Technology Reports

Data Decisions' technology reports are designed to give you the background you need to put your purchase decisions in perspective within the larger information processing picture. They tell you what products can and can't do—and explain how they'll fit into your overall operation. They also tell you what to look for when evaluating and purchasing products and how to install them with a minimum of interruptions and a maximum productivity increase.

The following technology reports include communication concepts, product and network technology on topics such as local area networks, modems and multiplexers and networking; the use of communication test equipment; and guidelines for selecting the carrier service to match user needs.

For information on the technology of voice systems, such as PBXs, key/hybrid systems, and call accounting and distribution systems, see the reports following the Technology Reports (1200) tab in Volume 4.

X.25 & Packet Network Concepts And Applications

A Summary of the Popular International Standard in Data Communication and an Evaluation of Its Application to the Problems of Corporate Data Processing

INTRODUCTION

While the technical aspects of data communication have evolved considerably since the first uses of communication lines to carry data to and from computer systems, the growth of the applied technology at the user level has been slowed by a lack of information, relating the capabilities of new products to the needs of users. Many descriptions of the internal structure of modern communication protocols are available, both from vendors who support them and from consultants who specialize in data communication, but many users find these descriptions are not helpful in either identifying areas where a new protocol can be applied or evaluating competitive products.

Data communication is not the specialty of most corporations who use it; it is a tool in the implementation of a corporate strategy for the movement and processing of information. Communication protocols represent rules for the movement of data. The benefits and restrictions associated with each protocol and each implementation of it must be evaluated in light of both the overall impact on company strategies and the quality of service users of information expect with each alternative. The purpose of this article is not to define the internal structure of either a packet network or the international X.25 standard. Such a definition is of limited value to those who wish to apply concepts, not develop them. Rather, it will present a summary of the use of packet concepts and the associated protocols in modern data communication systems and offer some guidelines for evaluating both potential applications and alternative vendors.

THE DEVELOPMENT OF NETWORK PROTOCOLS

In 1964, the RAND Corporation completed an 11-volume study on the potential for a high-reliability national network to carry both voice and data communications. The study noted that most communication users exchange data in units of about 130 characters. These data units became known as **packets**. Communication systems in which a series of computers are connected by communication trunk lines and cooperate to move packets from sender to receiver became known as **packet switches**.

Packet switching technology takes advantage of the fact that data communication users can rarely utilize the entire capacity of a connecting line. While a conventional telephone network links each user pair with a dedicated path, a packet switch provides users with dedicated paths only to a computer element within the network called a **node**. The user's node is connected to other nodes in the network by a series of trunk lines. These lines are shared among all users who connect to the node directly or indirectly to other stations in the network through the node. Each trunk line connects two nodes. The nodes cooperate in allocating trunk capacity to the user connections which the trunk must serve. **Figure 1** shows a packet network of four nodes, A, B, C, and D. The nodes are connected by the trunk lines AB, AC, AD, BC, BD, and CD. A user at point 1 in the network exchanges data with a user at point 2 through node A; the data route to node D can be established by a direct path or through other nodes. The ability of nodes to control sharing trunk lines allows data exchange at cheaper rates than can be achieved by circuit switches or leased telephone lines. The fact that any of several **routes** between users can be used to transfer information also makes packet networks highly reliable.

Packet networks do not gain such benefits without disadvantages, however. Because two users in a packet data exchange are not actually connected to each other, the network itself must provide means to receive subscriber requests for connection and to clear connections that are no longer needed. The network computers

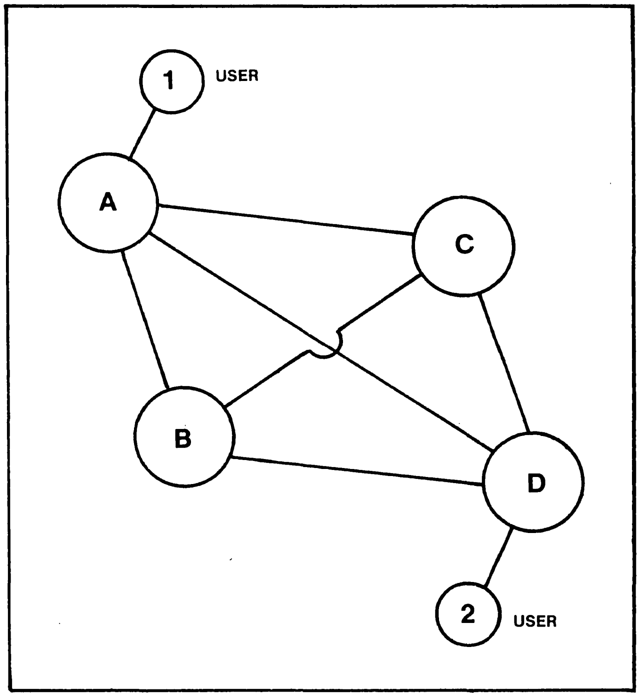


Figure 1 • packet network structure.

that send and receive messages in a manner similar to a relay introduce a delay in the passing of the message. This delay is often in the range of a half-second or more; a delay of that magnitude requires special techniques for acknowledgement of received data at the destination so the sender need not wait for an acknowledgement to make its way back. A further problem can occur when two users with dramatically different device speeds connect to each other. A small asynchronous printer operating at 30 characters per second can not keep up with a large mainframe operating at 2400 characters per second or more, so a network either prevents this type of connection or somehow regulates the flow of data between users or, as it is called, **end-to-end**.

In data communication terminology, a set of agreements on character codes, signals, and procedures that allows two users to communicate is called a **protocol**. Packet networks obviously needed a protocol to allow users to connect to the network and thus to other packet network users. The earliest packet networks, ARPANET of the Defense Department, TYMNET and TELENET in the United States, DATAPAC in Canada, TRANSPAC in France, and EPSS in England, had their own rules to connect and to request service. From the start, many network users and suppliers of communication equipment pressed for the development of an international standard to define the rules for connection to a packet network. A standard allows users to move freely from network to network without expensive changes to their communication equipment and allows vendors to build equipment applicable to any network.

Three more-or-less parallel developments occurred in the 1970s

X.25 & Packet Network Concepts And Applications

that define today's communication networks. In 1974, IBM announced its **System Network Architecture (SNA)**. This communication plan provides IBM users with software communication support in the IBM S/360/370/3000 series of processors, and in a growing family of controllers and terminal devices to support such business functions as data entry and funds transfer. SNA also defined a network protocol which is itself commonly called SNA. SNA divides the data link and network services functions into **layers**, the lowest is data-link-oriented and the highest is closely related to the user. The structure of SNA anticipated the other developments in network concepts: the **OSI Reference Model and X.25**.

The International Standards Organization began formal work in 1976 to define a model for network protocols to follow. The first step was to define all of the tasks required to permit two users to communicate across a network. The second step was to organize the tasks so that like functions were grouped together for most efficient interaction. The result was a model of seven layers, or levels, called the **Basic Reference Model for Open Systems Interconnect**. Finally published in 1978, the OSI model quickly became not only a template for the design of network protocols but a powerful tool in analyzing and understanding network functions. The names and numbers of the levels of the model have become a part of the definition of communication products from packet networks to local-area cable products. From the lowest, most **primitive** level, they are defined as follows:

- **Level 1—the physical interface, or connection.** RS-232 is an example of a level 1 interface.
- **Level 2—the data link protocol.** Such as SDLC or HDLC.
- **Level 3—the network level.** Responsible for routing between users and for control of the flow of data.
- **Level 4—the transport level.** Responsible for user-to-user error recovery.
- **Level 5—the session level.** Sets up and breaks connections on behalf of the user.
- **Level 6—the presentation level.** Provides for formatting of data so that both user and network can operate on the form most efficient for their purpose.
- **Level 7—the application level.** Or the user level.

As valuable as the OSI Reference Model is, it is a model and not a standard that can be implemented. In 1976, the CCITT (Consultative Committee on International Telegraphy and Telephony), which is a part of the International Telecommunications Union and hence of the United Nations, began work on an implementation standard for the attachment of packet devices to packet networks. Although the OSI model was not yet published, the level concept was applied to the standard for definition of the physical, data link, and network levels. In keeping with the CCITT structure for numbering standards, the packet attachment standard was given a number in the group of standards relating to digital signalling and identified by a prefix of "X" and called **X.25**.

X.25 & ITS RELATED STANDARDS

X.25 defines the way in which a packet device, called in the standard a **packet DTE** (Data Terminal Equipment), attaches to a network port or **DCE** (Data Connection Equipment). It defines the physical device interface (level 1), the link protocol, a form of High Level Data Link Control (HDLC), and a unique protocol to combine many user conversations onto a single data link (level 3).

Level 1 of X.25 is specified as a relatively new digital interface standard called **X.21**. This standard provides a relatively small number of control signal lines compared to the common RS-232 interface but gives full user control of the connection and status on call progress. Because X.21 is new and therefore not available on most equipment, X.25 permits another interim level 1 interface called **X.21 bis**, which is essentially the same as the RS-232 interface found on most communication equipment. This is currently the most popular X.25 interface.

HDLC is the link protocol for X.25. In the original 1977 publication of the standard, two forms of the protocol were supported, called **LAP (Link Access Protocol)**, and **LAP-B** for the balanced version of the protocol.

Some communication network planners felt that the LAP-B procedure had inherent potential for a condition called **deadlock**, a state where neither network nor user can properly control the line, so most users who adopted X.25 implemented the LAP form of HDLC. In 1980, an improved form of the LAP-B protocol was included as an update to X.25, and this is now the preferred form.

LAP-B is a **bit protocol**, which means that it separates user data and special protocol sequences by including a special pattern of bits. The bit pattern used by control sequences is a zero-bit followed by at least six one-bits. To prevent a user character from showing this pattern, user data is examined before it is sent and a zero-bit is inserted after five consecutive one-bits. At the other end, the zero bit is removed before giving the data to the user. The process is called **bit-stuffing** or **Zero Bit Insertion/Deletion** (ZBID), and requires special communication hardware. Units of link information in these protocols are called **frames** and are started and ended by special patterns called **flags**; a flag bit pattern is 01111110. **Figure 2** shows a sample layout of an HDLC/LAP-B frame.

The purpose of a link protocol is to exchange data over a line. To allow the line to be used for multiple data exchanges, the data/information frames, or **I-frames** as they are called, carry not only user data but information identifying and controlling the multiple "exchanges/conversations" that X.25 supports. Control and identification fields are elements of the level 3 protocol. Level 3 divides the link into a series of **logical channels**, each channel can be routed separately by the network to support separate user conversations. The user can make a data **call** on any channel, set up a temporary **virtual circuit** to another user through the network. The term "virtual circuit" indicates each logical channel serves a user in the same way as a dedicated copper or microwave path between users. Level 3 thus provides the user with the ability to combine transmissions over a trunk line within a packet network; this reduces communication costs by reducing both the number of lines from the network to a host computer and the number of computer ports.

Multilayer protocols that allow users to share a line are more easily applied to computers than to terminals. In fact, X.25 is most often used to attach computers to a network. Terminal attachment procedures are covered in another group of standards that define a facility to build packets for a device unable to meet the X.25 standard. This packet assembly/disassembly (PAD) facility offers a set of **18 parameters** to define how the terminal data can be converted to and from packets and how the terminal can request network services, such as connection to another user. An **X.3 standard** defines the PAD parameters and their functions in the translation between X.25 procedures and the terminal's native mode of operation. Another standard, **X.28**, defines the way the terminal and PAD interface, while a third standard, **X.29**, defines the interaction between the PAD as an X.25 subscriber and a real X.25 DTE at the opposite end of a data call. **Figure 3** shows the 18 PAD parameters and their use in defining the PAD process.

