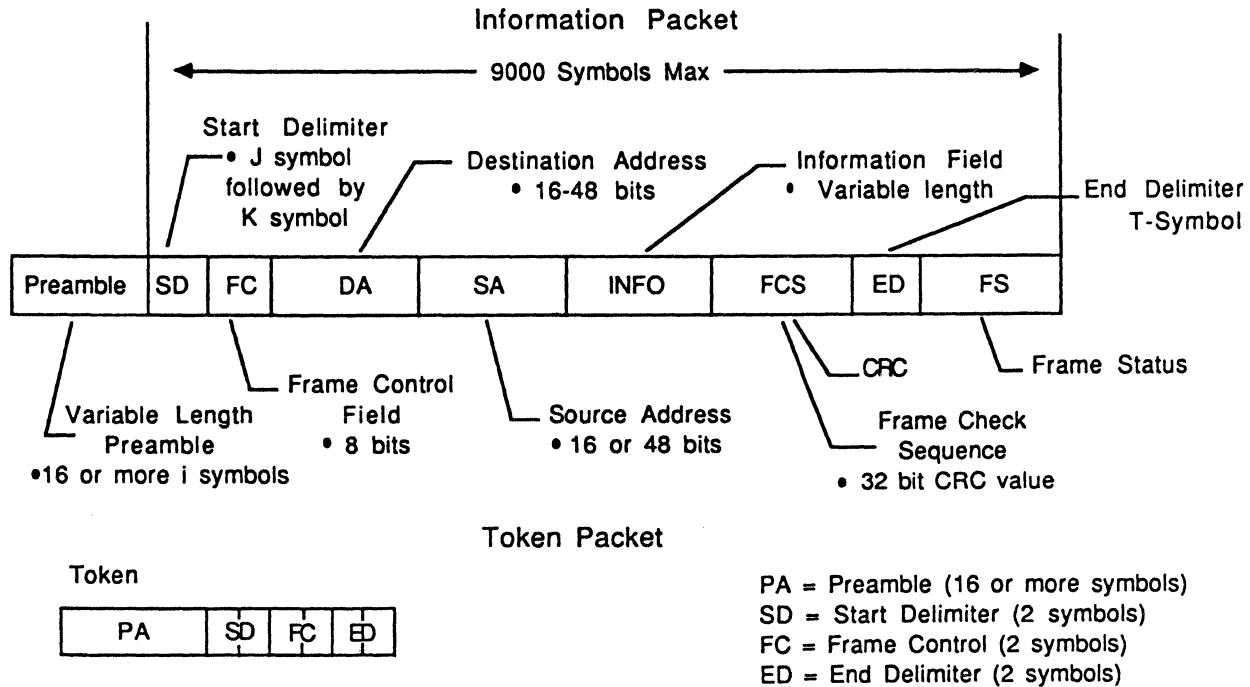
In Manchester code, there is a transition at the middle of each bit period. The mid-bit transition serves as a clock and also as data. A high-to-low transition represents a 1, and a low-to-high transition represents a 0.

In Differential Manchester code, the mid-bit transition is used only to provide clocking. The coding of a 0 (1) is represented by the presence (absence) of a transition at the beginning of the bit period.

Comparison coding schemes used in LANs and/or fiber.



**Figure 9.**  
**FDDI Packets**



**FDDI packet format.**

SMT Connection Management (CMT) operates at the logical level, controlling the interface between PHY and MAC entities in a given station and controlling SMT-to-SMT communications across the ring. When a session requires a connection to another station, CMT causes PHY to send a stream of symbols to the targeted station. Upon receiving the symbols, the receiving station's PHY returns a continuous stream of symbols (primitives) indicating line state, station status, and willingness to carry out the requested action (establish a link). QUIET symbols indicates a disabled link. HALT or alternating HALT and QUIET (MASTER) symbols indicate an operating link and the receiving station's status (master, slave, or peer). A stream of IDLE symbols indicates willingness to connect. Once the link is established, CMT configures the PHY and MAC.

Entity Coordination Management (ECM) controls the optional optical bypass switch and signals the Physical Connection Management (PCM) entity when the bypass is complete. ECM also performs the Path Test to determine a fault's location.

Physical Connection Management (PCM) initializes adjacent stations' PHYs and manages signaling. Maintenance support functions are also part of PCM.

Configuration Management (CFM) interconnects PHYs and MACs. It automatically configures these connections according to PCM flags. CFM is defined differently for stations and concentrators.

Ring Management (RMT) relays MAC and CFM status information. It detects stuck signaling beacons, initiates the trace function, detects duplicate addresses and resolves them to allow continued ring operation, and notifies SMT of MAC status.

The draft SMT standard was forwarded out of the X3T9.5 committee in April 1990. Letter ballots returned

in midsummer 1990 included comments that required some additional work. The draft is now technically stable.

**Higher Layers**

From LLC upward, ANSI intends FDDI to generally fit traditional protocol stacks. The ANSI FDDI committee has not yet formally drafted protocols for Layers 3 (network) and 4 (transport), which are needed for any type of internetworking. Many suppliers believe the TCP/IP protocol suite can serve this purpose. TCP/IP software was originally developed by the U.S. government for Arpanet, the worldwide packet switched network, but has been used with great success in commercial applications—particularly for internetworking among different LANs. TCP and IP follow layered networking concepts, occupying Layers 4 and 3, respectively, of the OSI Model. Vendors have adopted the protocols for Ethernet and other local area networks. Most commercial implementations of the TCP/IP protocol suite include three standardized upper layer protocols: Telnet (virtual network terminal), File Transfer Protocol (FTP), and Simple Mail Transfer Protocol (SMTP).

Although more versatile than TCP/IP, the OSI internetworking protocols—spanning Layers 3 through 5—are less practical to implement in the real world. Much effort, however, has been expended to migrate to ISO-based standards, particularly under the MAP/TOP thrust, as shown in Figure 5. These market demands are forcing investigations into upgrading the third and fourth layers. The OSI stack also has a well-developed suite of upper layer applications, including X.400 (electronic mail), File Transfer

**Table 1. 4B/5B Codes Used in FDDI**

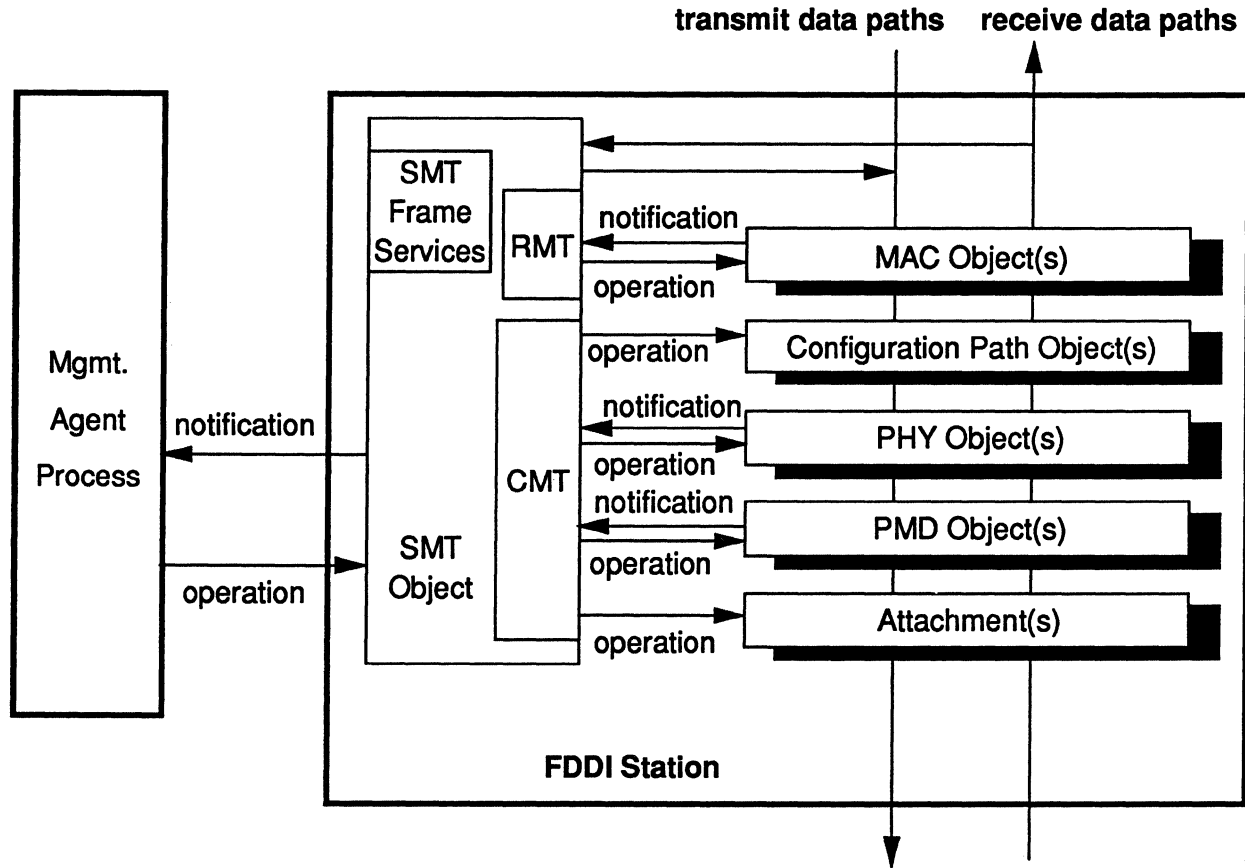
Function or 4-bit group (4B)	Group Code (5B)	Symbol
<b>Starting Delimiter</b>		
First symbol of sequential SD pair	11000	J
Second symbol of sequential SD pair	10001	K
<b>Data Symbols</b>		
0000	11110	0
0001	01001	1
0010	10100	2
0011	10101	3
0100	01010	4
0101	01011	5
0110	01110	6
0111	01111	7
1000	10010	8
1001	10011	9
1010	10110	A
1011	10111	B
1100	11010	C
1101	11011	D
1110	11100	E
1111	11101	F
<b>Ending Delimiter</b>		
Used to terminate datastream	01101	T
<b>Control Indicators</b>		
Logical Zero (reset)	00111	R
Logical One (set)	11001	S
<b>Line Status Symbols</b>		
Quiet	00000	Q
Idle	11111	I
Halt	00100	H
<b>Invalid Code Assignment</b>		
These patterns shall not be transmitted because they violate consecutive code-bit zeroes or duty cycle requirements. Some of the codes shown shall nonetheless be interpreted as a Halt if received.	00001	Void or Halt
	00010	Void or Halt
	00011	Void
	00101	Void
	00110	Void
	01000	Void or Halt
	01100	Void
	10000	Void or Halt

Access Management (FTAM), and X.500 Directory System Protocol. Moreover, the federal government has mandated its own version of the OSI Reference Model—called GOSIP—must replace TCP/IP in government procurements after August 1990.

Some believe that middle-layer OSI protocols will be changed to look more like TCP/IP. In today's commercial networking applications, however, vendors are blending different protocol stacks from various sources to match

user needs. One vendor's network protocol, for instance, might blend different layers from OSI, TCP/IP, or IBM's SNA. In reality, networking protocols are still evolving, and prospective users must evaluate them with an eye toward future standards.

Figure 10. SMT Architectural Model



The relationships between managed objects in an FDDI station and the SMT object.

**Open Issues**

In May 1991, five networking equipment and semiconductor manufacturers announced that they had joined together to define and publish an open, interoperable solution for transmitting FDDI over shielded twisted-pair (STP) cabling. Advanced Micro Devices Inc., Chipcom Corp., Digital Equipment Corp., Motorola, and SynOptics Communications published the proposed specification in hopes that additional companies would adhere to it—helping to lower the cost of the needed components—thus accelerating the installation of FDDI to the desktop. The vendors’ rationale for developing FDDI-over-STP technology is to allow users to employ existing cabling when migrating from existing LANs to FDDI without installing fiber optic cabling.

The first ANSI-chartered Twisted Pair-Physical Medium Dependent (TP-PMD) ad hoc meeting was held in August 1990. Chipcom, Digital Equipment, and SynOptics shared the results of their independent research into implementing FDDI over STP. The companies’ proposals were sufficiently similar that the companies agreed to work together to develop an open, interoperable proposal.

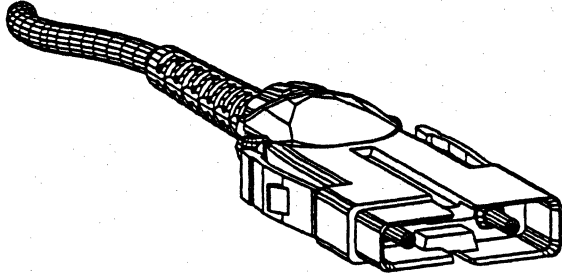
The proposed solution specifies a Physical Medium Dependent sublayer that uses 150-ohm, shielded twisted-pair cables. While the document does not specify a particular design, it does call for the complete replacement of the existing optical PMDs with electrical counterparts. Because all other aspects of the FDDI standard remain the same, no changes are required beyond the PMD level.

As mentioned earlier, Codenoll Technologies introduced a nonstandard version of the FDDI interface board for Extended Industry Standard Architecture (EISA) computers. While using an 850-nm. LED in the Codenoll design does reduce the cost of FDDI connectivity, users will be unable to operate with standard-based equipment containing 1,300-nm. transmitters and receptors.

**FDDI-II and Other Enhancement Efforts**

Work is nearly complete on defining a scheme where available network bandwidth is divided between voice and data using time-division multiplexing. FDDI-II, a superset of FDDI, is being defined as an upward-compatible, fiber-based LAN incorporating the current data capabilities in addition to the capability to handle voice and T1-compressed video traffic. The FDDI-II proposal will follow the original FDDI standard, adding a fifth document to the present standard. The new document, Hybrid Ring Control, describes a hybrid operating mode comprising the packet-switching scheme used in FDDI and an isochronous transport mode similar to that used in public switched networks. Adding the isochronous mode will enable networks based on the expanded standard to transfer pixel data—key to video and computer graphics transport—in addition to circuit switched voice signals. Because it builds on the existing FDDI framework, FDDI II should be completed and approved very quickly. According to Gene Milligan, chairman of the ANSI X3T9.5

Figure 11.  
Media Interface Connector



Example of Media Interface Connector (MIC) plug.

committee, the Hybrid Ring Control document should be an approved standard by the end of 1991.

Some people are already thinking of even higher rates, particularly when considering voice and data. The FDDI follow-on, or FFOL, is under development at ANSI. The FFOL project proposal covers the capability to operate as a backbone for multiple FDDI networks; interconnection to wide area networks, including Broadband Integrated Services Digital Networks (BISDN); a data rate between 600M bps and 1.25G bps initially, with intermediate data rates matched to the Synchronous Digital Hierarchy (SDH) underlying SONET and eventual support for data rates up to 2.4G bps; duplex links; support for existing FDDI cabling, where possible; and support for both single-mode and multimode fiber. If the ANSI effort keeps on schedule, the family of standards comprising FFOL should be complete by late 1995.

There is also a separate effort called High Speed Channel being undertaken in X3T9.3. This is a point-to-point system for digital data interface (channel extension rather than a network configuration intrinsic in FDDI). These systems are designed to carry 400M, 800M, or 1600M bps

on very short copper links, but they are being developed to be compatible with future fiber optic links operating over longer distances.

### FDDI Information Sources

The chairperson of the ASC X3T9.5 task group is Gene Milligan of Seagate (Oklahoma City, OK).

ANSI is involved in the form of the X3 Secretariat, as is CBEMA (Computer Business Equipment Manufacturers Association).

Specification X3.139, X3.148, and X3.166 can be obtained from Global Engineering Documents.

Draft Specification X3T9.5/84-49 is available from ANSI for review.

#### ANSI

1430 Broadway  
New York, NY, 10018 (212) 642-4900.

#### CBEMA

311 First Street, Suite 500  
Washington D.C. (202) 737-8888.

#### Global Engineering Documents

2805 McGraw Avenue, P.O. Box 19539  
Irvine, CA 92714 (714) 261-1455

Information regarding the FDDI/STP effort is available from:

#### Advanced Micro Devices Inc.

901 Thompson Place  
Sunnyvale, CA 94088 (408) 982-7880

#### Chipcom Corp.

118 Turnpike Road  
Southborough, MA 01772 (508) 460-8900

#### Digital Equipment Corp. (DEC)

550 King Street  
Littleton, MA 01460 (508) 486-5096

#### Motorola

6501 William Cannon Drive West  
Austin, TX 78735 (512) 891-2140

#### SynOptics Communications

4401 Great America Parkway  
Santa Clara, CA 95052 (408) 764-1013 ■

---

# Introduction to FDDI-II

---

## In this report:

Basic FDDI .....	2
Principles of FDDI-II Operation.....	3
FDDI-II Services .....	3
Hybrid-Ring Operation.....	5

## Datapro Summary

Fiber Distributed Data Interface-II (FDDI-II), designed for optical fiber medium, is poised to dramatically extend the functionality of the workstation. Perfectly suited for multimedia networks, FDDI-II supports the concurrent transmission and mutual synchronization of audio, image, video, and other multimedia elements. A suite of FDDI-II products are in the development and testing process; these include multimedia workstations, as well as FDDI-II routers and bridges. With its promise of workstation extensibility, FDDI-II may prove to be the high-performance LAN standard of the 1990s.

Fiber Distributed Data Interface (FDDI), with a data rate of 100 Mbps, has gained worldwide recognition as the high-performance LAN of choice. FDDI-II, a logical evolution of FDDI that has been long in the planning, is now itself on the verge of great success. Eventually, all FDDI deployments will be FDDI-II.

FDDI uses a token-ring access protocol, conforming to the structure and architectural concepts of IEEE 802 LANs. It is based on the use of an optical fiber medium. Standards for FDDI are being developed in X3T9.5, an ANSI-accredited standards committee.

FDDI-II satisfies the vital requirement of multimedia networks: the concurrent transmission and mutual synchronization of data streams such as video, audio, image, and voice on the same medium with conventional packet traffic. Thus, FDDI-II is an enabling technology that will bring many new capabilities to the workstation level.

---

This Datapro report is a reprint of "Get Ready for FDDI-II" by Floyd E. Ross, pp. 54-58, from *Networking Management*, July 1991. Copyright © 1991 by PennWell Publishing Company. Reprinted with permission.

## FDDI-II Perspective

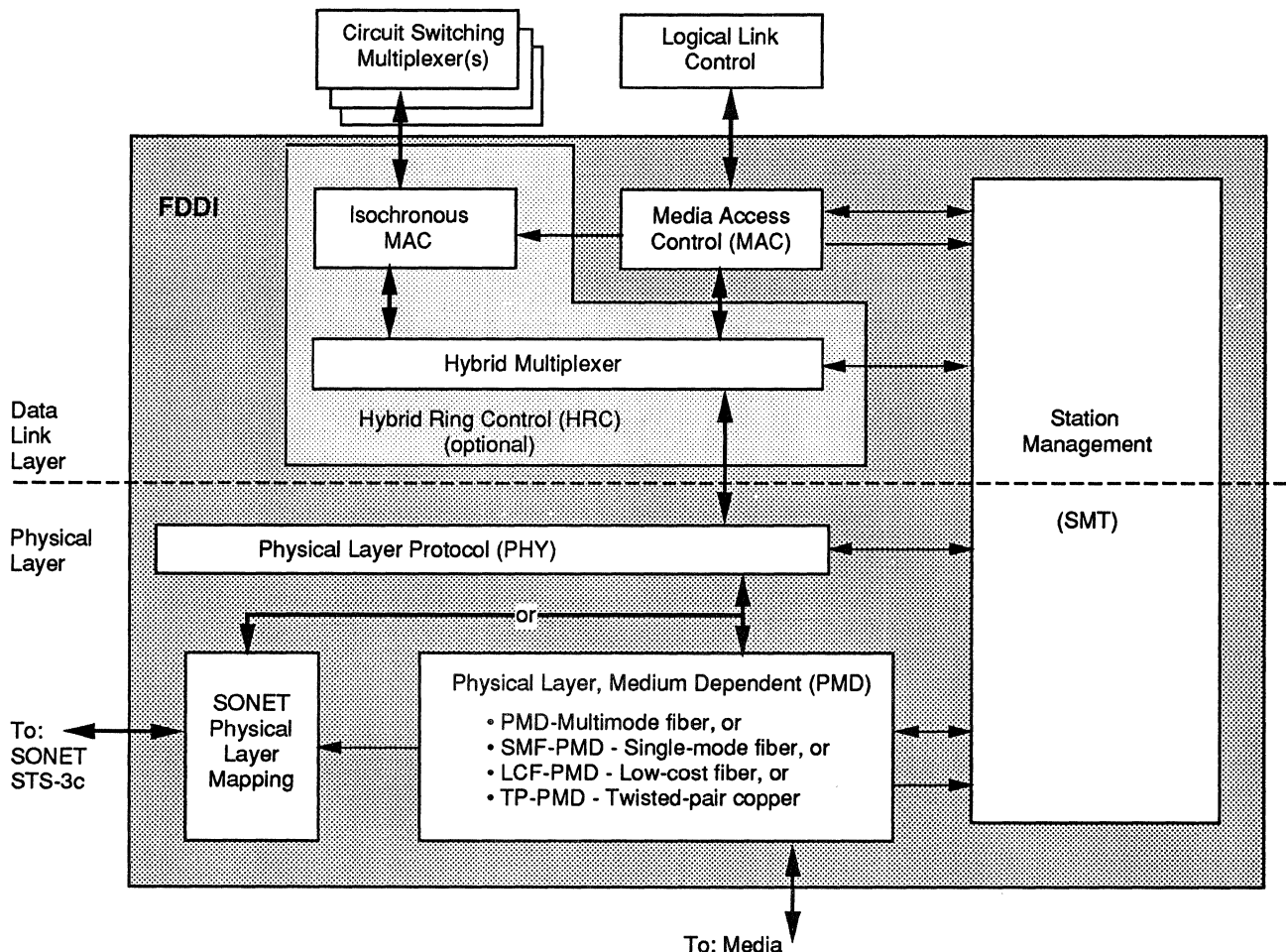
FDDI-II provides a superset of the services now offered by the basic FDDI. Thus, FDDI-II supports all of the current FDDI applications and considerably ends the range of applications that FDDI rings may address.

FDDI standards require that an FDDI-II station be interoperable with stations conforming to the current FDDI standards, thus setting the stage for graceful deployment of FDDI-II. Network managers can also expect to see products specially designed to support a mixed FDDI-II environment, further easing the transition to FDDI-II. Present FDDI equipment will be able to stay in place and perform the tasks for which it was intended.

Work on FDDI-II standards has been driven by a core group of contributors. Extensive evaluation and simulations have been done. And more than one U.S. chip vendor is known to be committed to early development of FDDI-II-capable chip sets. FDDI-II products are well into design, development, and testing. Expect to see initial products by mid-1992.

These products will include the long-touted multimedia workstations and other equipment that can benefit from FDDI-II's capabilities. Along with these will come FDDI-II routers, concentrators, bridges,

Figure 1.  
FDDI Reference Model



FDDI-II requires the addition of a hybrid-ring control (HRC) standard. HRC fits between MAC and PHY and multiplexes data between the packet MAC and the isochronous MAC. HRC requires use of the second-generation MAC and PHY standards. Otherwise, any combination of MAC or MAC-2 and PHY or PHY-2 is allowed.

and supporting communications equipment. The impact of FDDI-II's application to phone and other digital switching premises equipment will be felt across the communications industry.

### Basic FDDI

With basic FDDI, data is transferred in packets, called frames. Packets are recognized by their unique starting and ending delimiters. Each FDDI frame contains an information field from 0 to 4472 octets long. Frames are directed to the station or stations designed in a destination address field contained in the header portion of the frame.

FDDI provides for two kinds of packet traffic: synchronous and asynchronous. Synchronous service requires pre-allocation of the bandwidth needed to the requesting FDDI station. In return, availability of the bandwidth allocated and a response time not greater than twice the established target token rotation time (TTRT) value are guaranteed.

After the synchronous bandwidth has been allocated, the bandwidth remaining is available for asynchronous

frames. The pool of asynchronous bandwidth for asynchronous frames is shared by all stations in a fair and deterministic manner.

### FDDI Standards

The basic set of FDDI standards consists of media access control (MAC), physical layer protocol (PHY), physical layer medium dependent (PMD), and station management (SMT). Figure 1 shows the model for FDDI standards.

An early enhancement to FDDI added an alternative single-mode fiber PMD (SMF-PMD) to accommodate link distances of up to 60 kilometers in lieu of the 2-kilometer maximum allowed by the original PMD. Other planned enhancements include PMD standards based on twisted-pair copper cable and low-cost optical fiber to provide low-cost links between FDDI concentrators and workstations in the office. Another enhancement will provide direct connection to SONET STS-3c, allowing transport of the full band-width directly over a SONET link.

The American National Standards Institute has published standards for MAC, PHY, and PMD. These are ANSI X3.139-1987, X3.148-1988, and X3.166-1990, respectively. The equivalent international standards, which

have also been published, are ISO 9314-2:1989, 9314-1:1989, and 9314-3:1990, respectively. The remaining FDDI standards are being developed.

### The FDDI-II Enhancement

FDDI-II is an upward-compatible enhancement of FDDI that adds a circuit-switched service to the existing packet service of basic FDDI. FDDI-II is best viewed as a bandwidth allocation mechanism. Basic FDDI offers a bandwidth of 100 Mbps for packet traffic. FDDI-II enables portions of this bandwidth to be allocated exclusively for the transmission of circuit-switched (isochronous) data streams. The unit of bandwidth for allocation to isochronous traffic is a wideband channel. This is an allocation of synchronous bandwidth. When allocated, a wideband channel provides a full-duplex 6.144-Mbps data stream, a rate equivalent to four times the North American and three times the European basic access rate to the public network. Up to 16 wideband channels can be allocated, accounting for 98.304 Mbps of the FDDI data capability.

FDDI-II requires one additional standard, the hybrid-ring control (HRC), as shown in the FDDI reference model. HRC multiplexes data between the packet MAC and the isochronous MAC. The use of HRC requires the use of second-generation FDDI MAC and PHY standards, designated MAC-2 and PHY-2.

### Isochronous Data

In contrast to a packet service, a circuit-switched service provides a continuous stream of isochronous data at some fixed data rate between two or more stations. Isochronous data is defined as data that occurs on a regular basis relative to some timing marker and with a specified variation from that timing marker. Instead of using addresses to direct this data, the connection between these stations is established based upon some prior agreement. This agreement may have been negotiated using packet messages or established by any other suitable convention known to, or built into, the stations involved.