The PAD parameters provide a means of **tuning** the process of converting a string of user data into packets. One major issue in this process is the identification of one or more conditions to signal the end of a logical message from the terminal and initiate sending the packet to its destination. This **data forwarding signal** must be selected carefully. In certain applications, such as word processing, the computer reads each character a terminal enters and processes it immediately. Thus, a special function key may be used to delete a word because the computer interprets it as a request to do so and performs the action immediately, updating the display to the user. In a packet environment, a user may select the entry of a carriage return as a signal to forward the packet of data to the host. If the word processing application is run in this mode, the keyer may be shocked to find that the "delete word" key no longer works. The reason is that a "delete word" function is typically used in the middle of a line when an error has been made and as such is almost never followed by a carriage return, the selected signal to send the data to the computer. A frustrated typist may press the key continuously for several seconds, thinking it is stuck, and in doing so fill the packet with "delete word" commands. These, on reaching the computer, may delete several sentences. Seeking to avoid this pitfall, a PAD user may elect to forward a packet each 1/20 second so that a natural pause in keying which accompanies the use of a special function

X.25 & Packet Network Concepts And Applications

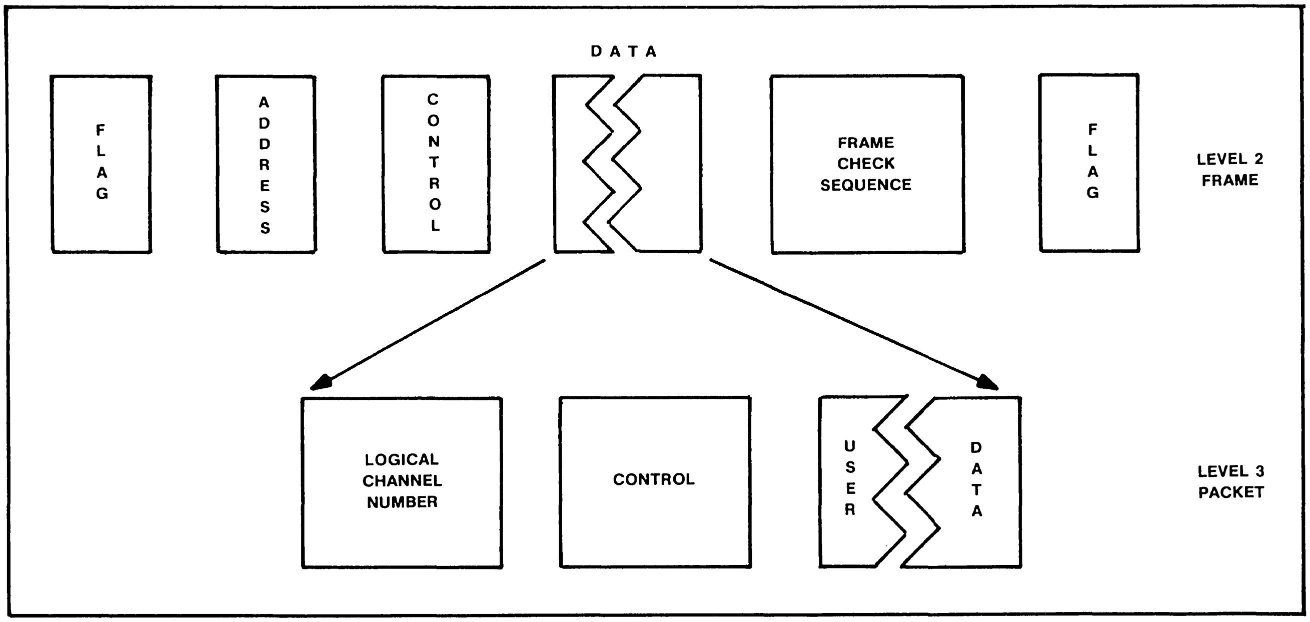


Figure 2 • X.25 frame and packet structure.

PAD PARAM-ETER	USE	PAD PARAM-ETER	USE
1	controls terminal's ability to escape to PAD control functions during a call	11	can be read by the computer to determine terminal speed
2	controls the local echo of characters sent by the terminal	12	enables the terminal to use flow control on the PAD
3	selects characters to be recognized by the PAD as a signal to forward any accumulated data as a packet	13	controls the insertion of a line feed after a carriage return is sent to a terminal
4	selects a time interval of terminal inactivity as a signal to forward data	14	controls PAD insertion of padding characters after a line feed is sent to a terminal
5	enables the PAD to flow control terminal data (XON/XOFF)	15	enables the PAD to recognize local editing functions so the operator can modify packet data entered but not sent
6	allows the terminal to receive PAD messages	16	selects the character the PAD recognizes as a "delete character" function
7	defines PAD action on receiving a break from the terminal	17	selects the character the PAD recognizes as a "delete line" function
8	controls discarding pending output to a terminal	18	selects the character the PAD recognizes as a "redisplay line" function
9	controls PAD insertion of PADding characters after a carriage return is sent to the terminal		
10	indicates the point where long lines are broken for print/display as multiple lines		

Figure 3 • X.3 pad parameters and their use.

will cause the data to be sent to the computer. With this type of data forwarding a word processor application will perform adequately, but the very small packets formed by each pause or hesitation may more than double the charges levied by public networks which charge users in part on the number of packets sent. Because of the sensitivity of the operation of a terminal with an application to the setting of PAD parameters, many users of public networks use the X.29 procedures to set PAD parameter values from the computer prior to running a program.

As difficult as selecting PAD parameter values may be, the asynchronous terminal user has at least a standard for interaction with a packet network. Users of terminals that conform to IBM's binary synchronous or **bisync** communication protocol have no such PAD standard. The CCITT as a matter of policy will not consider a standard that represents a vendor's proprietary protocol. This rule, if it continues to be applied, will also prohibit the standardization of a connection between a public packet network and IBM SNA. Some public networks have developed bisynchronous interfaces outside the standard to provide additional connection potential and therefore increase revenues. TELENET, TYMNET, and DATAPAC have agreed in principle to adopt a common standard for bisynchronous attachment called the **BPAD**. This permits users of the IBM 3270 display system or the 2780 or 3780 batch terminals to control these devices through one of the three networks from a packet host. Users of one network can even connect to terminals/hosts on another network; the common BPAD standard assures proper operation.

Interconnection of networks is another requirement that extends beyond the attachment of computers covered by X.25. Computer-to-network interfaces can have an implicit master/slave relationship, but two networks cannot. Also, station **addresses**, the packet equivalent of phone numbers, may require translation when network boundaries are crossed to prevent misrouting through address duplication. The **X.75** standard provides a protocol definition for **inter-network gateways**.

■ WHERE X.25 IS COMMONLY USED

The primary use of X.25 is to connect a computer to a public packet network. Most public networks charge for service based on a flat fee for each connection, or **port**, into the network plus a

X.25 & Packet Network Concepts And Applications

graduated system of charges per data packet sent. Because X.25 provides a user with a protocol supporting multiple data conversations on a single link, a computer connection through an X.25 port can support the same number of logical channels as several PAD ports into the network but at a lower cost. The single X.25 link will also probably cost less for the communication interface hardware on the user's computer.

Another example of this line-sharing principle is found in private communication networks built by users with lines leased from AT&T or other public carriers. These networks, used in such applications as airline reservations or credit authorization, may connect hundreds or even thousands of terminals to provide very quick access to a computer. In many cases, the total volume of information sent and received is very small. Such private networks may require expensive hardware additions to supply communication port connections to attach all of the many communication lines to the computer. A single front-end processor with one or more X.25 links to the computer can replace all of the individual terminal connections. Here X.25 is not required by the network, a private line can support any communication protocol the user requires. Instead, it is used as a concentration protocol either to save on the cost of computer ports or to permit connection of more terminals than a computer can physically attach.

X.25 can also be used to connect two computer systems together. Many minicomputer users are integrating their data processing operation to provide both users and programs access to databases resident on different computers, which may or may not be co-located. Because X.25 is an international standard, it is supported by most of the major computer companies and can be used to link computers of different speeds, types, and manufacturers. The multiple-channel capability of X.25 allows users to exchange data between several tasks at the same time and still support terminals on one system accessing programs or data on another. X.25 is **full-duplex** and can support simultaneous transmission in both directions, reducing interference between users.

Certain very large users, such as universities or research centers, can combine many of these uses and create a private packet network using special communication processors called **packet engines**. These devices function in the same way as the nodes in a public network, making connections and routing data among the various computers and terminals. X.25 can be used to link the computers to the network and to communicate between the packet engines while X.3 and its companion standards can be used to attach terminal devices. Such a network offers free connectivity between users and provides the ability to define **closed user groups** to restrict data access to sensitive applications, such as financial or classified data.

A company preparing to evaluate the possible benefits of X.25 should first classify its intended use of products or services based on it. The following questions form a guide in this evaluation.

• Will connection to a public packet network such as TELENET, TYMNET, or UNINET be a part of the communication application? If it is, a greater degree of standardization of the X.25 support from each vendor will be needed, because public networks require that the equipment used be **certified** for network attachment.

• Will both terminals and computers be involved in the application, or just computers? Applications involving terminals will require consideration of a PAD.

• Do the programs that send or receive data over the X.25 connection already exist or will they be written for the application? If the programs are to be written, they can take advantage of the features of both X.25 and the PAD with little additional effort. Existing programs may have to be modified at additional expense.

• Are the computers and terminals to be used already installed or will new equipment be purchased? New equipment can be purchased for X.25 compatibility. Current hardware may require special attachments for X.25 support, if such support is available at all.

• Does the company have personnel with data communication experience, especially in computer networks and X.25 itself? The implementation of an X.25 application can be difficult without staff experience to support it. Users without this experience should evaluate vendor ability to provide support as a key element in selection of services or equipment.

• Will computers be directly connected to each other via X.25 or only through a public or private packet node? All X.25 communication packages will not support direct computer-to-computer connection.

• If a private packet network is planned, will terminal-to-terminal connections be necessary? Some **packet engines** will not support a connection between terminals, but only between a terminal and a computer.

• What areas of geographical and volume growth are projected over the next five years in the application involved? Public networks cover varying geographical areas and rate structures may favor some areas over others. In addition, very large data volumes may cost-justify a private network. Growth from a limited to a large data volume indicates a possibility of migration from a public network to a combination of public and private packet communication. This possibility should be related to vendors involved to insure compatibility.

• Are there any related applications which may become candidates for sharing the communication facilities? Many users find that communication systems attract secondary applications, which may have different requirements. Potential secondary applications should be screened for effects on data volumes and for differences in functional requirements.

When answers to these questions have been obtained, the prospective X.25 user should review the available X.25 products and services to assure a match in needs and features. This task is especially critical for users who do not have extensive data communication experience within their organizations. The major areas to consider are the type of public network available, the X.25 capabilities offered by the computers themselves, the need for secondary X.25 products, such as test equipment, and the possible impact of the use of X.25 on the business itself.

■ PUBLIC NETWORK SUPPORT OF X.25

Several active X.25 public networks are available in the United States, several more are planned, and several specialized communication services related in some way to X.25 are sold. These services are often classified as **Value-Added Networks**, or **VANs**, to distinguish them from public carrier services providing only a circuit connection, such as the phone companies. Most of these networks offer the general capability of connecting user equipment and terminals, widely geographically disbursed but not heavily utilized, at a reasonable charge. Such networks use a different basic charge structure than the telephone company: connection rates are based on usage and not on the distance between the communicating parties. Although users associate such networks with X.25, most do not use X.25 for any purpose except to connect users who support that protocol; in fact, connections to these networks are more often made via PAD than X.25.

Public network users normally connect both computers and terminals to the network. Host computer connections are normally made through dedicated ports, often through a small communication device leased from the network to provide one or more ports for computer attachment. This interface is typically operating at a high speed (9600 bps or more) and can use any protocol, including X.25. Because X.25 requires less processing to convert it into even a proprietary internal packet protocol used by public networks, X.25 host attachment is often considerably less expensive than attachment using another protocol. Terminal attachment is still predominantly through the asynchronous PAD, and can be made with a dial-up connection into public ports, dial-up connection into a private port, or a dedicated port.