This prior agreement typically requires knowing the location of a time slot, or slots, that recur regularly relative to some well-known timing marker. A common timing marker used in North America is the basic system reference frequency, the 125-microsecond clock used by the public networks. In FDDI usage, this clock is referred to as the cycle clock.

The data rate of a circuit-switched service is appropriate to the service being provided. For example, 64 kbps is used for a digital voice data stream. Other data rates, up to many megabits per second in the case of video, are used for other applications. Once a connection is established, the data rate remains constant. A wideband channel concurrently supports any mixture of permitted data rates up to the total 6.144-Mbps bandwidth of the wideband channel.

The difference between packet and circuit-switched data is interesting. Most packet data traffic occurs in random quantities and at random times.

In contrast, isochronous data occurs in precise amounts on a precise time basis. It typically represents ordered and timed digital samples from a sensor (such as a telephone or video camera). Thus, isochronous data must be synchronized with clock information to ensure the accurate regeneration of the sampling clock (as distinct from the bit clock) at receiving stations for data reconstruction. The cycle clock is used for this synchronization.

## Principles of FDDI-II Operation

In the FDDI-II mode, all data is formatted in a rigid pattern of 125-microsecond elements known as cycles, rather than being transmitted as packets of variable length separated by variable periods of idle. The cycles are generated by a station known as the cycle master. The cycle master also inserts a latency adjustment buffer, enabling it to maintain an integral number of cycles on the ring.

The cycle format is shown in Figure 2. The starting delimiter marks the start of each FDDI-II cycle and precisely signals the cycle clock from station to station in the ring.

Each cycle contains a header, a data packet group, and 96 cyclic groups of data. Within the header, the C1 and C2 symbols are used to synchronize FDDI-II cycles and establish the cycle sequence value contained in the next octet of the header. The cycle sequence octet provides a modulo 192 cycle sequence count, which can be used to establish subcycle data rates (or synchronization signals). It also enables the recognition of cycle sequence errors, should they occur. Remaining values of cycle sequence are used in a monitor ranking protocol to establish backup cycle masters. The isochronous maintenance channel octet provides a 64kbps voice channel for maintenance purposes.

The 16 symbols of programming template (P0 through P15) within the header determine whether the corresponding wideband channel (0 through 15) is allocated to packet or isochronous traffic. Thus, each symbol ( $P_n$ ) controls the flow of the corresponding data octets in the cyclic groups. A symbol is 4 bits, equalling one-half of an octet. In Figure 2, P1 and octet 1 of each cyclic group, corresponding to wideband channel #1 being allocated, are shaded.

Cyclic groups are 16 octets long, with each octet corresponding to one of the 16 possible wideband channels that may be allocated. Cycles occur at an 8-kHz rate. The multiplication of 96 cyclic groups by 8 bits by 8 kHz yields the 6.144-Mbps rate of a wideband channel.

## FDDI-II Services

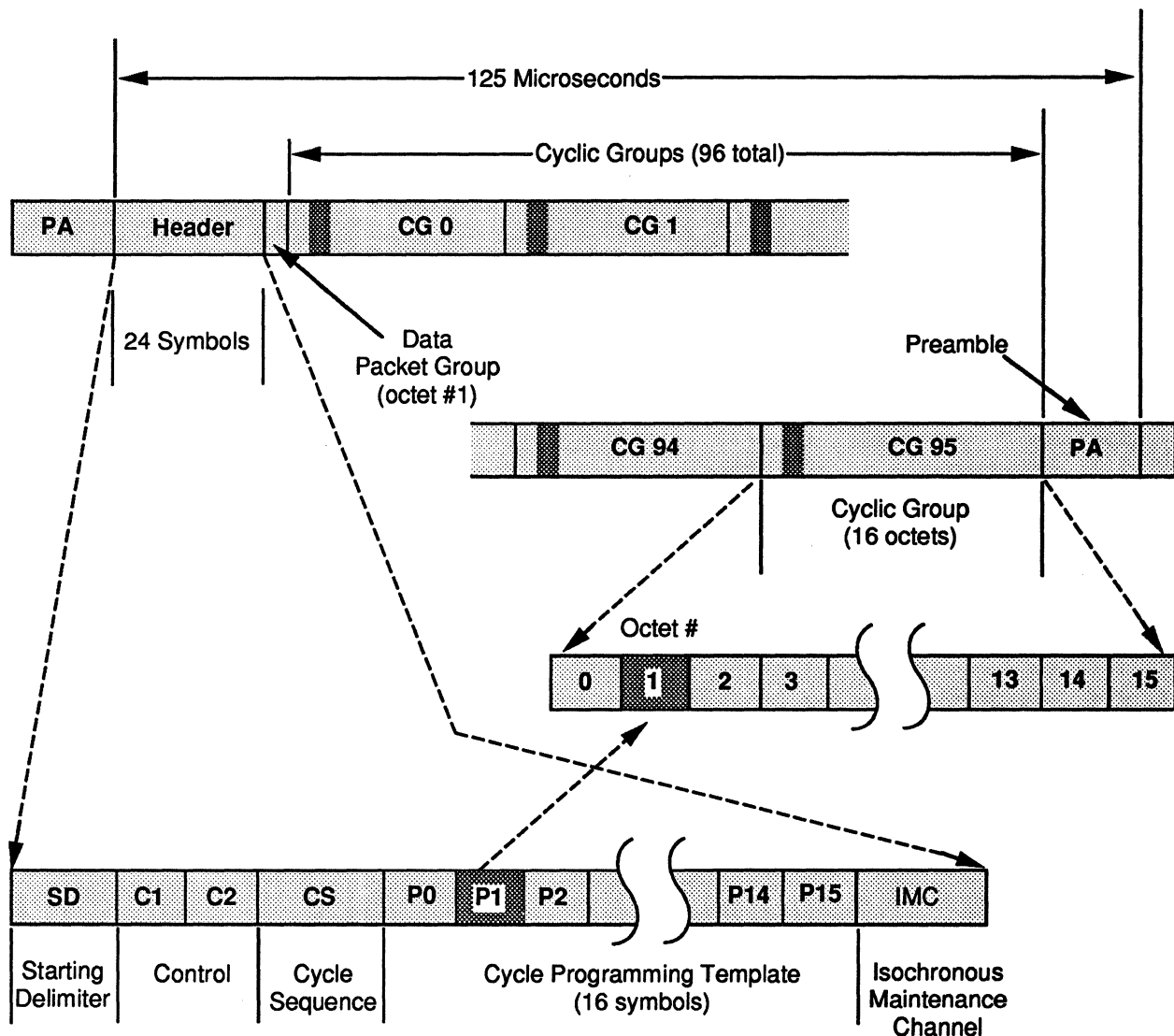
As shown in Figure 3, FDDI-II provides a service access point for the isochronous data stream when a wideband channel is allocated. Each wideband channel provides a 6.144-Mbps data stream to the attachment at its service access point, commonly referred to as a circuit-switch multiplexer (CS-Mux). Data can be repeated or replaced bit-by-bit or octet-by-octet.

How this data is handled depends on the design and programming of the individual CS-mux. The CS-muxes in each of the stations with access to a given wideband channel have access to the same isochronous data. Call setup or other procedures will have established which bits of this stream are to be examined and which are to be replaced at this service access point.

Considering all of the stations that can participate in a given wideband channel, a variety of services are possible. Using a single octet (64 kilobits) of wideband channel #1 as an example, this octet might be written by one station and observed by all others. This could be a broadcast phone message. Another octet might be observed and replaced by a station, then observed and replaced by another, with no other stations having access. This could be a normal one-on-one phone conversation. Using a larger bandwidth example, such as the whole wideband channel, would allow similar services for television pictures.

In addition to the 16 service access points for the wideband channels, FDDI-II provides one packet data service

Figure 2.  
FDDI-II Cycle Format



The starting delimiter of the cycle format marks the start of each FDDI-II cycle and precisely signals the cycle clock from station to station in the ring. Each cycle contains a header, a data packet group, and 96 cyclic groups of data.

access point for the FDDI frames that can be transported in the bandwidth not allocated to wideband channels. This is a physically discontinuous data stream because all octets belonging to each wideband channel have been diverted to the service access point for the corresponding wideband channel. As usual, MAC examines each data symbol and may repeat or replace it.

### FDDI-II Packet Data

The data packet group contains 12 octets of data. The first octet immediately follows the header. The remaining octets are distributed through-out the cycle, with one octet preceding each eight cyclic groups. The data packet group provides a residual 768-kHz (34-Mbps) packet channel when all 16 wideband channels are allocated.

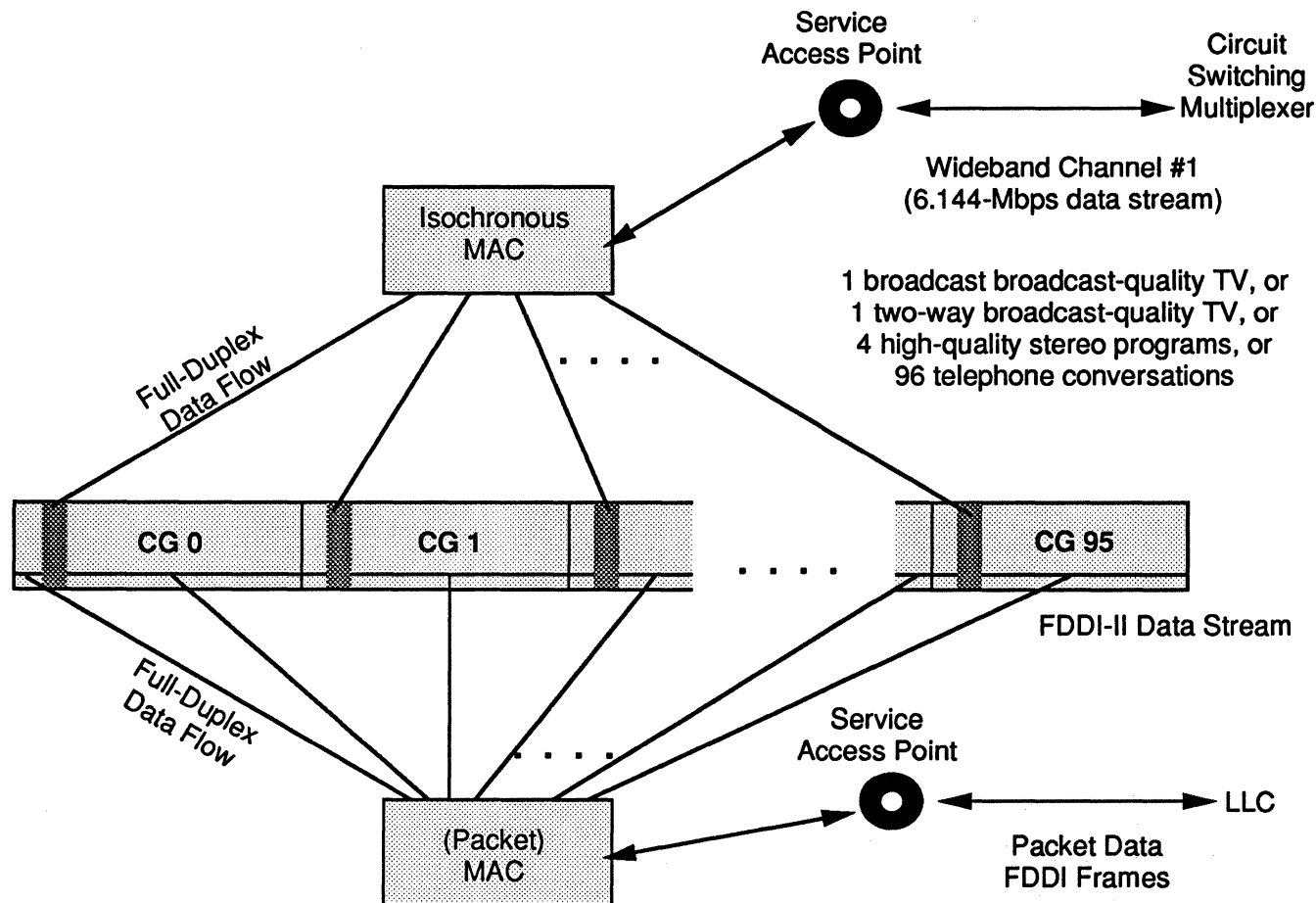
The total bandwidth available for packet traffic is the sum of the data packet group and any wideband channels

not allocated to isochronous traffic. Thus, the packet data stream consists of the 12 data packet group octets plus any of the 16 octets in the 96 cyclic groups for which the corresponding wideband channel is not allocated. Figure 3 shows the flow of data within an FDDI-II station with only wideband channel #1 allocated to isochronous data.

Within the packet traffic bandwidth, frames are transmitted in the normal token-ring protocol fashion, with a token circulating to indicate availability of the medium for transmission of a frame. The packet data stream, once separated from the octets of wideband channels that have been allocated, is exactly like the stream of frames in the basic FDDI, with two important differences. First, two logically adjacent octets of a frame may have a time gap between them. Second, a slightly modified starting delimiter is necessary to distinguish the start of a frame from the start of a cycle.



Figure 3.  
FDDI-II Services



FDDI-II provides service access points for the isochronous data stream when wideband channels are allocated. In the figure, wideband channel #1 is allocated to isochronous data. The allocation of additional wideband channels provides additional service access points, one for each channel.

The FDDI-II mechanism provides a very efficient way to multiplex packet and isochronous data on the same medium. For example, with eight (i.e., one-half) of the wideband channels totalling 49.152-Mbps of bandwidth allocated to isochronous service, 49.92-Mbps of bandwidth would still be available for packet traffic. Furthermore, the packet data is not subjected to the delay of the latency adjustment buffer in the cycle master. Thus, FDDI-II has retained basic FDDI characteristics essential to an efficient frame transport service.

### Hybrid-Ring Operation

An FDDI-II ring is initialized in basic (token) mode. The switch to a hybrid mode of operation, combining both packet and circuit-switched data capabilities, occurs only after a station has negotiated for and won the right to be cycle master along with the required synchronous bandwidth allocation. The cycle master then generates cycles at an 8-kHz rate (every 125 microseconds) and inserts the latency adjustment buffer.

Once a station has been allocated one or more wideband channels, that station may suballocate the combined

bandwidth of these wideband channels as desired. In FDDI-II, each subchannel allocation is designated as X bits beginning at octet M after the cycle clock marker in wideband channel number N. This definition allows connections at data rates of all multiples of 8 kbps (i.e., X=1) up to the 6.144-Mbps data rate of a wideband channel. Commonly used subchannels include 16-, 32-, 64- (B-channel), 384-, 1536-, 1920-, and 2048-kbps isochronous data rates. Each subchannel represents a full-duplex circuit-switched connection.

Mixtures of isochronous data rates within a wideband channel are allowed. If required, the aggregate bandwidth of several wideband channels can be allocated as one virtual service (e.g., uncompressed TV).

To reflect changing network needs, wideband channels can be allocated and de-allocated without any interruption of ring operation. A change in allocation is accomplished by the cycle master. To avoid any disruption to packet traffic, a token is first captured, giving the cycle master the right to transmit frames (and thus impunity to remove all incoming frames). A high-reliability protocol is used to

change the allocations of wideband channels to prevent the generation of cycle format errors as a result of errors in the data stream.

An important aspect of wideband channel allocation is that a wideband channel, once allocated, can be used as desired by the participating stations. In some cases, this may be in a well-known manner, understood by all stations. In other cases, the use of the wideband channel may be in a manner known only to the small set of stations privy to its use. In neither case is the operation of the ring compromised. Both kinds of use are permitted.

---

### **Status of FDDI-II Standards**

HRC is in the final approval process as both an ANSI and an ISO standard. MAC-2 and PHY2 documents, both considered technically sound by the X3T9.5 committee, are now in the approval process for ANSI standards. Any changes should be superficial.

The FDDI SMT document is the most sophisticated multivendor interoperability document in existence today for the lower levels of the OSI reference model. Its very existence sets FDDI apart from all other LAN standards.

SMT, now in an extensive X3T9 letter-ballot comment-resolution process, is fast becoming a technically firm document and will soon be submitted for final approval as an ANSI standard.

FDDI-II requires several extensions to SMT; the current document has been developed to easily permit these additions. Further, having learned from the past, contributors can be expected to present the committee with working solutions that, having already been proven in the FDDI-II working group, can be readily adopted into SMT. Thus, SMT considerations are not expected to cause significant problems in the deployment of FDDI-II-capable products. Indeed, despite the massive effort on SMT, deployment of FDDI products has not been impeded. Two previous revisions of SMT demonstrated interoperability at Interop '89 and '90.

Quietly, and without much fanfare, the critical pieces necessary for FDDI-II, including standards and chip sets, have been put in place. And by all indications, success is at hand. The pundit who quipped that "FDDI is the Ethernet of the '90s" didn't know it, but he or she really meant FDDI-II-and he or she was right. FDDI-II will become the dominant high-performance LAN standard of the '90s. ■

# Simple Network Management Protocol (SNMP)

## In this report:

SNMP Architecture.....	2
How SNMP Works.....	4
SNMP's Advantages .....	7
SNMP's Disadvantages .....	7

## Datapro Summary

The Simple Network Management Protocol (SNMP) is a viable alternative to the ISO CMIP over TCP/IP (CMOT) protocol. Originally defined to manage TCP/IP networks, SNMP can also be used to manage OSI networks. "Agents", "managers", and "Management Information Bases" combine to control network devices. Non-SNMP devices can be managed with proxy agents. SNMP's designers created a successful vehicle for multivendor network management; however, the protocol itself is less important than what users can do with the data. This report explains SNMP's architecture, details its history and implementation, and discusses its advantages and disadvantages.

Simple Network Management Protocol (SNMP) originated in the Internet community as a means for managing TCP/IP networks and Ethernet networks. During 1989, SNMP's appeal broadened rapidly beyond the Internet, attracting waves of users searching for a proven, available, multivendor-network monitoring method.

### SNMP History

SNMP evolved from the Simple Gateway Management Protocol (SGMP), formalized in November 1987 by Chuck Davin of MIT (formerly with Proteon); Jeffrey Case of the University of Tennessee/SNMP Research, Inc.; Mark Fedor of NY-SERNet; and Martin Schoffstall of NYSERNet. SGMP was an early attempt to address the issue of network

router management under TCP/IP. While SNMP is similar to SGMP in architecture and design philosophy, the syntax is different and the two protocols are incompatible.

In August 1988, the same four SGMP authors formalized SNMP as an Internet Draft Standard. In April 1989, SNMP became an Internet Recommended Standard (RFC 1098). Table 1 lists the current SNMP RFCs.

The Internet Activities Board (IAB) is currently examining both SNMP and OSI's Common Management Information Services and Protocol over TCP/IP (CMOT) as potential solutions for TCP/IP network management. Although many analysts view CMIP over the OSI stack as the preferred long-term solution for network management, TCP/IP implementations are widely available today and will continue in use for some time. Furthermore,

—By *L. Michael Sabo*  
*Communications Architect*  
*SSDS, Inc.*

SNMP is eclipsing CMOT as the interim TCP/IP solution.

### Using SNMP

Using SNMP, a network administrator can address queries and commands to network nodes and devices. It can be used to monitor network performance and status; control operational parameters; and report, analyze, and isolate faults. The protocol performs these functions by carrying management information between *managers* and *agents*.

## SNMP Architecture

SNMP operates on three basic concepts: manager, agent, and the Management Information Base (MIB) (see Figure 1).

A **manager** is a software program housed within a Network Management Station. The manager has the ability to query agents, receive agent responses, and set specific variables using various SNMP commands.

An **agent** is a software program housed within a managed network device (such as a host, gateway, terminal server, etc.). An agent stores management data and responds to the manager's data requests.

The **Management Information Base (MIB)** is a database of managed objects, accessible to agents and manipulated via SNMP to provide network management information.

### The MIB

The MIB conforms to the Structure of Management Information (SMI) for TCP/IP-based internets, as described in RFC 1155. This SMI, in turn, is modeled after OSI's SMI, as defined in Draft

Proposal (DP) 2684. While the SMI is similar for both SNMP and OSI environments, the actual objects defined in the MIBs are different.

SMI conformance is important, since it means that the MIB is capable of functioning in both current and future SNMP environments. In fact, the Internet SMI and the MIB are completely independent of any specific network management protocol, including SNMP.

### The Internet-Standard MIB

Each SNMP agent contains instrumentation that, at minimum, must be capable of gathering "Internet-standard MIB" objects specified in RFC 1156 (May 1990). Such objects include network addresses, interface types, counters, thresholds, and similar data for all network devices and NMSs involved. (Nonstandard MIB objects are manageable under SNMP, provided they are defined using SMI conventions specified in RFC 1155.)

Objects are defined using a subset of Abstract Syntax Notation One (ASN.1), the ISO SMI specification language. Also, SNMP's designers chose the ASN.1 basic encoding rules to align the protocol with the OSI environment.

The standard MIB's structure is logically represented by a tree. The root (which is unlabeled) divides into three main branches: ISO, CCITT, and Joint ISO/CCITT (see Figure 2). Within the Internet subtree, which is several levels down the ISO branch, exist four subtrees: Directory, Management, Experimental, and Private. The Experimental subtree is reserved for Internet research purposes. The Internet-standard MIB, now at revision level MIB-I, finds its root under the Management subtree. Under the Private subtree is a very important branch called Enterprises.

Figure 1.  
SNMP Architecture

SNMP has three architectural components: manager, agent, and MIB. Agents collect management information through instrumentation and store the information in a database called the MIB. The agent will provide management information to an SNMP manager upon request.

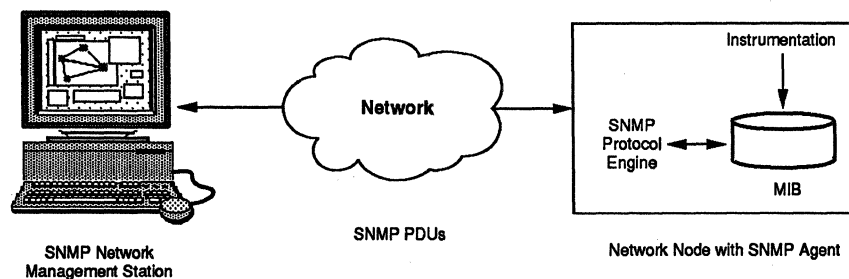
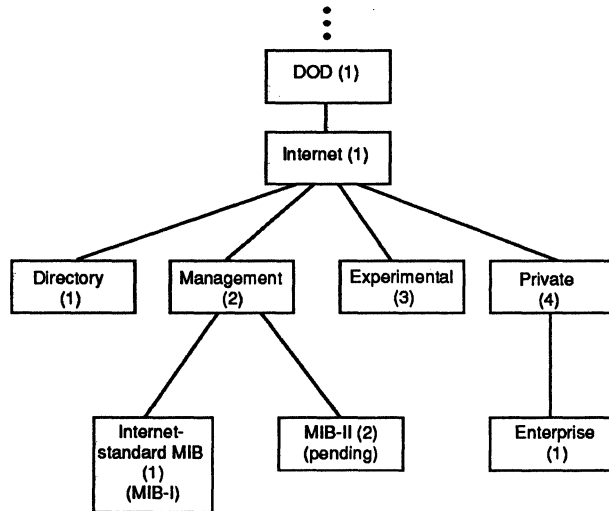


Figure 2.  
The Management Information Base (MIB)



This figure depicts the location of the Internet-standard MIB within the Internet tree.

### The Private Enterprise

The Enterprise subtree, with its root under Private, is reserved for organizations wishing to develop extensions to the Internet-standard MIB. Organizations may apply for a specific Enterprise number, which uniquely identifies the organization's management tree, and is essential if the device is to manage objects other than those defined in the standard MIB.

Organizations may obtain a Private Enterprise number (free of charge) by sending an electronic mail message on the Internet to [jkrey@isi.edu](mailto:jkrey@isi.edu). While the IAB controls the contents of the Internet-standard MIB, Private Enterprise MIBs are controlled by vendors or other special-interest groups. As of July 1991, over 250 Private Enterprise numbers had been assigned.

### Agent Responsibilities

Each agent possesses its own MIB view, which includes the Internet-standard MIB and, typically, other extensions. The agent's MIB need *not* implement *every* group of defined variables in the standard MIB specification RFC 1156. For example, gateways need not support objects applicable only to hosts. This eliminates unnecessary overhead, facilitating SNMP implementation in smaller LAN components with little excess memory capacity. If a device supports a specific protocol (such as UDP), however, *all* objects from that particular group (i.e., the UDP group) within the MIB *must* be supported.

## Obtaining RFCs on the Internet

RFCs are available through FTP from Internet host [NIC.DDN.MIL](ftp://NIC.DDN.MIL). Login using the username **anonymous** and the password **guest**. Once logged in, type in **get RFC:RFC-FCnnnn.txt**, where **nnnn** is the RFC number. For example, **get RFC:RFC1157** will retrieve *A Simple Network Management Protocol (SNMP)*.

RFCs can also be obtained through electronic mail. Send a message to [SERVICE@NIC.DDN.MIL](mailto:SERVICE@NIC.DDN.MIL) and place the RFC number in the subject field.

FTP to [NIC.DDN.MIL](ftp://NIC.DDN.MIL) with the **anonymous** guest login to obtain a current index of all RFCs. Once the session is established, type **dir RFC:RFC-INDEX**. A document name, such as **RFC-INDEX.TXT.nnnn**, will be returned. The **nnnn** represents the latest RFC number. Type **get RFC:RFC-INDEX.TXT.nnnn** to fetch the index for review. Type **quit** to logout.

An agent performs two basic functions:

- Inspects variables in its MIB
- Alters variables in its MIB

Inspecting variables usually means examining the values of counters, thresholds, states, and other parameters. Altering variables may mean resetting these counters, thresholds, etc.

It would be possible to actually reboot a node, for example, by setting a specially defined variable (assuming one exists) to **reboot=1**. Most "SetRequest" commands accomplish tasks such as modifying routes or interface types, however. (For more information on SetRequest and other SNMP commands, see *How SNMP Works*, following.)

### Manager Responsibilities

Managers execute network manager station (NMS) applications and often provide a graphical user interface depicting a network agents map. The manager also typically archives MIB data for trend analysis.