Public networks have not achieved the popularity in the United States that they have in other parts of the world. In European countries, most long-distance business data communication travels on such networks; in France, there is a move to give the TRANSPAC network a monopoly on such traffic. The entry of AT&T into the packet network market with its Net 1000 service indicates that packet switching use will grow considerably in the near future. Established VANs are competing to offer better facilities and new services to attract users.

X.25 & Packet Network Concepts And Applications

■ TELENET

TELENET, a division of GTE, is one of the best known of the VANs. Established in late 1973 by Bolt, Beranek, and Newman (best known as the contractors for the first packet network, ARPANET), TELENET offers service in over 300 cities in the U.S. and interconnects with other public networks in over 50 foreign countries. Like most of the public packet networks, TELENET first offered asynchronous communication service only, and gradually grew to support both X.25 and IBM binary synchronous (BSC) protocol connections.

Terminals attach to TELENET through TELENET Central Offices (TCOs), which are actually city locations classified into three categories according to traffic density, data rates, and/or type of traffic supported. Class A cities support dedicated users at rates up to 9600 bps in most locations and dial-up users at rates up to 1200 bps; Class B cities support dedicated users at rates up to 1200 bps in most locations, with a few locations supporting rates up to 9600 bps, and dial-up users at rates up to 1200 bps; Class C cities support public dial-in users only at data rates up to 1200 bps in most locations, with a few locations supporting rates up to 2400 bps. Dedicated facilities are also available to support private dial-in users. Most terminal users, however, elect to use the public dial-up ports. This access currently costs between $0.10 to $0.57 per minute plus $1.70 per thousand packets transmitted (packets received are billed to the station transmitting them). A dedicated port can be leased for a fee ranging from $425 to $750 per month; leased lines (dedicated access facilities, or DAFs) provide a permanent connection and are available from $600 per month for a 110- to 1200-bps line to $2,000 per month for 14.4K bps. Per-packet costs remain the same for all types of traffic, but discount plans are available.

Clusters of user terminals or a host computer must attach to TELENET through a TELENET processor (TP) unless the user device has an X.25 interface certified by TELENET. The TP Series can attach from 4 to 116 asynchronous ports, up to 16 IBM BSC ports, and up to 32 HDLC/X.25 ports at costs from about $800 per month to $2,300 per month. Some models will also perform local packet switching so that connected terminals and computers can communicate with one another without entering TELENET and paying usage charges.

Vendors of X.25-compatible products must apply for certification for their products to assure they comply with the standards of TELENET. There are now over 100 such certifications, and the rate of certification is growing annually. TELENET will supply users with a list of certified vendors. A host computer or communications front-end that is TELENET certified can attach directly to TELENET without requiring a TP interface, a savings of at least $1,500 per month.

■ TYMNET

TYMNET was originally the network portion of TYMSHARE, established in 1969. It became a public VAN in 1977; it offers the same basic service as TELENET with some technical differences in the network architecture. One difference between the two networks is that TYMNET packets can be made up of data from multiple users which happens to be sharing a part of the route. Packet sharing allows greater transport efficiency. TYMNET thus bases its usage charges on characters transmitted rather than on packets transmitted. A TYMNET charge of $.04 per thousand characters is equivalent to TELENET's charge of $1.70 per thousand packets if the average packet size is 25 characters. Selection of PAD parameters for optimum packet size is therefore less important with TYMNET than with TELENET.

TYMNET rates for terminal attachment depend on the "density" of the user service area. In high-density areas, the access charges vary from $4.25 per hour for low level use to $2.00 per hour for high-level use. Medium- and low-density cities incur a $2 and $5 surcharge, respectively, over high-density rates. Charges for private dial-up ports vary from $250 per month to $400 per month depending on the city density and data rate. Private port users still pay access charges identical to public port users, but quantity character discounts and off-peak time traffic charges are available to all users on a standard basis.

TYMNET provides asynchronous/synchronous terminal access and host access through a **TYMNET engine** that supports from 8 to 126 user ports. Engines can be purchased, leased, or rented; rental charges range from $600 per month to $2,700 per month depending on number of ports and protocols supported. Leased network connections are charged at rates ranging from $800 per month to $1,500 per month depending on the data rate (2400 bps to 14.4K bps). Per-character costs are the same as public access, but TYMNET also provides discount plans based on connect time or traffic volume for all types of traffic. Again, dedicated users with their own TYMNET-certified X.25 interface attachment will save considerably over the lease of an "engine," although the vendor recently made them available on a purchase basis. To date, over 50 non-TYMNET products have been certified for operation over the network.

TYMNET covers much the same geography as TELENET, and the services offered are similar. Users deciding on which network to use should evaluate the differences in the pricing and its effect on the user's total charges.

■ UNINET

The third major VAN, UNINET, was originally the private packet network of United Computing, a subsidiary of United Telecom. It consists of a series of switching centers linked by 56K-bps trunk lines of AT&T-IS Dataphone Digital Service, with smaller service centers linked by other DDS links. Synchronous users connect directly to a node, while asynchronous terminals and hosts can connect to a Local Node Asynchronous with 8- to 64-channel capacity.

Fees for interactive terminals on UNINET are not based on traffic but on connect time. Batch terminals such as IBM 2780/3780 or host computer exchanges are charged more than interactive terminals for access and are assessed 1.25 cents per thousand data bits transmitted. Access charges to public dial ports for interactive terminals vary from $2.25 per hour for high-density areas and 300-bps data rate to $11.50 per hour for low-density areas and 1200-bps data rate. Batch terminals are charged $4.00 to $8.00 per hour.

Host access to UNINET is through an interface processor that supports from 12 to 128 ports and leases for $1,500 and $5,500 per month. X.25 access is available at rates based on line speed. A 2400-bps X.25 interface is $850 per month while one at 19.2K bps costs $2,600 per month. UNINET covers approximately 200 cities at present, less than TELENET or TYMNET, but its rate structure again offers advantages to certain types of users, particularly those with interactive terminal traffic.

■ AT&T ACCUNET PACKET SERVICE

AT&T Communications has filed with the FCC for a new tariff on its ACCUNET Packet Service (formerly BPSS for Basic Packet Switching Service), a public packet transmission service. AT&T considers ACCUNET packet service to be a carrier service rather than a VAN, because regulations forbid AT&T to offer any value-added service as a part of its regulated operation. BPSS has become part of AT&T Communications ACCUNET Digital Services. Analog and digital access is available at 4800, 9600, and 56K bps. The network supports the CCITT X.25 standard and is slated to add X.75 to interconnect networks. ACCUNET Packet Service is available as a backbone transport to resellers such as VANs as well as corporate users. Although the original rates were set too high and not competitive with other carriers, discouraging usage, the new tariff cuts the rates substantially and will put the service in line with its competition.

■ OTHER PUBLIC VANS

ADP AUTONET is a network connecting about 130 cities; it provides public dial, private port dial, or dedicated service. Host computers with X.25 interface can connect directly while non-X.25 hosts require an asynchronous facility based on the number of ports required. Public service began in early 1982.

COMPUSERV is a network offering dial-in terminal access in 130 cities. Host connections are through a micronode that costs from $1,000 to $14,000 depending on port characteristics. An X.25 interface is available at a monthly rate of $1,000 and supports up to 61 connections.

X.25 & Packet Network Concepts And Applications

GRAPHNET is a combination packet switch and message service.

ITT FAXPAK is a network serving data facsimile users.

■ HOST COMPUTER SUPPORT OF X.25

Support for X.25 among computer manufacturers is growing, and computers from micros to mainframes support X.25. **Figure 4** shows the vendors that either currently offer or plan to offer X.25 support.

All X.25 packages do not provide an equivalent level of user support. Even though the protocol it standardized, variations in the features available can conflict with the requirements of public networks or other vendor products. Buyers of X.25 packages should evaluate the products based on those basic requirements already defined, and should check specific host features for restrictions and level of support:

• **number of logical channels which may be assigned on a link.** A small number restricts the number of terminals or other users the line can support.

• **maximum speed of the line.** Attaching to a network at a higher speed will provide improved performance levels or allow more users (within the restrictions of logical channel assignment noted above) for the same level of performance.

• **support of the X.29 PAD control protocol.** Special packet types are used to control a terminal attached via a PAD. Host support for X.29 will simplify user programming.

• **support of optional facilities.** X.25 has a number of special features, such as closed user groups to provide restricted access to a terminal or host computer, may not be supported by all computers.

• **operating system and application support.** Not all operating systems available for a computer system may support X.25. Specialized application packages, such as database query languages, may not support users connected via X.25.

• **memory and other matching requirements.** Some computers will require special hardware or expansions to support X.25.

• **certification.** Many public networks allow user-supplied host computers or communication processors with X.25 capability to substitute for network interfaces, which are otherwise leased from the network supplier. In most cases, the operation of these devices must be **certified** by the network. This is normally done by the system supplier and not the end user, but it must be done for the device to qualify for attachment in place of a network interface. Certification is also a good way to verify the ability of two devices to communicate: if both are certified against the same network they have a greater chance of being compatible.

• **selection of station type.** As already noted, X.25 divides stations into two types, DCE and DTE. Most computers are assumed to be DTEs while the network is the DCE. If it is desired to connect two computers via X.25, one must be able to act as a DCE.

• **address length.** International standards for addressing within public networks permit addresses up to 15 digits in length. Users of X.25 who wish to connect via gateway from a domestic public network to an international network must be able to accommodate as many address digits as the called network and international routing conventions require.

• **level of support.** Any X.25 implementation must support the CCITT recommendations, 1980 revision. There are nearly always exclusions in areas of optional facilities, but both level 2 HDLC and the basic level 3 functions are required. Terms such as "based on," "similar to," or "at level 2" in qualifying support may mean that the full standard is not implemented.

■ PRIVATE PAD FACILITIES

Just as it is possible to purchase host computers with X.25 access capability, it is possible to purchase devices which can act as a PAD for the connection of multiple asynchronous terminals. These devices must also be certified against the network to which they attach, and they normally have no internal processing capabilities, but they allow a user to attach multiple terminals at what is often a significantly lower cost than the equivalent direct attachment of the devices.

A typical private PAD is a small interface unit capable of fitting on a shelf or desktop and connecting up to 8 asynchronous lines to a public or private X.25 network. The link to the network is usually a single line operating at 9600 bps or higher. It also can provide a supervisory terminal connection so the user can change the PAD parameter values on each attached terminal or to display some statistics on the operation of each.

The major question on the function of private PADs is the way they appear to the X.25 network. If the PAD appears as a packet DTE such as a computer, but does not inform the network that it is operating as a PAD, it need not support full PAD features or it may offer extensions to PAD functionality. It may not respond to X.29 control however, from the packet DTE to which it connects. This will prevent host control of PAD operation. If multiple applications are accessed from the terminal, the computer will be unable to condition the PAD for each application, forcing any parameter adjustments to be made by the terminal. For PADs to appear to the network as PADs, they will normally be identified as **supporting full X.3/X.28/X.29 capability.**

Another variation in private PAD devices relates to the X.25 functions available to the PAD. This applies primarily to PADs that appear as such to the network, because a packet DTE is assumed to have full access to X.25 features. Some networks, and some PADs, will not permit a PAD to receive a call, to use a permanent virtual circuit, or to call another PAD. Information on such restrictions can sometimes be obtained by reference to the device's documentation, but often requires specific questioning of the supplier.

Some specific features of private PADs should be researched by prospective buyers or users:

• **number of asynchronous devices supported.** If the PAD supports relatively few connecting devices, it is more difficult to recover its cost in savings in public network attachment charges.