**Table 1. RFCs Applicable to SNMP**

RFC Reference	Title	Date
RFC 1155	Structure and Identification of Management Information for TCP/IP-based Internets	May 1990
RFC 1156	Management Information Base for Network Management of TCP/IP-based Internets	May 1990
RFC 1157	A Simple Network Management Protocol (SNMP)	May 1990
RFC 1158	Management Information Base for Network Management of TCP/IP-based internet: MIB-II	May 1990
RFC 1161	SNMP over OSI	June 1990
RFC 1187	Bulk Table Retrieval with the SNMP	October 1990
RFC 1215	A Convention for Defining Traps for use with the SNMP	March 1991
RFC 1227	SNMP MUX Protocol and MIB	May 1991
RFC 1228	SNMP-DPI—Simple Network Management Protocol Distributed Program Interface	May 1991
RFC 1229	Extensions to the Generic-Interface MIB	May 1991
RFC 1230	IEEE 802.4 Token Bus MIB	May 1991
RFC 1231	IEEE 802.5 Token Ring MIB	May 1991
RFC 1232	Definitions of Managed Objects for the DS1 Interface Type	May 1991
RFC 1233	Definitions of Managed Objects for the DS3 Interface Type	May 1991
RFC 1238	CLNS MIB—for use with connectionless Network Protocol (ISO 8473) and End System to Intermediate System (ISO 9542)	June 1991
RFC 1239	Reassignment of Experimental MIBs to Standard MIBs	June 1991
RFC 1243	AppleTalk Management Information Base	July 1991

## How SNMP Works

### SNMP Protocol Data Units (PDUs)

To carry out these duties, SNMP specifies five types of commands, or verbs, called *Protocol Data Units*:

1. GetRequest
2. GetNextRequest
3. SetRequest
4. GetResponse
5. Trap

### GetRequest and GetNextRequest

An agent will inspect the value of MIB variables after receiving either a GetRequest or GetNextRequest command (PDU) from a manager.

### SetRequest

The agent will alter MIB variables after receiving a SetRequest command. An NMS, for example, could instruct an agent to modify an IP route using SetRequest. It is a powerful command and could corrupt configuration parameters and seriously impair network service if used improperly. Due to SNMP's lack of inherent security measures (see Concerns, following), some component vendors have not implemented or enabled SetRequest within their SNMP agent implementations. Many vendors are working to enhance security features within their products in order to offer a more secure SetRequest implementation.

### GetResponse

An SNMP agent responds to an SNMP manager's GetRequest, GetNextRequest, and SetRequest

## GNMP: New Kid on the Block

On May 31, 1991, the National Institute of Standards and Technology (NIST) issued its first version of the *Proposed Government Network Management Profile (GNMP)*, a document to provide "the standard reference for all Federal Government agencies to use when acquiring Network Management (NM) functions and services for computer and communications systems and networks." The proposal discusses:

- the scope of the GNMP
- its applicability
- its development
- its specification sources
- its relationship to the Government Open Systems Interconnection Profile (GOSIP)

### Scope

The GNMP mandates CMIS and CMIP as the management information exchange protocol. Managed Objects (MOs) are

included from DMI, NMSIG 90/197, IEEE 802.3 HUB Management, and other international standard publications. The proposal also details five systems management functions:

1. Object Management Function
2. State Management Function
3. Attributes for Representing Relationships
4. Alarm Reporting Management Function
5. Event Reporting Function

### Applicability

GNMP is mandatory for federal agencies. This presents problems since some of the standards it adopts (CMIP, for example), and the GNMP itself, are still under development. The intent, however, is to provide guidance in selecting from current NM tools while evolving tighter specifications.

### Development

NIST conducted a survey of federal agencies in the summer of 1990. The results indicated that the management of local area networks and their interconnecting bridges was a prime concern. GNMP was proposed specifically to address that concern. The Phase I implementation, proposed as GNMP Version 1.0, focuses on management of OSI model layers 1 and 2.

### Specification Sources

NIST cites part eighteen of the *OIW Stable Implementation Agreements, December 1990* as the primary specification source for GNMP Version 1.0. Other sources, however, provided the additional specifications needed to provide the minimum-required management capabilities.

### Relationship to GOSIP

GNMP is the management information specification of the networks defined by GOSIP. GOSIP specifies the protocol stacks and system services that convey the management information between managed objects and their managing systems. GNMP cites GOSIP

heavily, and later versions of each document will continue to cross-reference each other.

### Conclusion

The GNMP will have a significant impact on the network management marketplace in the United States for two primary reasons. First, it is a guideline for implementing accepted international standards for network management. Second, the use of GNMP is mandated for federal agencies and encouraged for companies that do business with federal agencies.

The second reason is causing some trouble in Corporate America. GNMP specifically omits Simple Network Management Protocol (SNMP), widely used for network management. It is conceivable, though unlikely, that use of GOSIP and GNMP could become mandatory for firms wishing to conduct business with the federal government and for colleges and universities that receive federal aid, causing a costly retooling of entrenched network systems.

PDU with a GetResponse PDU. The GetResponse includes the original request followed by the requested information. Returning the original request with the response implements a stateless protocol where the manager need neither track outstanding requests nor correlate replies.

### Traps

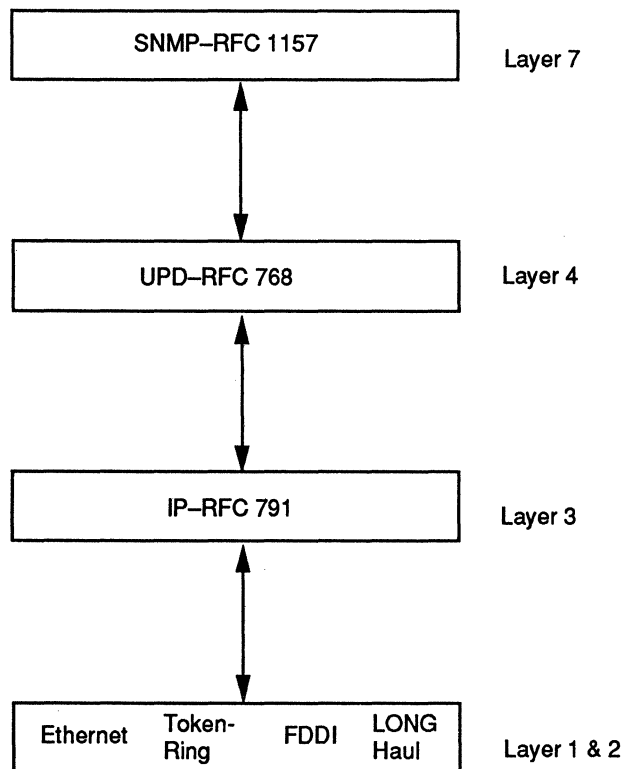
Trap is a special, unsolicited command type that agents send to a manager after sensing a prespecified condition, such as ColdStart, WarmStart,

LinkDown, LinkUp, AuthenticationFailure, EGP-neighborLoss, or other enterprise-specific events. Traps are used to guide the polling timing and focus, which SNMP employs to monitor the network's state.

### Transport Mechanisms

As mentioned previously, managers and agents exchange commands via messages. SNMP's monitoring and control transactions are not actually

Figure 3.  
SNMP Protocol Suite



This figure details the relationship between SNMP RFCs and the OSI Seven Layer Model.

TCP/IP dependent—SNMP only requires the datagram transport mechanism to operate. It can therefore be implemented over any network media or protocol suite, including OSI. There are currently two standard SNMP transport mechanisms: User Datagram Protocol (UDP) and within Ethernet frames (as defined in RFC 1083). Currently, there are no commercial implementations of SNMP directly over Ethernet. All commercially available SNMP NMSs use UDP to exchange SNMP PDUs. Figure 3 diagrams SNMP's relationship with its transport mechanisms in terms of the OSI model.

### UDP

Each SNMP message is represented entirely within a single UDP datagram. This lessens the message processing burden and helps to minimize the agent's complexity. The SNMP message consists of:

—version identifier—SNMP community name—PDU

The *version identifier* refers to the RFC version (currently at 1). An SNMP *community* consists of an agent and its associated applications. As mentioned before, a PDU is one of five command types. The SNMP protocol entity receives most messages at UDP port 161 on its associated host. Traps are received on UDP port 162.

### Ethernet Frames

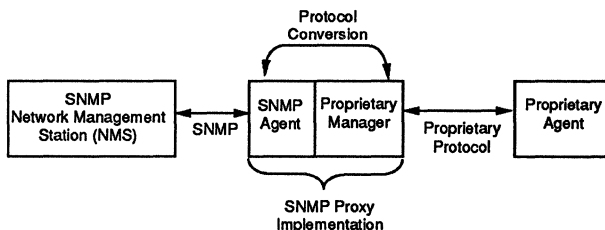
SNMP over Ethernet frames, while implementable, is not recommended by the SNMP specification authors. While the SNMP message looks the same, an SNMP NMS must be configured to accept SNMP PDUs directly from the Ethernet driver.

### SNMP Proxy Agents

Proxy agent software permits an SNMP manager to monitor and control network elements that are otherwise not SNMP addressable. For example, a vendor wishes to migrate its network management scheme to SNMP but has devices on the network that use a proprietary network management scheme. An SNMP proxy can manage those devices in their native mode. The SNMP proxy acts as a protocol converter, translating the SNMP manager's commands into the proprietary scheme. This strategy facilitates migration from the current proprietary environment to the open SNMP environment (see Figure 4).

Proxy agents are well suited for vendors with an existing base of non-SNMP devices communicating efficiently under a proprietary scheme. By using a proxy agent, the vendor can reduce the investment risk of putting SNMP equipment into the field.

Figure 4.  
SNMP Proxy Implementations



SNMP proxies help a vendor migrate a proprietary network management scheme into the SNMP environment.



## SNMP's Advantages

SNMP's major advantages are the following:

- Its simplicity eases vendor implementation effort
- Its memory and CPU cycle requirements are lower than CMIP
- Its protocol has been used and tested on the Internet
- Its products are available

### Simplicity

SNMP's designers successfully kept the protocol simple, easing vendor implementation and thereby encouraging widespread implementation.

### Memory and CPU Use

By using only a subset of ASN.1 to define the MIB and implementing only five command types, agent implementations require far less memory and fewer CPU cycles than most network management protocols, including CMIP.

### Tested and Used

SNMP has a distinct advantage over CMIP in that it was tested and actually used by the Internet community *before* it became a standard. The RFC process of standardization, in fact, requires that a critical mass of users employ and then comment on a particular protocol or other specification before the authors submit it for approval. In contrast, ISO's method is to develop a preliminary standard after much commenting, with implementation and testing as a poststandard process.

### Availability

Finally, SNMP products are available now. While SNMP is not sophisticated, its availability will give many network managers the opportunity to try out multivendor network management and possibly discover what they need to manage their networks. SNMP developers and proponents know that the SNMP tools available now fall far short of satisfying user needs. Yet, the best way to clarify those needs is to get experience and redefine requirements on an ongoing basis.

## SNMP's Disadvantages

SNMP has several disadvantages, including the following:

- Lack of global vision
- Weak security features
- Problems with the Trap command

### Lack of Global Vision

While SNMP's widespread deployment before standardization is an advantage in one respect (see SNMP's Advantages), the protocol's definition is more or less cast in stone—essentially without giving vendors and users outside the U.S. the opportunity to voice their needs, concerns, and suggestions. The Internet Activities Board (IAB) is, of course, under no obligation to do so. Yet European vendors, commercial users, and academicians are now embracing SNMP, since they share our same need for an open, multivendor network management protocol. SNMP may spread throughout the world and leave some non-U.S. users at a disadvantage.

In contrast, ISO/OSI standards developers have a truly global vision and put a high priority on accommodating the viewpoints expressed by all nation representatives. In the OSI community, developing nations are heard on an equal par. The price of emphasizing universal applicability within OSI standards is a much slower pace of standards development, as compared to rapidly developed standards such as SNMP.

### Security Issues

There are very few security mechanisms defined as part of the SNMP protocol specification. For example, there is no capability defined to ensure that SNMP PDUs received by an agent actually originated from an actual manager—and not from an unauthorized interloper.

Thus, vendors are reluctant to support the SetRequest verb on their agent implementations. SNMP does, however, support access modes of read-only/read-write classifications for MIB variables. To employ this capability, the user may configure a variable such as RouteTable as read only. This prevents the agent from setting RouteTable to a potentially harmful value, if an unauthorized perpetrator tries to fool the agent with a SetRequest command. However, using read-only when the variable should be defined read-write effectively

disables the SetRequest verb on those variables and reduces SNMP's functionality.

### Trap Problems

SNMP does not define the mechanism for where a Trap should be sent, nor what the agent should provide as part of a Trap (even for standard Traps). The specification merely notes that a Trap should include "interesting information." Thus, Trap is implementation specific.

---

### Conclusion: Meeting the Goals

SNMP's authors adhered to several design goals during the protocol's comparatively brief development time:

1. **Keep the agent simple**—Minimize the number and complexity of the agent's management functions.
2. **Make SNMP monitoring and control extensible**—Accommodate unanticipated aspects of network operation and management.

---

This report was prepared and updated exclusively for Datapro by L. Michael Sabo, a communications architect with SSSD, Inc., Littleton, CO. Mr. Sabo is currently consulting on various networking and internetworking projects. Previously, he participated in porting TCP/IP to the emerging ANSI High-Performance Parallel Interface (HIPPI) Gigabit/sec LAN standard. Mr. Sabo has been active in integrated network management. He participated in developing an object-oriented and SNMP-based network management architecture for Lockheed Integration Services. This effort included defining numerous private enterprise management information base (MIB) objects to support system management functions.

Mr. Sabo is a member of the SNMP working group and has been active in the Internet for six years. He is a member of the board of advisors for Datapro *Network Management*. He holds a master's degree in data processing management from the University of Denver and a bachelor's degree in Computer Science from Wright State University.