• **speed of the network connection.** A high-speed line connecting the PAD to the network will carry more traffic and can

COMPUTERS/CONTROLLERS		SOFTWARE OR COMPONENTS	OTHER EQUIPMENT
Bolt Beranek & Newman	Northern Tel	National Semiconductor	Atlantic Research
Burroughs	Perkin-Elmer	Pacific Software Mfg	Codex
Com-Pro	Prime	Western Digital	DCA
Control Data	SESA/Honeywell		Dynatech Packet Tech
Data General	Siemens		Gandalf
DEC	Tandem		ICOT
Harris	Telenet		Infotron
Honeywell	Tymnet		Memotec
IBM	Univac		Spectron
NCR/Comten	Varian Assoc		

Figure 4 • vendors supporting or committed to support X.25

X.25 & Packet Network Concepts And Applications

improve performance, but the speed of the link is not normally the major source of delay. Most PADs provide a network connection at speeds to 9600 bps.

• **input device speed limitations.** Some PADs will allow asynchronous devices to connect at speeds to 9600 bps, while others will limit them to 1200 bps or even less. Private PAD devices may be the only way to attach high-speed terminals to public networks that do not support terminal access at such speeds.

• **ratio of sum of input speeds to network connection speed.** An X.25 link can normally support many asynchronous connections of the same speed because the terminal devices involved are rarely used at full capacity. A typical CRT, for example, is capable of nearly a thousand characters a second transmission rate but may average less than 10. Most units will allow 4 to 8 times the network connection speed for the sum of the input device speeds.

• **certification.** Private PAD devices must normally be certified in the same manner as computers, but not all networks certify the PAD portion of the logic with the X.25 interface. The technical documentation may not provide assurances that the X.29 interface is standard. Many vendors will furnish a list of users from which similar applications can be researched, but where this is not possible, specific assurances should be obtained on the applicability of the device to the application intended.

■ NON-ASYNCHRONOUS CONNECTION

Although most terminals are asynchronous, many users have terminals that use various synchronous protocols such as IBM binary synchronous, Honeywell GRTS, or CDC UT-200. Some very specialized asynchronous protocols such as NCR-279 or the asynchronous form of Burroughs Poll/Select require that the terminal devices be **polled** or periodically interrogated by their host computer for traffic. These protocols cannot be handled by the X.3 PAD function because the control of the devices is more complex. Also, they represent a specific vendor's product, and CCITT will not develop a standard for their attachment to a packet network.

Users with non-PAD terminals or host computers can attach to a public network in several ways. First, the network itself can support access by the terminal type involved even though there is not an accepted standard for such attachment. Most public networks will allow IBM 3270-type terminals or **batch** terminals such as the IBM 2780 or 3780 to attach through their native BSC protocol. The degree of support of this attachment varies with the network; some will permit the BSC terminal to be controlled by a BSC host or by a packet host while others may require that a packet host control the terminal. Attachment via a non-standard network feature is convenient, but a user requiring inter-network call support may find that the destination network does not support the special interface.

Another way to connect non-PAD terminals is to attach them to a host computer that supports X.25 and allow the computer to serve as the network gateway. The user program in the host computer to perform the required switching function is a task similar to writing a simple PAD interface. Some computer communication access methods and some front-end processors will have features to simplify this task.

■ OTHER X.25 PRODUCTS

Most users will find that either X.25 computer support or a private PAD are the only X.25 products required to support their communication applications, but for users with more complex networks there are other X.25 products available:

• **private packet engines.** Provide the same features as public networks but in an environment more economical to those with very high data volumes.

• **front-end processors.** Give users X.25 network access on computer systems with either no X.25 capability or with limited capability.

• **concentrators.** Combine many different types of terminal or computer communication protocols into one or more X.25 links.

• **multiplexers.** Function as private PAD devices and provide a "gateway" from a user-defined network based on switching multiplexers into an X.25 network or into a computer supporting X.25.

• **data analyzers.** Display the X.25 protocol itself in a form that can be easily analyzed and to point out problems or violations in procedures.

• **protocol testers.** Function as active X.25 devices by simulating a network connection or a user DTE to permit testing the X.25 products or diagnosing problems.

• **protocol converters.** Allow non-X.25 devices to connect to X.25 networks or computers supporting X.25.

• **software packages.** Provide X.25 support for any mainframe, minicomputer, or microcomputer that has hardware to transmit and receive bit-oriented protocols, such as HDLC or SDLC.

• **integrated circuits.** Provide partial X.25 (level 2) support on a single chip.

■ APPLYING X.25 TO A BUSINESS ENVIRONMENT

Even the most enthusiastic supporters of X.25 do not claim it is the best solution or even a good solution for all user data communication problems. There are many uses for the X.25 capability. Several primary uses have already been discussed, and these will serve as the environment for discussion of the impact of X.25 on user communication costs, application programs, and long-term communication strategies.

■ COST CONSIDERATIONS

The most difficult aspect of evaluating the cost and savings associated with X.25 use is determining what elements of network cost are actually affected by the use of X.25. Public network access is possible without X.25, as is computer-to-computer communication and protocol conversion. Defining the financial advantages of X.25 requires accurate costs for providing the service required by the user with X.25 and without it. It is important to assure that the level of service and limitations in features under both alternatives are comparable; differences are highlighted for review and assignment of values for comparison.

X.25 is normally a substitute for multiple connections of another type. For example, a single X.25 connection to a public network might serve 30 remote terminals, replacing 30 local computer ports connecting to the network or one network engine. The following sample lists cost elements that make up the total cost of a connection with and without X.25:

WITH X.25—

• **cost of the X.25 feature.** Both hardware and software.

• **cost of the X.25 interface.** To the network.

• **cost of the high-speed line connection.** 9600-bps line.

• **cost of modem(s).** On the line.

• **cost of programming.** For the X.25 interface.

WITHOUT X.25—

• **cost of the alternative protocol handler.** For the computer.

• **cost of additional ports.** For multiple lines.

• **cost of connecting ports.** To the network

• **cost of modem(s).** For the lines.

• **cost of programming.** For the alternate interface.

• **cost of a network interface.** For the host.

Each cost should be determined through consultation with the vendor or network operator; it should be fully discussed to insure that the conditions under which it is valid are representative of the conditions in the network. Users planning to replace existing network connections with X.25 should verify any assumptions about equipment, such as modems and computer ports to retain under X.25. Equipment designed for operation with asynchronous or binary synchronous terminals may not be compatible with the X.25 link protocol because of its full-duplex operation. Users with network technical control equipment or data line monitors should also insure they will operate in an X.25

X.25 & Packet Network Concepts And Applications

environment. If not, include the cost of their replacement in the comparison.

When X.25 is being considered as a local connection protocol to connect a computer to a front-end processor or another computer, the cost evaluation consists of comparing the cost of an equivalent data rate connection in another protocol with that of X.25. Here, it is important to be sure that the connections are actually equivalent. X.25 is a **full-duplex** protocol. Other protocols such as IBM's binary synchronous are "half-duplex." In addition, protocols such as BSC are not inherently able to support multiple simultaneous users and can require additional custom programming to do so. On the other hand, X.25 may not be available for all the user's computers.

Users may consider developing X.25 support for their computers rather than purchasing it from the computer manufacturer. This is almost certain to be a costly undertaking; industry experience places the development time for X.25 support in terms of multiple man-years and development costs in excess of one hundred thousand dollars. Modern protocols such as X.25 and IBM's SNA are very complex and development of custom versions is not normally practical. Users with special requirements should consider purchasing a vendor package and custom modifying it, assuming the source program statements are available from the vendor. Some software suppliers will make X.25 source programs available for user modification. Implementation cost will be lower than for full user development, but it is also normally more expensive than buying X.25 support from the computer manufacturer.

■ IMPACT ON EXISTING APPLICATIONS

The use of X.25 as a communication protocol may not be transparent to the normal business data processing applications that run on user computers and use data communication lines as a source of input or output for information. Many programs that support asynchronous terminals are written to include the terminal handling within the program itself. X.25 communication support requires a different type of handling in many cases, and changes may be required to the programs to use X.25. This is particularly true with minicomputers without special communication packages, the equivalent of IBM's VTAM or TCAM. Telecommunication packages can provide a user interface less specific to a particular communication protocol than a user-written handler will be. Another source of application impact is the possibility that X.25 support may not be available on the operating system currently in use. Many vendors who offer X.25 support will not do so on all their operating systems. Although most vendors attempt to provide users a migration path from one operating system to another, it is not normally possible without some level of application program changes.

X.25 usage in any form implies that asynchronous connections to the computer will be via a PAD. Although it is possible to define a set of PAD control parameters to permit most applications to operate satisfactorily, some users will find certain programs will not run efficiently or may fail to run at all. The primary problem area is the process that forms packets from user data streams. Each application program expects the user to key some sequence of characters constituting a request or message. The computer reads the message character-by-character but acts on it only when it has all been received. A message can be a line of text, or it can be a single special keystroke that invokes a function such as "delete word." The PAD must establish boundaries when forming packets.

A packet must be sent if it is full of data. Some partial packets should be sent as well, because they contain an entire message or the end of a message, and the computer will not operate on the data until it is complete. Essentially, this means that proper operation of the application program with the terminal over a PAD/X.25 connection depends on whether the **packet** boundaries are set so the end of a user **message** is not held by the PAD in a partially filled packet.

For example, if a user message consisted of an account number, an amount, and a carriage return, the data will probably not fill a packet. Most networks set at least 128 characters as a full packet's size. Unless the user defines the PAD parameters so the entry of a carriage return or the elapsing of a specified time interval causes the packet to be sent, the operator will wait indefinitely for a computer response. Normally, a combination of time interval and special character such as carriage return allows a terminal to access an application program without modifications.

Some applications require a more-or-less continuous entry of data; waiting for the expiration of a time interval slows the production rate of the operators. If these applications end each message with a character such as carriage return, the timer method of causing a packet to be sent can be eliminated. If the messages terminate on some other characteristic of the data such as message length, however, it may be necessary to change the application so it operates properly.

For example, if a transaction keyed into an accounting system requires the entry of an account number of ten digits and an amount of six digits, the transaction will not be forwarded to the computer because it does not terminate with a recognizable special character (such as any of the ASCII control codes). Changing the application program to accept a carriage return as termination of a transaction allows the setting of PAD parameters to send a packet when a carriage return is entered. Changes of this type are often quite simple; they can be made while adapting the application program to operate with any differences in communication procedures in X.25. The cost of the customizing effort must be considered.

■ HIGHER OSI LAYERS—THE MISSING LINK TO THE USER

The OSI reference model defines 7 layers of communication functions to support the connection between users through a network. X.25 implements only 3 of the 7; so the functions normally performed by the higher levels must be either absorbed into the user programs or evaluated and put aside as unnecessary in the user environment.

Level 4 of the model is the transport layer, and it normally provides what is termed **end-to-end assurance**. This means the transport level operates at each user connection to control the flow of data and to retransmit data lost due to the failure of a line or node in the network. This level of error protection is beyond that offered on a line-by-line basis by level 2; level 2 cannot recover from a line or node failure that causes loss of data after it is received at a node but before it is sent to the user.

Level 4 also can ensure that a connection is not broken while data still remains in the network. Without level 4 support, users must realize that some X.25 connections can lose data if a node or line fails and the connections must be rerouted. This is not normally the case with public packet networks; they provide this service to users without special user programming, but it may be true for private networks. Users can perform level 4 recovery functions within their own applications during computer-to-computer communication over an X.25 network. Terminal PADs will not provide this ability in most cases, and it is possible that private X.25 networks with multiple nodes can lose data if a node or line fails.

Data can also be lost if a user sends a request to "clear" a connection while data is held in the network. There are various ways to prevent this, even when a PAD is used, but if one of these methods is not employed, several packets can be lost at the end of a transmission when a user clears the connection immediately after sending data.