3. **Make SNMP architecture independent**—Do not code to any particular host/gateway architectures.

SNMP's designers achieved the first goal by limiting functions to five and by requiring only an unreliable datagram service such as UDP. Simplifying the agent reduces vendor development costs—making it more attractive for component vendors to support SNMP. Widespread availability of SNMP agents is a prerequisite both for user acceptance of SNMP *and* for stimulating NMS vendors to integrate SNMP manager implementations.

The second goal is realized by employing the MIB. The current Internet-standard MIB will evolve and expand to include more items—thus giving network managers more control over their networks. New objects are added by integrating new subtrees into the MIB. SNMP can easily traverse the new structure by using the GetRequest or GetNextRequest command.

The third goal is realized via the open communications architecture above which SNMP operates and by SNMP's capability to operate over many different transport mechanisms. In addition, through the use of ASN.1 basic encoding rules, SNMP is not tied to any specific machine architecture.

SNMP's designers successfully created a new vehicle for multivendor network management, a first step toward solving an increasingly critical problem. SNMP is merely a vehicle. The protocol itself is less important than the tools and applications that must be developed; how the data is gathered is less important than what users can do with the data. Most of the work lies ahead, as users gain experience with managing multivendor nets and determine what works and what does not. ■

---

# Dueling Protocols: SNMP vs. CMIP

---

## In this report:

Differences .....	2
The Field of Battle .....	3
Who's Who in CMIP? .....	4
Network Detente? .....	4

## Datapro Summary

Two network management protocols—SNMP and CMIP—are attracting attention for their abilities to effectively manage heterogeneous networks. Although the protocol families have similar goals, they differ in transport layers, data access philosophy, polling/reporting methods, and other ways. Supporters from both camps are now coming together with the realization that SNMP and CMIP will each play important roles in managing LANs of the future.

One of the hottest topics in networking today is network management. Now that most of the connectivity and interoperability issues have been or are being resolved, you can turn your attention to keeping track of the devices on your networks, checking on the network's performance and load, and diagnosing and correcting any problems.

While products that manage homogeneous networks have been available, managing heterogeneous networks is more complex. Some people fear that if you depend on one vendor's proprietary solution to manage your network, that vendor could try to steer the blame for any problems toward third-party products.

Therefore, much of the attention in managing heterogeneous networks has focused on two families of network management protocols: the simple network management protocol (SNMP), which comes from a de facto standards-based background of TCP/IP communication, and the common management information protocol (CMIP), which derives from a de jure

standards-based background associated with the Open Systems Interconnection (OSI).

Until fairly recently, debates comparing the two sets of protocols have raged in conferences, electronic discussion groups, and scholarly papers. Now, however, proponents of both sides are beginning to admit that *both* protocol families have a role to play in managing the networks of the future. And discussions are currently moving toward trying to figure out which duties each protocol family is best suited for.

## Similarities

In many ways, SNMP and CMIP are more similar than they are different—a view that even die-hard proponents of one or the other admit. “They’re similar in that both have the same goal: to move network management information from one place to another, so the network manager can retrieve information from a device, make changes, and find out what’s broken,” says Jeff Case, one of the authors of SNMP and president of SNMP Research (Knoxville, TN), a company that supplies core software to SNMP vendors. “For problem diagnosis, for capacity planning, for report generation—both would be useful in that regard.” Both protocol families use the concept of a management information base. An MIB consists of a set of variables, test points, and

---

This Datapro report is a reprint of “Dueling Protocols” by Sharon Fisher, pp. 183-190, from *Byte*, Volume 16, Number 3, March 1991. Copyright © 1991 by McGraw-Hill, Inc. Reprinted with permission.

controls that all devices on the network support and that a network manager can control.

In addition, both protocols allow for vendor-specified extensions to the MIB. These extensions could allow you to control devices more specifically without requiring that you use a least-common denominator approach to network management. They could also enable the management of heterogeneous networks.

In some cases, SNMP proponents have bowed to CMIP and taken advantage of ways in which the CMIP specifications are superior. For example, vendors are making the SNMP MIBs and extensions to them compatible with those that are used by CMIP. In addition, the content and structure of SNMP packets are defined using the abstract syntax notation (ASN.1) OSI-standard protocols.

## Differences

The differences between SNMP and CMIP are also present in a number of areas, and they often end up being as much a matter of "religion" as anything else. "The differences between SNMP and CMIP are in the category of differences between C and Ada," Case says. Specific differences depend on "who you ask—whether they're a frothing-at-the-mouth SNMP lunatic or a frothing-at-the-mouth CMIP lunatic." The following list contains some of the differences.

### Data Access Philosophy

SNMP is oriented more toward retrieving individual items of information; CMIP is oriented more toward retrieving aggregate information, Case says. "Suppose we had a database of employee records, and the employee record consisted of name, number, department, and salary. In SNMP, we would say, 'What is the value of *employee record*? What is the value of *employee name* for a particular employee, and what is the value of *social security number*? What is the *salary* for a particular employee, and what is the *department*?' SNMP would say, 'The value of this field is *this* and the value of that field is *that*.' In CMIP, we say, 'Tell me about *employee*, such that *employee* is so-and-so.'

In other words, Case continues, "in SNMP, you ask for just what you want, and what you asked for is just what you get. In CMIP, you say 'give me the class of what I want, subject to certain constraints,' and it gives you everything except what you threw out. In SNMP, you ask for and receive answers to more focused questions, where CMIP deals with data more in bulk."

Both approaches have their advantages, Case points out. "It depends on what problem you're trying to solve. If you're trying to deal with individual information objects, then you want to use SNMP. Suppose that I wanted to find out about a particular individual's salary. The CMIP approach is to get the whole database and throw out everything you don't want. It's not terribly efficient. But if you want the whole [database], then CMIP is going to be better."

### Polling Versus Reporting

Similarly, SNMP works by polling, or regularly asking each device for its status, while CMIP uses *reporting*, or having the device inform the manager of its status when it changes. "SNMP polls devices to find out if they're dead or alive, while CMIP relies on the device itself to communicate to the management system that something has happened," says David Mahler, formerly responsible for

marketing activities for OpenView (a network management product from Hewlett-Packard) and now vice president of marketing for Remedy (a start-up company in Palo Alto, CA, developing protocol-independent network management products).

CMIP's approach has both advantages and disadvantages. "If you have a large number of devices that you're polling all the time, you can consume net bandwidth [with SNMP]," Mahler points out.

For example, the SNMP demonstration at the 1990 Interop (see the sidebar "The Field of Battle") featured an admittedly unusual 26 network management devices, each doing its own polling. These took up some 15% to 20% of the Ethernet show network, according to Rich Fitzgerald, the western region support manager for Xyplex (Borlboro, MA) who helped to arrange the demonstration.

In general, the whole issue of what percentage of the network the load management should be allowed to take up is unresolved. Many people in the networking community are concerned about it. However, with SNMP's philosophy, "you can have stupid devices that don't have to be smart enough to tell you they have problems," says Mahler. This, combined with SNMP's smaller size requirements, makes it more useful for smaller devices such as PCs.

### Functionality

CMIP is generally thought of as having more specific features and capabilities. But, notes Case, they may well be capabilities that you neither want nor need.

"For example, take the ability to move a table of 10,000 information items from one location to another. CMIP will do that better. But an SNMP person says, 'Why would I want to move a table of 10,000 items? All I want to do is scan the table and ask for the things I wanted in the first place.'"

Karen Auerbach, president of Epilogue Technology (Ventura, CA), agrees. She points out that while CMIP may have more capabilities built in, most of them can be accomplished elegantly with SNMP if you're familiar with the protocol.

### Size and Performance

Case says that "SNMP is going to tend to be smaller, faster, and less expensive than a CMIP implementation. CMIP will require more processor capacity, run slower, require more memory, and be more expensive." This also relates to the polling versus reporting issue, because polling requires less intelligence from the devices being managed than reporting does.

For example, in most cases, vendors can implement SNMP in a pop-up TSR program on an MS-DOS PC, Case says. "You can't do that with CMIP. Of course, if you buy extended memory and a big extended memory board, sure, someone can come up with a counter-example and make a liar out of me."

SNMP's simplicity also gives devices "more bang for the buck" in terms of CPU load, Case says. A standard measure of CPU use in network management is "management operations per second," or MOPS, a similar measure to millions of instructions per second (MIPS) and floating-point operations per second (FLOPS). "You'll get a lot more MOPS out of SNMP than out of CMIP."

Because of SNMP's smaller size, it has even been implemented in such devices as toasters, compact disc players, and battery-operated barking dogs. At the 1990 Interop

show, John Romkey, vice president of engineering for Epilogue, demonstrated that through an SNMP program running on a PC, you could control a standard toaster through a network.

Similarly, Simon Hackett, of the University of Adelaide in Australia, working with TGV (Santa Cruz, CA), demonstrated a CD player with an X Window System interface through which you could select discs from a library, choose which songs to play, and adjust the volume. And you could perform all these functions over a network, as TGV staffers learned one evening when a bored Hackett cranked the volume to 11 in the Santa Cruz offices from his computer in Australia.

At the same Interop conference, Case used a battery-operated barking dog to demonstrate how two SNMP network managers could control the same device. Playing on one of Case's signature expressions, "That dog won't hunt," the dog would walk and yap when directed to by either program, but would return an error message if the two programs tried to control it simultaneously.

Although these demonstrations were just for fun, they pointed out how SNMP could be applied to noncomputer devices as well. Hackett's CD player was connected to the network through a "black box" he built with 64K bytes of RAM. Several people asked to buy copies of the box since it wasn't specific to the CD player. He and Romkey joked about designing a "home-appliances MIB" that, all kidding aside, could be implemented to automate any number of devices.

### Transport Layers

For its underlying transport mechanism, which is what transmits data between nodes on a network, SNMP requires only "unreliable datagrams," which means it can be used with Ethernet, Novell's IPX, UDP, and other simple communications protocols. CMIP, in comparison, requires a reliable transport layer, such as TCP/IP or OSI's connection-oriented TP-4 transport protocol.

While this sounds shaky, "unreliable" in this case means only that the data is sent with no guarantee of delivery. If the receiving device doesn't acknowledge that it's gotten the data, the sending device simply transmits the message again.

The usual analogy for unreliable versus reliable is a letter versus a phone call. The phone call sets up a circuit between the communicating nodes, while the letter is simply sent. On the other hand, letters require less equipment and overhead than do phone calls. What's true for the letter is also true for unreliable datagram transport.

While the reliable transport layer makes CMIP better at retrieving large amounts of data, it may also make the network harder to manage when trouble occurs. And that is when managers need network management the most, Case says. "Say that the network is in a fault state. Use of unreliable transport allows the network management station to retry until it gets through. A connection-oriented network tends not to be able to deal with that, and it may not even be able to get a connection in the first place."

In other words, Case explains, when troubles occur, "you want network management to run on an all-terrain vehicle. You don't want a more fragile vehicle—[the messages] have got to get through."

### Standards and Testing

CMIP, like other OSI protocols, is an international standard controlled by international standards bodies such as the ISO. Vendors can test their implementations, says

## The Field of Battle

### The Field of Battle

Many of the milestones associated with the SNMP-CMIP debate have occurred at Interop, an annual exhibition and technical conference dedicated to helping people set up and manage heterogeneous networks. In 1987, discussion began on developing SNMP based on the earlier, less complex SGMP, while CMIP was still being defined.

User organizations, in general, were interested in getting products with which to manage their already burgeoning networks. Vendors, in general, were hoping to have to develop only one family of network management products rather than two, since even SNMP proponents conceded CMIP would probably be the final result.

Then, at the 1988 Interop, competing panels of SNMP and CMIP vendors each insisted that their protocol was

superior and that future Interop conferences would demonstrate that. At the same time, the Internet Activities Board, which acted as a governing body for the TCP/IP networking community and provided some measure of control over the de facto standard protocols, was debating which of the protocols it should approve. The eventual result was that both were approved: SNMP for the present, and CMIP as a future goal.

CMIP proponents were far less in evidence at the 1989 show. Only a couple of vendors demonstrated products, while numerous SNMP products were shown. Then, at the 1990 show, SNMP exploded: Nearly 50 companies took part in a demonstration referred to as the SNMP Solutions Showcase. Again, only few CMIP products were shown.

Mahler, against a conformance test suite from the Corporation for Open systems (COS), which also performs conformance tests for other OSI protocols. In addition, through public demonstrations such as Interop's as well as more private ones, vendors can demonstrate that their products interoperate.

SNMP, in contrast, is not an international standard, although, like TCP/IP, it is controlled by the Internet Activities Board. Vendors primarily check their implementations with interoperability testing.

Some organizations, international ones in particular, may find that they are required to go with protocols that meet international standards. "The reality is that we're in North America, where TCP/IP is very popular, and so SNMP is [very popular, tool]" explains Mahler. "But that's not true on a worldwide basis."

For example, although the Government OSI Profile of the U.S. government does not yet cover network management, it does require using other OSI protocols in cases where they're applicable and available. It's logical to assume that future implementations of GOSIP will require CMIP.

### Availability of Products

If practicality is the most important principle to you, SNMP has one undeniable advantage: There are a lot more products supporting it than CMIP. "There's certainly a lot

## Who's Who in CMIP?

Different groups of vendors have carried the CMIP torch. These days, the controlling group, formed in 1988, is known as the Open Systems Interconnection Network Management Forum (OSI/NM Forum). It includes "essentially all the major computer and telecommunications companies in the world," says David Mahler of Remedy.