Session services, level 5, is normally used to make and break connections on behalf of the user. Its primary benefit in most environments is to simplify the task of calling another user and breaking the connection when data exchange is completed. Most users will not find it necessary to implement custom logic to make up for a lack of session services in X.25.

Level 6 offers presentation services. Without this level, each device that accesses the network must exchange data in the code set and format native to the device. In protocols with a level 6, such as IBM SNA, a common form can be defined for exchanging data between, for example, a computer and a terminal. This "virtual terminal" format may not relate to the form of any of the actual devices, but level 6 will perform translation to and from the format for each device type on the network. X.25 has no such level for conversion, and connections between devices with

X.25 & Packet Network Concepts And Applications

incompatible data formats are impossible unless user programs provide for it.

If a user has, for example, a DEC VT-100 CRT and a Hazeltine 1500 connecting through a PAD into the same computer X.25 connection and both ran under control of the same application program, cursor-positioning commands cannot be used by the program; the two terminals do not respond to the same form of cursor control. Likewise, the same program cannot service both an ASCII terminal, such as the VT-100, and an IBM terminal, such as the 3270 which uses the EBCDIC character set; data patterns for one are incorrect for the other. The program must select the proper control procedures and code set based on the identity of the calling terminal, a form of user-provided level 6 presentation service.

■ FUTURE DEVELOPMENTS IN X.25

The selection of a communication protocol or a network architecture should never be made without considering the future of the selected technology. In the case of X.25, the developments likely to shape the form its use will take are the evolution of the international standards themselves and the growth in the new products and services to support X.25.

■ THE PROGRESS IN STANDARDS

X.25 itself may be updated, as it was in 1980, to correct small problems, provide clarifications, or define additional features. Extensions, if the example of the 1980 revisions continues to hold true, will apply primarily to interaction between true X.25 devices, rather than between packet DTE and PAD. It is unlikely to undergo major changes in the near future. The addition of functional levels of the OSI model to X.25 can come as an amendment to X.25 itself or as an additional X-series standard that builds on X.25's three-level base.

The International Standards Organization is considering a proposal for the definition of level 4 and 5 standards which originated with the National Bureau of Standards in the U.S. and the European Computer Manufacturers Association (ECMA). Copies of the proposed standards can be obtained by interested users or systems developers from the National Bureau of Standards. The proposal for level 4 establishes end-to-end error and flow control, and level 5 facilities can combine multiple network connections to gather data for a single-user connection. Even a positive action on the proposal will not bring these facilities to the user immediately; even limited support will probably require at least a year to appear and several years to become widespread. It is significant, however, that such a proposal has been made. It indicates a full implementation of the ISO model based on international standards will be available in the future, providing an alternative to SNA for networks supporting distributed processing.

■ NEW PRODUCTS & SERVICES

The increased acceptance of X.25 is reflected in a growing interest in it among the developers of communication products. Many announcements of new services and new products in the area of both software and hardware have been made.

■ NET 1000—AT&T-IS' NEW VALUE ADDED NETWORK

One of the most significant developments in value-added networks (VANs) is AT&T Information Systems' Net 1000 Service, formerly known as Advanced Communications Service, or ACS. ACS was first announced in the late 1970s as an **ultimate** network, connecting users regardless of device type, speed, code set, or protocol. It was planned to be a proprietary implementation of all levels of the OSI model; users could have the relative connection freedom available to a limited set of SNA-compatible devices, but expanded to include all major devices and protocols.

ACS generated a great deal of user interest and much consternation among the operators of VANs. Rumors of technical problems and legal difficulties with the entry of AT&T into this new area surrounded the service, and it seemed to fade by 1980. In late 1981, AT&T filed with the FCC to form a separate corporation to market non-traditional products and services such as ACS, and a new ACS with a more limited scope that its predecessor announced. The new ACS was named Net 1000 Service and was made available for use at the end of the third quarter of 1983.

Net 1000 is a value-added, packet-switched network that, besides supporting interactive and message communication, provides shared storage and processing support for customer programs and data. It is a flexible resource which supports customized user applications through user-created application programs or programs made available by AT&T-IS or independent organizations.

High throughput nodal processors are connected by channels routed over the ACCUNET Packet Service, formerly called Basic Packet Switching Service (BPSS). Within the network, the nodal processors use the X.25 protocol (finalized in 1980 by the CCITT) as the standard transmission control procedure. X.25 contains an error-checking system that ensures that the message sent by one node is identical to the one received by the adjacent node. The network is designed to be self-monitoring to minimize customer involvement in trouble reporting. According to AT&T, major network components will be duplicated to minimize system malfunctions.

Net 1000 defines two types of data communication service, Interactive and Message. Interactive service is similar to the type of user connection offered by packet networks. Two users can communicate with each other through Net 1000 and the VAN will perform protocol conversion to permit the connection of devices with unlike characteristics which ordinarily bars communication. Net 1000 is expected to support most popular communication protocols; however, ASCII and BSC (3270 and 2780/3780) are the only supported protocols initially. SDLC/SNA support is planned for the future to meet customer application requirements.

Net 1000 supports an array of terminal and host interfaces which include: asynchronous contention (Teletype Model 33), synchronous contention (IBM 2780/3780/3725), and synchronous polled (IBM 3271, 3274 BSC, and 3276 BSC). An initial SDLC terminal interface will also be provided. Host plug-compatible interfaces include asynchronous contention, synchronous contention (IBM 2780/3780), host synchronous polled (IBM 3271 BSC), and an initial SDLC. Within the network, Common Mode operation provides compatibility between unlike devices by defining least common denominator formats TTY amenable to both endpoints. A Class-Specific Mode provides for transmission of data which supports the standard features of defined terminal stations. Net 1000 also offers a Transparent Mode of operation which does not translate data. Interrupt and disconnect are the only format features supported in this mode. The two forms of data transmission service available are Call Service which provides for 2-way, session-oriented (inquiry/response) transmission; and a 1-way Message Service which provides 3 grades of transmission (priority, standard, and delayed) and 2 delivery options (delivery confirmation and earliest delivery time). Although it is probably not a major customer cost factor, traffic volume is a component in the continental U.S. pricing structure for the Call and Message Services.

Customers, using the facilities of a local common carrier, access Net 1000 through port interfaces at nationally distributed service points. Contingent on network design economics, a service point may, or may not, be a network node. The port interfaces are available in a variety of types and speeds; up to 9600 bps for private line analog and digital ports, and up to 4800 bps for dedicated and public dial ports.

Although the chosen equipment complement may change as the network evolves, current nodal hardware includes a DEC VAX-11/780 running on VMS to handle code processing, database assignments, and call routing. IBM Series/1 computers are utilized for protocol conversion and customer terminal interfacing.

As alluded to earlier, Net 1000 is much more than just a packet-switching system. It is a shared, expandable, user-programmable, and user-controlled data communication network. Besides providing communication interchange between previously incompatible terminals and host computers, Net 1000 can, for example, be used in distributed communication processing systems to perform the preprocessing and postprocessing tasks inherent in many business applications.

X.25 & Packet Network Concepts And Applications

Users may store and execute COBOL application programs and their attendant data files, a feature which suggests an imminent Net 1000 foray into the distributed processing, database inquiry, front-end processing, and network management fields.

Customer billing is distance insensitive and is based on usage. Net 1000 charges are based on traffic volume forwarded over the network, type of service, and customer-specified transmission priorities and deliver options in addition to charges for a service point port interface, computer resources used, program storage, and data storage facilities allocated.

■ NEW X.25 PRODUCTS

The acceptance of X.25 in the United States was slowed by the time required to develop support of the protocol by the major computer systems, a function of the development cycle and the caution exhibited by vendors when faced with a new protocol standard. Adding AT&T and IBM to the list of suppliers of X.25 products and services assures industry support, and many companies are responding by rushing products using X.25 to market.

Major computer vendors have largely embraced X.25, as shown on the chart in figure 2. A second round of enhancements is under way for these vendors to integrate X.25 fully into their communication network products, reducing the effort necessary to change to X.25 from other protocols.

Some vendors are also upgrading their support of optional facilities of X.25 to increase its application to specialized communication environments. **Datagram service**, a means of sending single-packet messages without the complex steps of setting up a **virtual circuit**, is useful where user interactions consist of a single, short entry. Datagram service is being implemented by some vendors. It is likely that many of the suppliers of large mainframes will eventually support all of the facilities described in the 1980 revision to X.25.

Suppliers of integrated circuits are working to provide devices to simplify the processing required to implement X.25 by moving some of the logic onto a chip. At least one company offers a chip with all logic necessary to implement level 2 of X.25. This trend will result in direct X.25 support by less complex computers and even terminals. It is unlikely that such devices will eliminate the need for a PAD function in the near future, but an inexpensive packet-mode terminal can be used economically in many new applications.

Software is also becoming available to provide X.25 support outside areas where the computer vendor has elected to implement it. Although the cost of such products are high, it is far less than the cost of custom implementation of X.25, and its use reduces development time a fraction of that required for a special X.25 programming effort. This can help users with older computers that will not be upgraded to provide X.25 support but will operate in an X.25 environment. Users considering this migration path to X.25 should make sure their hardware can support the protocol, in performance, memory size, and communication adapter characteristics.

■ X.25 & DISTRIBUTED PROCESSING

Many specialists feel that X.25 represents the best path toward a communication protocol to support true distributed processing. It represents no specific vendor and is supported by nearly all vendors. It can provide a multivendor user with a way to connect equipment for a distributed processing environment. Realization of distributed processing goals requires more than a method of communication. The scheduling of tasks in multiple systems, storage of distributed databases, gathering of information for inquiry and processing, and the management of file and record-level security are all outside the scope of X.25 or any likely extension. As a **peer protocol**, X.25 is a networking system primarily designed for connection of equals; it provides a better vehicle for networking computers cooperating to service a population of users than does a system defined with a fixed hierarchy of resources such as IBM SNA. SNA was designed to address both communication and distributed processing needs from its inception; SNA is evolving toward an architecture that will fully support connections of large host and terminal populations. SNA is one vendor's proprietary answer to the data communication/distributed processing problems faced by computer users, but it is not effective in the short term for users who must support equipment of several vendors.

X.25 is not an answer to the needs of all data communication users, but it can serve as the basis for a communication architecture that provides flexibility and economy now and also will support the highly integrated distributed networks of the future.

• **END**

The OSI Reference Model—A Blueprint For Network Communication

A Summary of the Basic Reference Model for Open Systems Interconnection (OSI) Developed by the Organization for International Standardization and its Comparison with Other Prominent Network Architectures

■ INTRODUCTION

Since its publication in 1978, the International Organization for Standardization (ISO) Basic Reference Model for Open System's Interconnection (OSI) has become the basis for the design and evaluation of modern communication protocols. Applications for the model range from satellite communication systems serving international corporations to local-area networks used to link computers and terminals within a single building.

The OSI model does not define HOW a protocol should be implemented, but its definitions of WHAT a protocol should do have served both users of data communication products and designers of new products, providing a sound basis for exploiting the advances in networks and transmission facilities.

■ THE CHALLENGE OF NETWORKS

Data communication applications developed from the initial business use of computers and communication and were characterized by designs to deliver data through connections normally used to carry voice—the telephone system. This kind of system provides the two parties with a "point-to-point" connection. The communicating devices operate at opposite ends of a circuit to allow each virtually instantaneous access to the other without interference from intermediate parties. Later, it became economically advantageous to connect several "slave" stations on a single circuit, under control of a single "master" station. This arrangement is referred to as multipoint communication. Although more than two users are served by such lines, they support only one dialog between users at a time. The master station either "polls" a slave to send a message or "selects" it to receive a message. Slaves do not transmit except by invitation, and never at the same time as another slave. Figure 1 shows both point-to-point and multipoint lines.

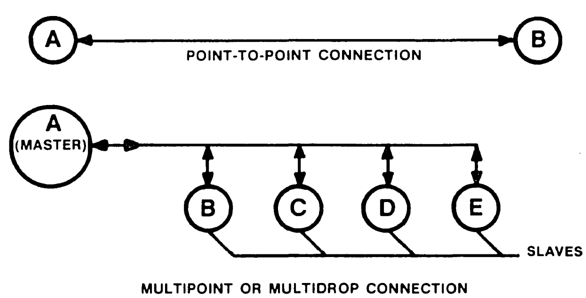


Figure 1 • direct connection arrangements.

Packet-switching communication concepts were developed in the late 1960s to provide a more economical and fault-free communication system. In a packet-switched network, many intelligent devices called "nodes" are interconnected via "trunk lines" to provide a series of possible paths between two points. User terminals or computers connect to a node and request the network to make connections to other users. In this type of network, the actual users are not directly connected to a circuit, and the path between them is vulnerable to delay and interference.

Figure 2 shows a typical packet network configuration. Two users must first connect to a node of the packet switch and request the service of connection to exchange data. The node cooperates with other nodes in the network to establish a path to the requested destination, and the node serving the destination connects to provide notice of an "incoming call." The process is similar to that of dialing a telephone. The dialing process requests the telephone company to provide a circuit to the called party. The ring at the other end is notice by the phone system that a call is waiting. Answering the phone, or accepting the packet, "incoming call," completes the connection phase for a conversation, either in the phone system or in a packet switch. The similarities between the two systems at this level mask the beginnings of a very complex issue: service requests versus data exchange.

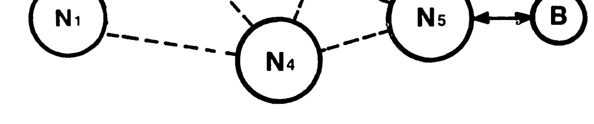


Figure 2 • network connections between users.

In the telephone example, the service request established by dialing is logically separate from the conversation; the mechanical task of dialing the phone does not relate to the vocalizations that make up the main part of the call. But, if the information exchange and service requests both use the same medium, a set of rules is needed to separate the two functions and avoid confusion and erroneous connections. Packet data networks communicate with the devices they serve in the only media available—data.

Another difference in the network versus circuit communication path can be seen by comparing Figures 1 and 2. A circuit system provides no servers between sender and receiver so all data the sender transmits can be acknowledged by the receiver. The acknowledgement procedure for correctly received data is included in the set of procedures the two parties agree upon to govern their exchange; it is called the "communication protocol." Data in such a system cannot reside in intermediate storage

The OSI Reference Model—A Blueprint For Network Communication

points; it is either received or it is known to be lost and must be retransmitted. A network, on the other hand, has active nodal elements to move information between users. A user is connected to a node and not to other users, so the acknowledgement of correct receipt of data may mean only that the node has received the data and not the final destination. Data can be entered in the network, and lost if a network element fails. An error recovery procedure between users can detect and correct errors on all the circuits in the path that connects them, but the delay inherent in node handling can become excessive, and the correct data must be transmitted over the entire path between users. If user A in Figure 2 has such an "end-to-end" error correction procedure with user B, an error in sending a message between A and Node 1 will not be recognized until the message is delivered to B, wasting the resources of the network in sending a bad message over several additional paths. User B must then send a request for another copy of the message all the way back to user A, and the correct copy must make the return. This waste is contradictory to the economic goals of packet networks. Errors that occur between user and network and between network elements must be corrected immediately; this creates another element in the "service protocol" the user must implement to connect to the network and exchange information.

If the network of Figure 2 has a set of rules for connecting to the network and the users have a similar set of rules to govern their exchange of information end-to-end, we have a workable basis for data communication through the network. Is it an optimum solution? Figure 3 shows the same user A and B pair communicating as before, but adds user C who wants to exchange data with user A in the same way that user B had exchanged data. User C, however, connects through a different network. If this second network does not follow the same rules as the first, the communication program that allows user A to converse with user B will not support conversations between user A and C. User A will need a new program; the amount of work involved depends on how well the programmers separate the procedures for talking to the network from the procedures for talking to other users. What must be done is to provide a program "layer" for communicating with users and "under" it another layer to handle the communication with the network. This lower layer must be changed for communication with a network using a different service protocol.

■ LAYERED PROTOCOLS—A GENERAL SOLUTION

The concept of a "user layer" and a "network layer" is an easy visualization for most communication users. The lower "network layer" provides a transparent path for the user, so that changes to the network protocol due either to changes in the network's operation or to migration to a new network can be limited to this layer. This two-layer model is the simplest model of a layered network architecture. The top layer specifies user tasks and programs, and the bottom layer specifies connection to the network and the network itself.

Reflection on the example of user A, who must communicate with a second partner, points out other opportunities for layering. Many service functions required in the new network, such as requesting a connection, are also found in the old, but the procedure for using them is different. If the user program could include an "intermediate layer" to request these generic services, the program could be simplified. The basic logic to request a connection would reside in the intermediate layer, and the network-dependent logic in the lower layer. Within the network itself, the nodes can also be considered "users" of the point-to-point links that connect them, adding a "link layer" to the "network," an "intermediate layer," and a "user layer."

This process of layering can be continued, eventually reaching a point where the boundaries between layers are difficult to justify. Many factors influence the selection of functions for each layer and the optimum number of layers in the network architecture. Many layers provide modularity of function and transportability, while few layers simplify the task of "connecting" the functions that satisfy a user request for service. In the process of defining layers, the functions directly related to the user tend to "migrate" to the top of the structure where the user is presumed to reside; those associated with the physical aspects of network connection drop into the network layers.

In 1974, IBM announced its System Network Architecture (SNA), a layered protocol that defines sets of layers. It provides basic network services, operates in user computers to help users access these services, and simplifies the user connection to the network. Figure 4 shows a model of the SNA concept. Although it was a revolutionary development in the eyes of most computer users, many other equipment vendors recognized the value of the concept and introduced their own proprietary communication network architecture. SNA, as Figure 4 shows, when introduced had 6 functional layers. Competitive offerings had different structures; Digital Equipment Corporation's DNA had 3 layers and Sperry's DCA had 5. Clearly, the question of where to divide functions was being decided on a different basis, and users trying to evaluate the offerings had difficulty determining which functions were implemented and where they were implemented. SNA, DNA, and DCA have now changed so that all have roughly equivalent layers—7 major ones in all—comparable to the OSI model.

A more serious problem than the one of evaluation faced both users who wished to implement mixed-vendor networks and network operators who wished to supply the lower-layer functions to users as an alternate communication service to telephone connection. The various network architectures did not even agree on a distribution of function, and there was little chance they would implement standard protocols or even protocols that could be translated one to the other. There seemed no chance that any vendor standard could be accepted across the industry because it would give the selected vendor both technical and marketing advantages. A standard was required and it had to be supported by an organization with sufficient influence to shape the

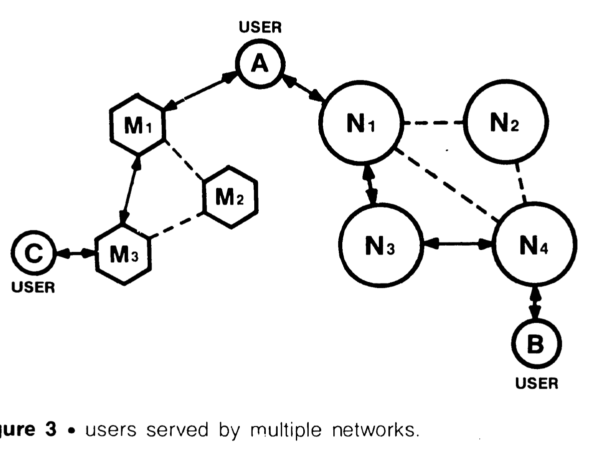


Figure 3 • users served by multiple networks.

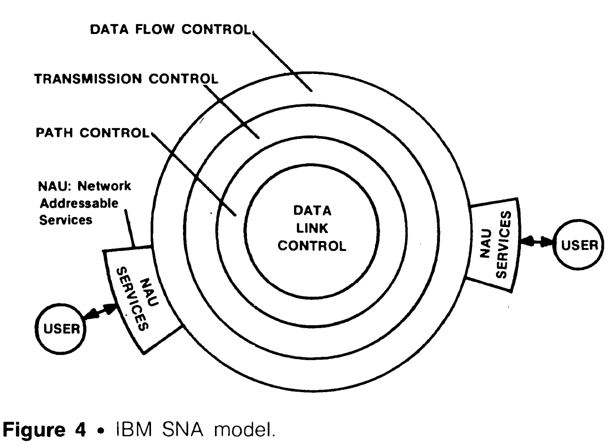


Figure 4 • IBM SNA model.

The OSI Reference Model—A Blueprint For Network Communication

development of network architectures and protocols to the ultimate benefit of users.

■ THE OSI REFERENCE MODEL

In 1975, the International Organization for Standardization (ISO) began work to develop the first step in standardization of network architectures: a model that defined the number of layers and the functions to be assigned to each layer. The goal was to provide a framework for specific standards of implementation for each layer in subsequent considerations, beginning with the most primitive functions and building toward those more related to the user. This "bottom up" approach reflected a desire to ensure that standards for network services were defined in time to influence the direction of growth in public packet transmission services. The result of the effort was published in 1978 as the Basic Reference Model for Open Systems Interconnection, commonly known as the OSI Reference Model.

The OSI model defines seven layers or levels consecutively numbered from the physical interface connection layer (1) to the application layer (7) representing the actual user. The lower three levels collectively define a basic network mechanism similar to the inner grouping of SNA layers. Levels 4 and 5 provide a reliable user path across the network, and level 6 a means of attaching various users to the common service facilities. Figure 5A shows the structure of the OSI model.

The first element a user normally addresses in connecting to a network is the physical, or electrical interface. The OSI model level 1 is the "physical layer"; it defines the mechanical and electrical interface between user and network. This includes not only the connector type and voltage levels, but the signalling procedures on the various interface lines for proper control of the connection. The OSI model requires that the physical level be satisfactory for the transmission of "transparent data"; it can use no procedure that restricts the characters users can send across the connection.

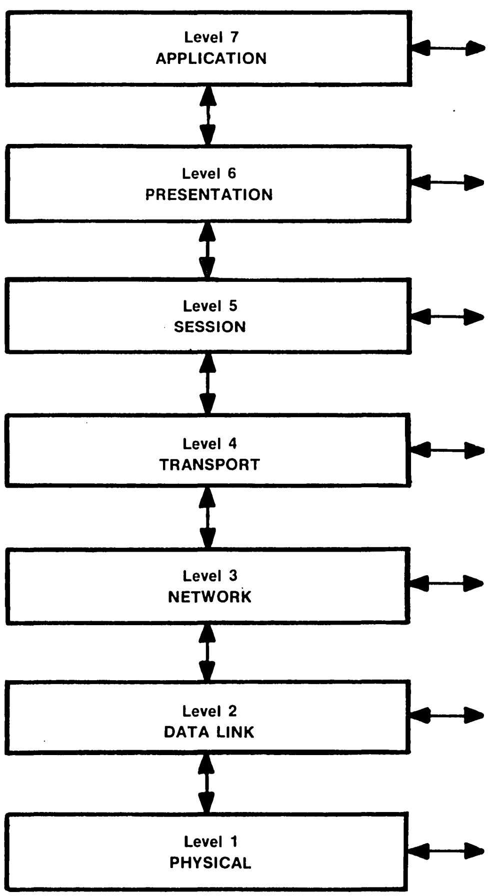


Figure 5A • the OSI model.

Level 2 is the "data link layer"; it controls the actual point-to-point exchange between two nodes of the network or between the network and the user. Level 2 provides an error-free channel that higher levels can use to exchange information.