"We got together two years ago because the feeling from our customers was that even if you still have proprietary networks out there, it is necessary for the network management systems to talk to each other. It would be wonderful if everyone moved to totally standards-based networks, but reality suggests that that isn't going to happen overnight. So, the companies got together to devise a way to let the management systems talk to each other."

Mahler points out, too, that the fact that the organization chose the OSI network management protocols doesn't imply anything about the structures of the underlying networks. "We chose OSI rather than anything else. We could have picked SNMP or a proprietary protocol, but because [the Forum] was designed as an international organization from the start, we picked an international standard. It has almost nothing to do with whether the network itself is OSI; it just uses the OSI mechanism."

The purpose behind the OSI/NM Forum is to make sure the vendors all make the same choices along the way toward developing products from the protocol specifications, Mahler says. "There are three phases to forming a standard. One is generating the base standard, which is

performed by standards bodies. When you generate the base standard, you build in options for implementing it." So, when the implementation takes place, vendors produce "implementation agreements" to make sure everyone selects the same options.

The third part is developing a suite for testing interoperability and conformance, Mahler explains, "The OSI/NM Forum didn't want to get involved with that, so we contracted with the Corporation for Open Systems," which also helps develop conformance testing for other OSI protocols. "We're squarely an implementors' group." Almost constantly, some subcommittee is meeting somewhere; in addition, the organization has plenary meetings every quarter. "It's different from a lot of organizations, because it's much more like a multicompany project."

Much of the CMIP versus SNMP rhetoric present in earlier years was due to a different group of vendors. The CMOT group proposed

running CMIP protocols over TCP/IP networks—hence the name: CMIP over TCP/IP, or CMOT. "The CMOT spec is [from] quite a different group of people with different attitudes," Mahler says. "My particular opinion, and I think you'll find it to be the general consensus, is that CMOT is dead. It lost its market window, and SNMP has very well filled the role of management protocol for TCP/IP. The SNMP community delivered more functionality, faster, to the marketplace."

The OSI/NM Forum's role is different, Mahler insists. "The CMOT community was working on the problem of managing TCP/IP devices. That's the same thing that SNMP was doing. The Forum worked on a very different problem. [Its members] didn't care what network you were trying to manage. They said the management systems had to talk to each other and were largely independent of the kind of network out there."

of interest in CMIP, but it doesn't have many interoperable implementations today," says Case. "People can go off and buy lots and lots of products based on SNMP: routers, Ethernet hubs, fiber devices, Ethernet devices—the list goes on and on. I don't think the same is true for CMIP."

For this reason, Case is less concerned about the standards issue. "The standards that are the most interesting to me are the one that are *used*, not the ones that are blessed but not implemented."

Vendors confirm the dichotomy. "We talk of a 'selling standard' and a 'buying standard,'" says Steve Saltwick, area manager for network products at Tadem Computers in Cupertino, California. Customers insist on current or future support for OSI protocols, but they aren't buying such products yet, he says. "OSI is a 'selling' standard—customers want to be able to move to it—but customers are buying SNMP."

Mahler concedes that fewer CMIP implementations exist, but he says it's only a matter of time. "The first year that SNMP came out, there were only three or four [implementations]. The second year, there were 12 to 14. The third year, 30 some. [CMIP] will go through a similar pattern," he predicts. Case agrees that CMIP products are in development and on their way: He's even implementing one himself. (To tell who the CMIP players are, see the sidebar "Who's Who in CMIP?")

## Network Detente?

The future, Mahler and Case agree, will see CMIP and SNMP devices working together to manage networks. "SNMP is focused a little more on the manager-to-device area, whereas the Forum implementation of CMIP [see the sidebar "Who's Who in CMIP?"], which is the most active area right now, is focused on communications between management systems," Mahler says. "We think that the two have largely complementary roles."

"What you may find is that SNMP will be used for some parts, and CMIP will be used for others," concurs Case. "You may find a time where the 'manager of managers,' based on CMIP, is interacting with the SNMP manager to control a particular LAN—SNMP within Dallas, but CMIP between Dallas and Chicago. It's not an either-or."

While this may sound complex, it's not all that different from the way programming languages work now, Case points out. Programmers don't try to write everything in the same language; they work with a "toolbox" of languages, each designed for a specific purpose. "There are dozens of network management protocols today," he says, citing IEEE 802.1, FDDI's SMT, IBM's NetView, DEC's NICE, and IBM/3Com's CMOL as just a few examples. "I wouldn't be surprised if there were more in the future. SNMP and CMIP are the two that happen to be getting the most attention [right now]." ■

---

# Modem Standards

## In this report:

Low-Speed Standards .....	2
High-Speed Standards Grow .....	2
V.32 Specifics .....	2
Data-Manipulation Standards .....	4
Data Compression with V.42bis .....	5
Standards to Watch For.....	7

## This report will help you to:

- Be aware of the most common modem standards currently in use.
- Know the steps that have been taken to improve modem performance and increase the data rates.
- Understand where modem standards come from and how they are made.

---

Sending and receiving files over modem connections is a routine procedure for most personal computer users. It's not unusual, however, to find modems that can't communicate effectively because of compatibility problems—they don't all follow the same standards.

For users, just *understanding* modem standards can be a problem. The maze of modem standards grows constantly. Look at modem advertisements and you'll see a long list—Bell 103J, Bell 212A, V.22, V.22bis, V.32—not to mention proprietary technology and protocols that are licensed by individual companies.

These standards cover a variety of transmission speeds and such features as error correction and data compression. The modem standards in use today come primarily from three sources: Bell Standards, CCITT Recommendations, or EIA/TIA Standards. (For definitions and an explanation of how modem standards are established, see the sidebar.) Table 1 shows the most common modem standards for data rates of from 300 bps to 14,400 bps, over leased-line and dial-up telephone lines.

---

This Datapro report is a reprint of "Modem Business" by Steven E. Turner, pp. 353-360, from *BYTE*, Volume 15, Number 12, November 1990. Copyright © 1990 by McGraw-Hill, Inc. Reprinted with permission.

**Table 1. Modem Standards**

Many standards and recommendations govern how modems are designed. These standards allow modems from many different manufacturers to communicate with one another. An asterisk indicates "with echo cancellation."

Data Rate (bps)	Standard	Line	Duplex
300	Bell 103J	Dial-up	Full
1200	Bell 212A	Dial-up	Full
	Bell 202	Dial-up	Half
	Bell 202	Leased	Full
2400	CCITT V.22	Dial-up	Full
	CCITT V.22bis	Dial-up	Full
	CCITT V.26ter	Dial-up	Full*
4800	Bell 208	Leased	Full
9600	CCITT V.29	Leased	Full
	CCITT V.32	Dial-up	Full*
14,400	CCITT V.33	Leased	Full

### Low-Speed Standards

The most common low-speed standards in use are the Bell 103J standard for 300-bps transmission and the Bell 212A standard for 1200-bps transmission. Almost every modem sold in the U.S. supports these standards, either as the primary rate or as secondary *fallback* rates. Fallback rates are used when the modem is unable to connect at higher rates, usually because the telephone channel is too noisy to provide error-free communication at that rate. For example, if a modem attempts to connect at 2400 bps but determines that the line will not support that rate, the modem may try to connect at 1200 bps or 300 bps instead.

The Bell 103J and 212A standards are two-wire, full-duplex standards. This means that modems that support those standards use ordinary telephone lines, and they transmit and receive data in both directions simultaneously. Even at 1200 bps (212A), the data rate is low enough that the data channel for both directions of transmission can fit comfortably within the 3000-Hz-wide voiceband telephone channel.

Because the CCITT was developing international standards during the 1960s (while Bell was defining U.S. standards), most 1200-bps modems in the rest of the world operate using a standard known as V.22. This is similar to the Bell 212A standard, but the carrier frequencies at which the data channels are modulated are different. Thus, V.22 modems and 212A modems are not compatible, unless special design changes are incorporated.

For 2400-bps transmission, most personal computer modems in use today implement V.22bis. The Bell Standard for 2400-bps data was never completely accepted, because at the time the telephone company's monopoly was dissolved, 2400-bps transmission wasn't yet perfected. As a result, there is almost universal compatibility among 2400-bps modems based on V.22bis.

Like the lower-speed standards, V.22bis is a two-wire (dial-up line), full-duplex standard. To fit two 2400-bps data channels in the 3000-Hz-wide voiceband telephone channel, the data bits are encoded into 4-bit bytes before transmission. Each data signal is then transmitted at 600 baud, and the two modem channels can again fit comfortably within the telephone-line channel.

### High-Speed Standards Grow

Prior to 1984, modem transmission at speeds above 2400 bps was possible only by transferring the data over expensive four-wire (leased) telephone lines. Special standards, such as Bell 208 for 4800 bps, V.29 for 9600 bps, and V.33 for 14,400 bps, were available for use with these leased lines. However, only users needing to transfer very large amounts of data could justify the cost of leasing the telephone lines and buying the more expensive modems.

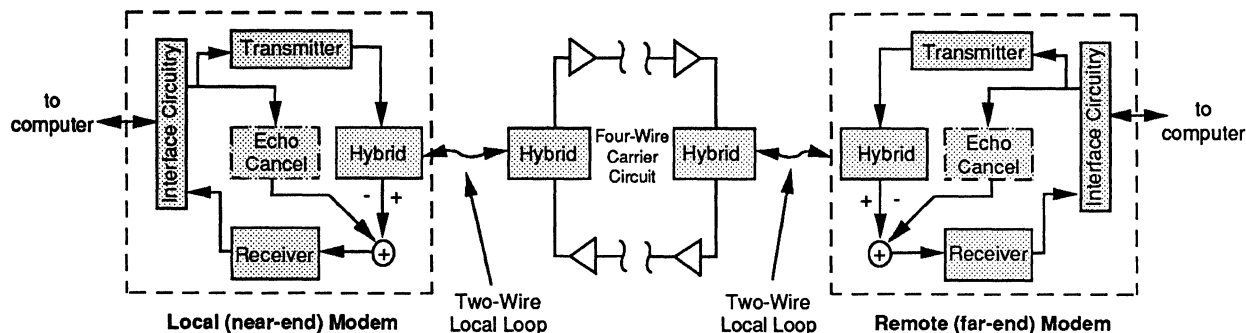
In 1984, the CCITT approved V.32 for use with standard dial-up telephone lines. V.32 leapfrogged from 2400 bps to 9600 bps, representing a 4-to-1 increase in throughput over modems using V.22bis. Using advanced technology to provide 9600-bps transmission over ordinary telephone lines, V.32 put the everyday personal computer user in the high-speed data business for the first time by opening new doors to sharing files and programs rapidly over modem connections.

### V.32 Specifics

The technology required to implement V.32 modems did not come easily. The level of technical expertise needed in developing V.32 modems has been conservatively estimated to be 100 times greater than for V.22bis modems. As a result, fully functional V.32 modems did not become widely available until late 1986—two years after V.32 was adopted.



Figure 1.  
Echo Cancellation



In all modem connections, the computer output signal is sent directly to the modem for transmission over the telephone line. A typical dial-up telephone-line modem connection requires the modem signal to pass through two-wire local-loop channels to the telephone company's nearby central office. From there, it travels over a four-wire circuit to the other end of the connection. The signal then passes through the distant modem's receiver to the computer at the other end. Hybrid circuits connect the two-wire/four-wire links and isolate the transmit and receive signals in the modems. In high-speed modems, such as V.32 modems, an echo canceler is used to further isolate the transmit and receive signals and improve signal reception.

To send 9600-bps data, V.32 modems group the data into 4-bit bytes and transmit them at 2400 baud. Since there is room for only one 2400-baud data channel within the 3000-Hz-wide telephone channel, V.32 calls for both modems to transmit over the same channel at the same time. Each modem must then sort out its own transmitted signal from the signal it is receiving from the other modem. To do this, V.32 modems use *echo cancelers*. Figure 1 shows a typical modem connection, with the echo cancelers included in the modems at each end.

Hybrid circuits inside all modems are designed to match the characteristics of the modem to the telephone line. Since the nature of the telephone network changes constantly, this match is never ideal. This results in part of a modem's transmitted signal being reflected through the hybrid and back into the modem's receiver.

In addition, echoes of the transmitted signal from the hybrid circuits out in the telephone network bounce back into the modem's receiver. To get a good strong received signal, these reflected echoes must be removed before the modem receiver processes its input.

The echo canceler, which is driven by the known transmitted signal, models the echoes produced by hybrid circuits in the modem and the network. The output of the echo canceler is subtracted from the received signal before it goes into the modem receiver for processing, thus eliminating the effects of the echoes. This is not a simple task. The

precision that is required in the echo canceler to remove the echoes is substantial. Since the transmit signal is constantly fluctuating with changes in the data, the echo canceler must continuously adapt to those changes as it mimics the transmitted signal's echo.

Since, at any given moment, a V.32 modem is transmitting more data than a lower-speed modem, the individual V.32 data signals are much weaker and harder to detect. For this reason, V.32 incorporates advanced coding techniques such as *trellis encoding*. Trellis encoding allows the modem to examine several consecutive received signals and look for known patterns before deciding the value of the signal.

This memory effect can produce dramatic reductions in the error rate. The end result is that well-made V.32 modems produce very low error rates and provide reliable, high-speed data transfer between modems. This allows personal computer users to trade programs and download files at rates unimagined in the early 1980s.