Level 3 is defined as the "network layer"; it routes data between nodes in the network to establish a connection between users. It also identifies service levels, particularly in terms of connection throughput.

The "transport layer," level 4 of the model, is the first layer with a clear end-to-end function. The "transport path" appears to the user as a point-to-point error-free connection. The transport layer also segments or blocks messages that are not economical for network transmission, and allows users with the same end points to share a transport path. An example of this is 2 application pairs residing in an end-point computer. Various levels of transport service are proposed to account for differences in user sensitivity to data loss and reliability of the lower level network.

Level 5 is the "session layer"; it establishes a logical connection for the user to exchange information with another user. The session can share a transport path with other sessions.

The "presentation layer" is level 6 of the model; it provides each user with an exchange dialog with the network (and therefore to other users); it is compatible with local operating and data format requirements. Level 6 also provides a means to request connections through the session layer and to terminate connections when communication is complete.

Level 7 of the model is the "application layer"; typically identified as the "user," it is more properly the user interface to the host computer's communication access method. The application layer provides the user with a logical means to address partners, validate system security, and coordinate lower level services.

It is important to remember that the definitions of the model are guidelines for developing implementation standards in each area; the guidelines will be refined in each application. For example, the original definitions for the network layer, level 3, were more comprehensive than the popular level 3 X.25 implementation. X.25 now follows the OSI model for connection mode networks. The OSI model is being changed to accommodate connectionless mode networks, such as LANs.

The development of the OSI model took over 3 years, a long time for the simple structure that resulted. The primary work of the organization, however, was not the drawing of the model; it was the determination of the number of layers and the way functions were to be divided, the same issues on which vendors disagree. Each layer in a layered architecture has three "exchanges" of information as shown in Figure 5B. The first exchange is with the next higher level in the model used to service the requests from the higher level. The next exchange is with the next lower layer, used to request service from and through it to the other lower level layers. For example, if we examine level 5, the session layer, we see that it exchanges information with level 6, the presentation layer. The presentation layer will use this exchange to request service and deliver information. In a similar manner, level 5 interfaces with level 4, the transport layer. For each layer in the model, the higher level exchange represents its "user" and the lower level its "server."

The third exchange in a layered protocol is not so easily defined. Returning to the example of the 2 users communicating through a network, both users had similar layered structures. In error detection, a part of one user's structure is interfaced with a part of the network's structure to detect and correct an error. This interface between layers of the same number at opposite ends of the connection is called a "peer protocol exchange." Peer protocols are the most important concept of the OSI model. If every layer treats the layers higher as a user, then every layer must cooperate with its equivalent layer at the opposite side of the connection to deliver data from its own user to the other user. Stated another way, each layer in the model is a user in itself,

The OSI Reference Model—A Blueprint For Network Communication

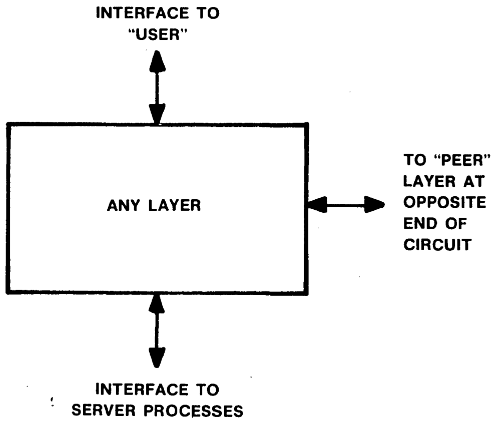


Figure 5B • OSI layer interaction.

exchanging data with another user via the combination of services provided by the lower layers.

Figure 6 shows the model as it applies to a user-to-user connection through a network. Note that each user has a 7-layer structure while the network has a 3-layer structure. The difference in layers between user and network exists because the boundary between levels 3 and 4 marks the point where services stop being user-to-node or node-to-node and become user-to-user. The peer protocols of the lowest three layers are point-to-point, while those of the higher levels are end-to-end. A point-to-point protocol is used between 2 communication network users or network nodes, connected directly by a communication link. An end-to-end protocol is used between a data source and a data receiver connected through multiple network nodes without being used or changed along the way. Referring back to the network model in Figure 2, point-to-point protocols are used between a user and its connected node, or between nodes. If an error occurs in such an exchange, it is corrected by one of the two connected elements of the network without the knowledge of the rest. On the other hand, if 1 element fails, the point-to-point protocol fails as well because only 1 station now operates on the line. Errors cannot be corrected. In this case, an end-to-end protocol between levels operating at each user site causes the network to choose another path and repeat transmission of any lost data so the connection can continue.

Discussion of multiple protocols can cause visions of a network with a maze of connections that individually transfer elements of each peer protocol from one place to another. Fortunately this inefficient concept is not accurate. The user's data element or message is the basic unit of information sent over the network. It originates in level 7 and is passed down through the levels at the sender's location. At each level, it has a control element called a "header" added to it to store parameters necessary to handle the

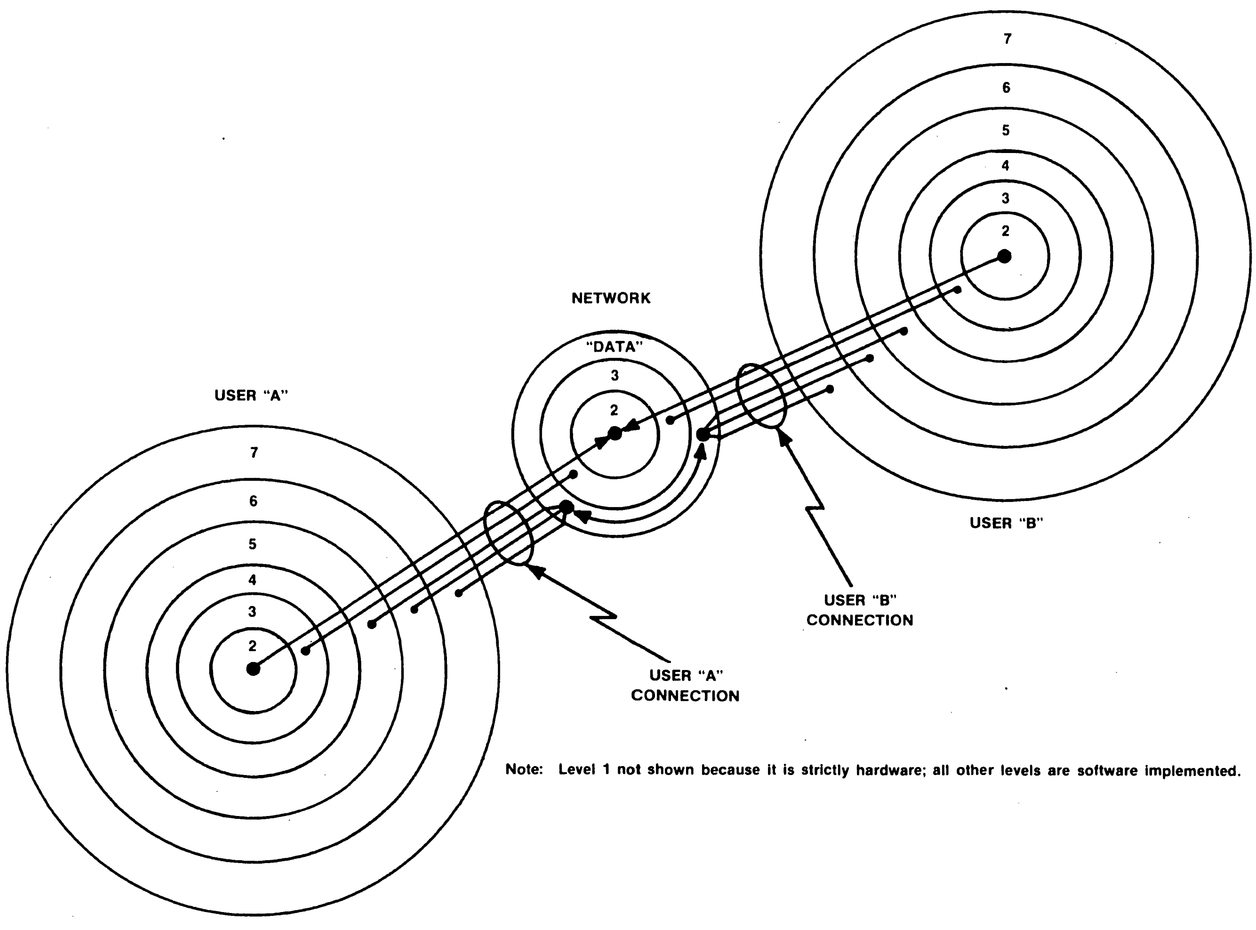


Figure 6 • an OSI-model connection.

The OSI Reference Model—A Blueprint For Network Communication

information at that layer at the opposite end of the peer protocol. When the data with its collection of headers is received at the destination, it migrates upward to the user level, with each intermediate level removing its header as it passes the data through the layer. The headers allow the peer protocols to pass information without forming a separate message, although not all peer requirements can be met in this way. When a level generates a message without a corresponding request from its user layer and having nothing to append it to, it sends its header alone to the next lower level, which treats it as data as before. In this case, the message will arrive at the destination and migrate upward to eventually disappear before reaching the user layer. Figure 7 shows the header structure of a model protocol. Note that each layer treats all data and headers from higher levels as data.

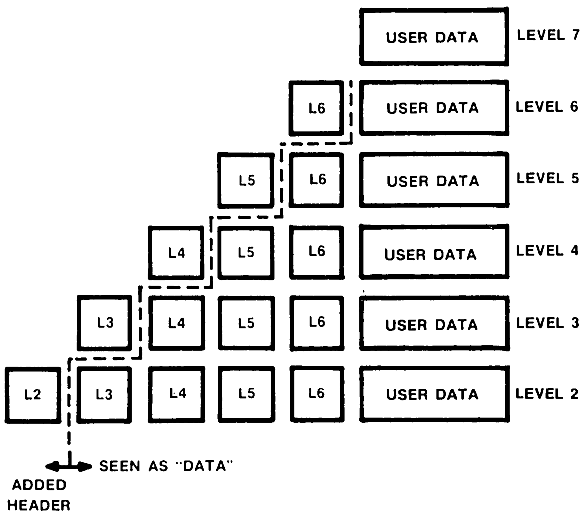


Figure 7 • header usage in an OSI-model protocol.

The individual functions of the layers of the reference model can best be explained in terms of popular computer communication protocols. The following sections will describe the function of each layer and explain its use through a brief description of how it is implemented in several sample network protocols.

■ THE PHYSICAL LAYER

The physical layer of the OSI model defines the electrical interface between the user device and the network, or between elements of the network. In most practical environments, this means control of the modem at the point of attachment or control of a direct user-to-network tie. The physical level definition includes a specification of the mechanical connection (the type of connector, number and location of pins, the use of each pin, and the identification of the male/female mating arrangement for each entity in the attachment) and voltage levels as well as the signalling procedure to allow both parties to recognize that the other station is active. For example, the domestic RS-232C standard calls for a terminal device to enable Data Terminal Ready (CD) and for a connection device to enable Data Set Ready (CC) when their respective interfaces are active. Further signalling procedures indicate that data is being sent or received.

The ideal physical interface standard must be unambiguous and capable of supporting the type (full- or half-duplex, for example) and speed of exchange users require. Some examples of interface types available are the common domestic EIA RS-232C standard, the very similar CCITT V.24 standard, the two new higher speed standards EIA RS-449 and CCITT V.35, and the CCITT digital standard X.21.