In an attempt to push technology barriers even further, the CCITT began, in 1989, to study the idea of extending V.32 up to a 14,400-bps rate. This standard was named V.32bis, since it represented an outgrowth of V.32 rather than a new idea. V.32bis requires even better echo cancelers than does V.32. It also requires an overall improvement in receiver quality. Testing has shown, however, that 14,400-bps transmission over standard telephone lines is quite feasible with proper modem design.

## Where Modem Standards Come From

To grasp the conglomeration of modem standards, some understanding of where they come from and how they are made is important. The modem standards in use today come primarily from three sources. The modulation and coding standards are normally Bell Standards or CCITT Recommendations. The interface standards are either CCITT Recommendations or Electronic Industry Association/Telecommunications Industry Association (EIA/TIA) Standards.

The Bell Standards are holdovers from the 1960s when all domestic modem standards were set exclusively by the telephone

company. In those days, the telephone company had a monopoly on anything connected to its lines, and, by law, it was the only one allowed to sell modems. As a result, it set its own design standards.

The 1968 Carterphone court decision opened the door for other manufacturers to begin making modems, and the method of standards-making changed. Since modems used in other countries at that time generally followed international standards, U.S.

manufacturers became involved in helping develop those standards instead of creating a new set of standards specifically for the U.S.

Today, most new modem standards are created by the CCITT, based in Geneva, Switzerland, and affect modem users worldwide. In the U.S., modem experts participate in national standards development groups, such as the TIA, to create those standards needed solely for U.S. interests. They also generate technical papers and proposals for the international CCITT organization and join technical experts from other countries in attending CCITT meetings.

Together, these groups work out the fine details of new international modem standards. However, when it comes time to vote on the new standards, each member country is granted only one vote. An official of the U.S. Department of State (the formal representative to the CCITT) casts the U.S. vote.

Many standards efforts never make it to a vote because of technical problems or political snags along the way. As a result, those that do reach approval are usually well-tested and proven techniques that can be applied around the globe. Once a standard is adopted by the CCITT, modem makers begin implementing it in their products.

One interesting feature of CCITT "standards" is that they are called *Recommendations*. The CCITT cannot force modem manufacturers to comply with its procedures; rather, it recommends an approach. However, in many countries where the telephone network is operated exclusively by the government, CCITT Recommendations have the full force of telecommunications law. In such cases, all modems connected to the network must comply explicitly with the appropriate

V.32bis is expected to be formally approved by the CCITT by mid-1991. Once adopted, V.32bis will open the door even wider for very fast data transfer between personal computers. A summary of new and evolving modem standards and their status is detailed in Table 2.

### Data-Manipulation Standards

With the basic modulation rates approaching the theoretical limits of telephone-line channels, modem makers and the CCITT have turned to new ways of improving performance and increasing the data rates. The two most important steps in this direction are V.42 for error correction, and its companion, V.42bis, for data compression.

The error-correction and data-compression functions are applied to the data before modulation and stripped off before the modem receiver decodes the data at the other end. An expanded view of these functions inside the modem is depicted in Figure 2.

At high speeds, modems are prone to making more errors, not only because of the reduced power in high-speed modem signals, but also because they use the edges of the bandwidth (which tend to be noisier) to carry data. V.42, formally approved in 1988, provides error correction using the automatic repeat request (ARQ) principle.

Under ARQ, data is grouped into blocks at the transmitter, and an advanced cyclic redundancy check is applied across each block. This is the same CRC concept already used to ensure the

CCITT Recommendations. In practice, worldwide adherence to CCITT Recommendations is the norm.

To grasp the conglomeration of modem standards, some understanding of where they come from and how they are made is important. The modem standards in use today come primarily from three sources. The modulation and coding standards are normally Bell Standards or CCITT Recommendations. The interface standards are either CCITT Recommendations or Electronic Industry Association/Telecommunications Industry Association (EIA/TIA) Standards.

The Bell Standards are holdovers from the 1960s when all domestic modem standards were set exclusively by the telephone company. In those days, the telephone company

had a monopoly on anything connected to its lines, and, by law, it was the only one allowed to sell modems. As a result, it set its own design standards.

The 1968 Carterphone court decision opened the door for other manufacturers to begin making modems, and the method of standards-making changed. Since modems used in other countries at that time generally followed international standards, U.S. manufacturers became involved in helping develop those standards instead of creating a new set of standards specifically for the U.S.

Today, most new modem standards are created by the CCITT, based in Geneva, Switzerland, and affect modem users worldwide. In the U.S., modem experts participate in national standards development groups, such as the TIA, to create

those standards needed solely for U.S. interests. They also generate technical papers and proposals for the international CCITT organization and join technical experts from other countries in attending CCITT meetings.

Together, these groups work out the fine details of new international modem standards. However, when it comes time to vote on the new standards, each member country is granted only one vote. An official of the U.S. Department of State (the formal representative to the CCITT) casts the U.S. vote.

Many standards efforts never make it to a vote because of technical problems or political snags along the way. As a result, those that do reach approval are usually well-tested and proven techniques that can be applied around the globe. Once a

standard is adopted by the CCITT, modem makers begin implementing it in their products.

One interesting feature of CCITT "standards" is that they are called *Recommendations*. The CCITT cannot force modem manufacturers to comply with its procedures; rather, it recommends an approach. However, in many countries where the telephone network is operated exclusively by the government, CCITT Recommendations have the full force of telecommunications law. In such cases, all modems connected to the network must comply explicitly with the appropriate CCITT Recommendations. In practice, worldwide adherence to CCITT Recommendations is the norm.

integrity of file transfers in techniques such as XMODEM. The main difference is that V.42 provides error-corrected operation for all information exchanges, not just file transfers using specific computer software programs. Since the technique for checking the received data and retransmitting flawed blocks is contained directly in the modem itself, it is completely transparent to the user and speeds up the transfer process.

The main drawback of V.42, as with any error-correction technique, is that when numerous errors are detected, the throughput rate suffers as blocks of data are retransmitted. However, this only comes into play when errors are actually present, and even then the slowdown in the transfer rate is a small price to pay for the capability to identify and correct those errors.

Modems equipped with V.42 were originally introduced in late 1988 in V.22bis products. It is now widely available in V.32 modems as well.

### Data Compression with V.42bis

Approved in late 1989, V.42bis provides the first "official" method for compressing and decompressing data in modems. (Several proprietary compression techniques have been available for some time, the most notable being Microcom's MNP level 5 technique.)

As with V.42, the CCITT adopted a technique similar to those already in use in the computer industry when it selected a method for

**Table 2. New and Evolving Standards and Recommendations**

*The U.S. TIA and international CCITT committees continue to develop new modem recommendations. Here is a list of recent recommendations and important ones currently under development. The V.32bis Recommendation will likely be approved in mid-1991; it will provide for 14,400-bps file and data transfer over standard telephone lines.*

Standard	Purpose	Status
CCITT V.32bis Fallback Procedure	Provides a standardized way of negotiating fallback data rates from 14,400 bps down to 2400 bps.	Technically agreed upon but not yet formally adopted.
U.S./TIA Fallback Procedure	Provides a standardized way of negotiating fallback data rates from 14,400 bps down to 300 bps, including the Bell 103J standard.	Under study by the U.S. TIA TR-30.1 Committee.
CCITT V.32bis Recommendation	Provides standardized dial-up modems at rates of up to 14,000 bps; an extension of V.32.	Under study by the CCITT Study Group XVII Committee. Possible approval by mid-1991.
CCITT V.42 Recommendation	Provides standardized error correction in modems via either MNP Level 4 or LAPM (Link Access Procedure for Modems) protocol.	Approved, April 1988.
CCITT V.42bis Recommendation	Provides standardized data compression in modems via a version of the Lempel-Ziv data-compression algorithm.	Approved, September 1989.
CCITT 19.2K-bps Dial-Up Modem Recommendation	Provides standardized dial-up modem communications at rates of up to 19,200 bps.	Under study by the CCITT Study Group XVII Committee.

V.42bis. This method is a variant of the Lempel-Ziv compression algorithm, the same type of compression used in the familiar .ARC and .ZIP techniques.

However, instead of applying only to files compressed in advance, V.42bis performs automatic, real-time compression and decompression on all the data flowing between the modems. This can bring about dramatic reductions in the amount of time needed to send and receive data. For example, it is possible to achieve up to 4-to-1 compression ratios with V.42bis. That could mean effective rates of up to 38,400 bps with a V.32 modem or rates even greater than the 56,000 bps offered by digital leased-line service when used with a V.32bis modem. The advantages of reducing the time required to transmit files across a modem connection by a factor of four are obvious, especially if the telephone call is long distance.

The amount of compression that V.42bis can actually provide depends on the type of data being transmitted. Compression algorithms work by recognizing repeated patterns in data and substituting shorter symbols for them. This reduces the number of characters needed to represent a given set of information. The more repetition a data file has, the greater the compression. On the other hand, purely random data contains no patterns at all, and it is noncompressible.

Figure 3 provides a comparison of how well V.42bis works on various types of data. Assembly language and computer source code contain many short, repeated commands, since the language has a limited command set. As a result, data compression ratios on these types of files are generally quite high. Conversely, precompressed files such as .ARC or .ZIP files have already been processed to remove redundancy. Passing them through V.42bis usually does not provide much more improvement. Data files that have been encrypted through a randomization process will also show little reduction in file size and transmit time, because the data has been preprocessed to remove identifiable patterns. For the average personal computer user, however, V.42bis should reduce modem signaling time and expense considerably.

V.42bis began appearing in modem products this summer, first in V.22bis modems and later in V.32 modems. Many of the first V.32bis modems will have V.42bis compression capability as soon as they hit the market.

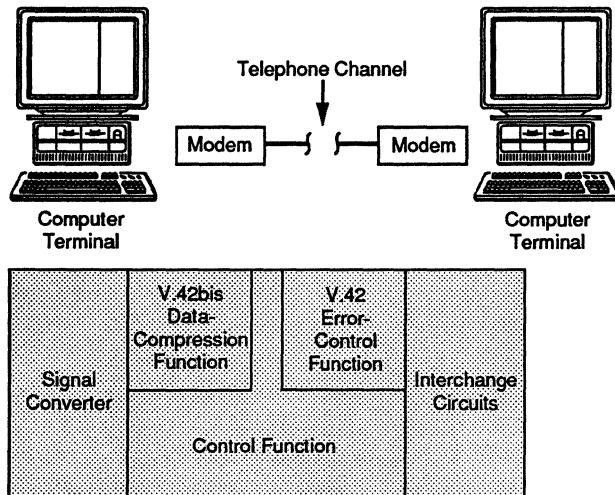
V.42bis relies on V.42 for its modem protocol and control functions. Because of this, only those modems that have V.42 will contain V.42bis. Fortunately, since V.42bis is a software-intensive technique, it doesn't require extensive modem redesign, and most modem makers are offering it in their products at a minimal increase in cost.

### Standards to Watch For

The CCITT is continuing to develop new modem standards, pushing the technology envelope a little further each time. A new effort is under way to standardize a 19,200-bps modem. Another CCITT standard currently under development will provide a uniform interworking procedure to ensure that modems implementing a number of different standards can communicate.

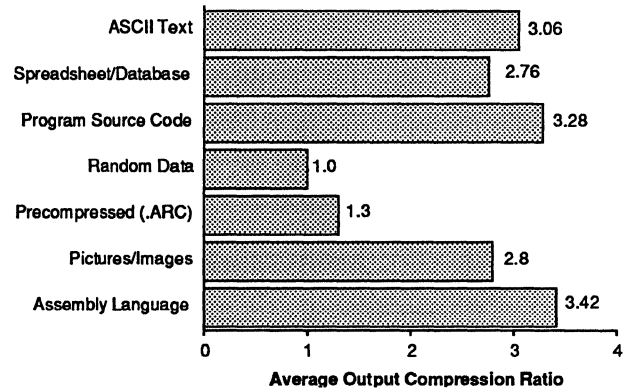
For example, if a V.32 modem calls a V.22bis modem, the new interworking protocol provides a way for the V.32 modem to identify the receiving modem's standard and fall back to V.22bis mode

Figure 2.  
Error Correction and Data Compression



Many new modems now use the CCITT V.42 error-correction and V.42bis data-compression Recommendations. In the modem, these functions are located in the overall control processor and are applied to the signal between the computer and the signal converter (modular/demodulator). Modems at both ends of the connection must have V.42 and V.42bis capability for these features to be used.

Figure 3.  
Data-Compression Comparison



The amount of data compression that V.42bis can provide depends on the type of data being sent between the modems. The more random the data, the less the compression, since truly random data follows no clear pattern. Data that has lots of repetition (such as text, source code, or pictures) can often be highly compressed. A compression ratio of 2.0 on this chart indicates that the data can be compressed by a factor of 2 and transmitted in half the time needed to transmit it uncompressed.

to match it. While many modems are already capable of this, there is no standardized format to ensure that all modems do it in the same way. The new interworking standard should improve compatibility by increasing conformity. Expect the new interworking scheme to begin appearing in modems by 1991.

Another important standards issue that the CCITT expects to take up soon involves interworking between cellular modems and regular telephone-line modems. There is currently no accepted way to guarantee that these modems can communicate, but with the explosive growth of cellular technology and the increased mobility of laptop computers, this will become a major issue in a few years. Hopefully, the CCITT will finalize a standard to solve this problem soon. ■