EIA RS-232C and CCITT V.24 are simple standards designed for use when low speeds and short distances are specified. Cable electrical characteristics limit these interfaces in theory to a distance of about 50 feet and in practice, with special cabling and interface drivers to about 200 feet, and the data rate to 19.2K bits per second (bps). The EIA RS-232C interface is by far the dominant standard in use today. Unfortunately, many vendors do not follow the signalling procedures exactly, and users often must use special cables to support communication. The speed limitations of the EIA RS-232C interface prevent its use in high-speed environments, and many users find the cable length restrictions unreasonable, particularly for connecting a terminal to a computer or to a communication processor where physical distance between devices is limited.

For higher speeds, CCITT recommendation V.35 provides an interface capable of supporting data rates to 72K bps. V.35 is commonly used for connections to wideband analog (voice-type) lines, and to AT&T's Dataphone Digital Service (DDS).

The EIA RS-449 is an alternate form of high-speed interface that supports data rates up to 10 million bps. This interface stirred much controversy when introduced because it uses a dual-cable mechanical attachment for full capability. The EIA RS-449 interface can support cable lengths up to 4,000 feet and at high data rates, although the maximum data rate depends on the cable length. EIA RS-449 has two companion standards, EIA RS-422 and EIA RS-423. These define two methods of applying a signal to critical interface leads such as transmit data or receive data. The EIA RS-423 method, called an "unbalanced" interface driver, uses a common ground return for half of the signal path, in the same manner as EIA RS-232. In the balanced form, EIA RS-422, the critical signals are sent using a pair of leads operated in a kind of "push-pull" fashion. The balanced interface form supports data rates to 10 million bps.

The CCITT X.21, a more recently introduced digital interface, is specified by the CCITT as the preferred level 1 for X.25; it has already been supported by IBM for SNA. X.21 differs from the other standards in that it supports a digital rather than analog connection through modems. It specifies a 15-pin connection with two data leads and 4 control leads. It can inform a user of the exact state of a digital connection and is superior to other interfaces for a digital environment, but its operation is limited in an analog telephone circuit environment. It is unique, however, in that computer signalling functions such as dialing are performed by sending data in its normal ASCII form, and response conditions such as "busy" or "number changed" are likewise represented as data so that the computer can directly read and interpret them. In many ways, X.21 acts as a combination data and automatic call unit interface.

IBM SNA supports all of these interfaces. X.25, while preferring the X.21 interface, supports the older CCITT V.24/RS-232C and other interfaces if 2 parties mutually agree to use them. Most users will find that the details of physical interfaces are so buried in the operation of their telecommunication software that the choice is transparent so long as the equipment being connected supports the same standard.

The OSI model will accommodate all the standard interfaces developed by other standards groups such as the Consultative Committee for International Telegraph and Telephone (CCITT) of the International Telecommunications Union and the Electronics Industry Association (EIA).

■ THE DATA LINK LAYER

Most users think of data link protocols when they think of data communication. A data link protocol is the set of rules or procedures 2 users of a communication link operate under to assure proper communication. Early data link protocols such as IBM's Binary Synchronous Communication (BSC) procedures had elements of higher levels of the OSI model, but these protocols were severely limited and not intended for use in a complex network environment.

IBM developed a new protocol for its System Network Architecture, called Synchronous Data Link Control (SDLC). SDLC was submitted to both the ISO and the American National Standards Institute (ANSI) for approval. Both bodies modified SDLC somewhat and issued protocol standards that are slightly different, High Level Data Link Control (OSI's HDLC) and Advanced Data Communication Control Procedure (ANSI

The OSI Reference Model—A Blueprint For Network Communication

ADCCP). All of these protocols predated the OSI reference model, but all embody the elements that the model defines as level 2 requirements.

The OSI model requires that level 2, the data link control layer, provide procedures to set up a link or agreement that both parties are ready to exchange data, transfer information, detect and, if possible, correct errors, and terminate the connection when it is no longer needed. Key elements in the design of level 2 protocols are transparency, efficiency, and control.

• **Transparency** means the user is not restricted from sending any character because it has significance to the protocol itself.

• **Efficiency** demands the protocol to minimize the effects of throughput delays within the physical path and to minimize the link capacity loss associated with control exchanges and other elements not actually connected with user data.

• **Control** requires the protocol to prevent random communication errors and ambiguous states where neither party is clear on the procedures to follow or where the action of one party can be misinterpreted by the other.

The transparency requirement is best understood in relation to BSC protocol. In this protocol, special control characters are defined for marking the Start-of-Text, End-of-Text, and so on. These characters had special significance to the communicating hardware/software and could only be transmitted in a special class of procedures called "transparent text." This was an unwieldy solution to the transparency problem, and a better solution was sought. It was found in a concept called "bit-stuffing." When a byte of user data is to be sent, bit-stuffing requires that any sequence of five "1" bits be followed with a "0" bit inserted behind it. At the receiver, the "0" bit that follows a string of five "1" bits is deleted, so there is no effect on user data. Characters that have 6 or more consecutive "1" bits are thus available for special link control functions and cannot duplicate user data. One such character, a binary pattern 01111110, is called a "flag" and is used in these "bit protocols" to mark the start and end of a data message or "frame."

Bit-stuffing serves a purpose beyond transparency. When modems are used for communication over phone circuits, problems in maintaining the synchronization of the sender and receiver can occur when the transmitted data stream contains a long string of "1" bits or "0" bits. This condition is called a "transitionless line" and confuses the receiving modem because there are no changes in line state to help detect a "drift" in receiver timing with respect to the received characters. Bit-stuffing prevents a string of more than five "1" bits, thus it eliminates half of the transitionless line problem. The other half is commonly addressed by changing the way data is sent on the line.

In normal data coding, a "1" bit is sent as a "1" and a "0" bit as a "0." Another form, called "Non-Return to Zero Inverted" or NRZI, sends by changing transmitted line state (0/1 or 1/0) when a "0" bit is found in the user data stream. Thus, the binary value 11010111 is sent as 11001111, assuming the line starts in the "1" state. The value "1" is sent until the third bit when a "0" causes it to invert; the "1" in the fourth position allows the line to retain its state; the "0" in the fifth position inverts it again; and the "1"s in the sixth, seventh, and eighth positions do not change state. When NRZI data coding is added to bit-stuffing it makes a transitionless line impossible; a string of "0" bits inverts the state with each bit transmitted, resulting in a line pattern 10101010. NRZI coding is therefore the preferred way to send data with bit protocols.

Not all modern protocols use bit-stuffing. Digital Equipment Corporation's DDCMP provides a count of user data characters as part of the message. This eliminates the need for a special character or bit pattern to mark the start and end of a message. This type of protocol is sometimes called a "byte count" protocol. Byte count protocols operate on many devices intended for use with older synchronous protocols such as IBM BSC, while the bit-stuffing protocols require special hardware for the "0" bit insertion and deletion.

Control for the bit protocols requires more control characters than are available with a sequence of 6 or more "1" bits. Therefore, control "header" concept was introduced. If each user data message has a precise number of characters preceding it for link control, extensive control exchanges can be supported without requiring the user to give up control characters. Link overhead is slightly increased, but if user messages contain 100 characters or more, the loss is not significant. A message header allows each station to identify itself, indicate whether it is making a request or providing a response, and indicate its send or receive status to simplify station synchronization.

Error detection and correction in the newer protocols had special issues to resolve. One of the key issues was the effect of message acknowledgement on throughput delay. A protocol that accommodated several unacknowledged and outstanding messages would substantially reduce throughput delay. The older "ACK" protocols required the receiver to send an "ACK" for each message received correctly; transmission throughput was considerably increased when the path delay between stations increased, as a result of the delay in the sender's receiving the ACK. Another key issue was the error potential of high-speed communication which requires the error detection facility to provide the user with as close to an error-free link path as practicable. This requirement was met using a mathematically derived 16-bit polynomial called a Cyclic Redundancy Check (CRC) value based on the actual data sent. CRC assures an error-free link and is calculated by the sender and appended to the data before the final "flag."

The first issue, message acknowledgement, was resolved by a "sliding window protocol," which allows several unacknowledged messages to be outstanding. In sliding window protocols, the sending station places a sequence number in each transmitted frame, the number cycling automatically to zero when it reaches a value called the "modulo" of the system. Modulo-8, typically used, sequences numbers from "0" to "7," then returns to "0." When the receiver acknowledges data, it sends a message containing the sequence number it expects to receive next to the sender. The sender then considers all frames with numbers less than that value as acknowledged. If the receiver detects an error, it sends a rejection message with the sequence number of the first frame it wishes to have retransmitted, and the sender resumes sending with that frame. The sender can continue to send as long as one number remains between the number of the frame being sent and the number in the oldest non-acknowledged frame, a total number of frames equal to the modulo minus one. This relatively large number of potential unacknowledged frames prevents delays in the communication path from affecting the data rate. The sender must receive an acknowledgement from the receiver only once in seven frames, for example, so a sender-receiver-sender delay that is less than the time required to send seven frames will not affect the data exchange.

Using sequence numbers to control errors allows another step in increasing protocol efficiency—"piggybacking": 2 stations are sending data back and forth and the acknowledgement sequence numbers a receiving station must return can be "piggybacked" on data frames the station is sending. Unique messages to acknowledge data are thus reduced or even eliminated.

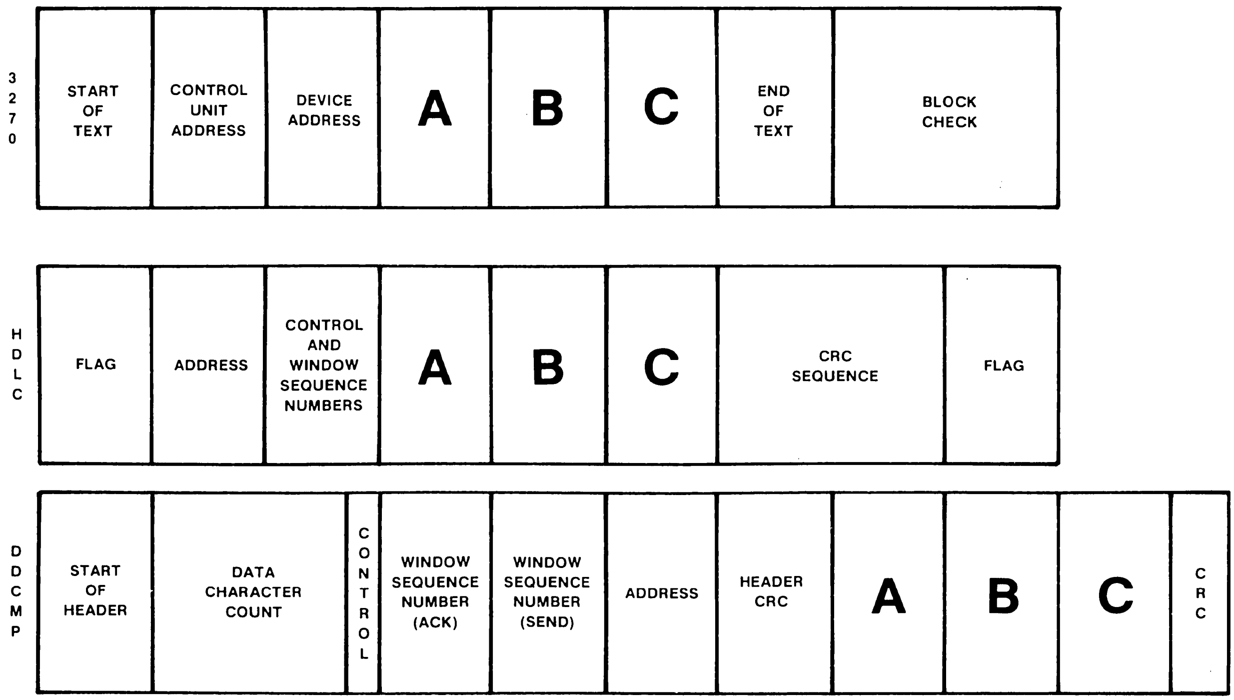
SNA and X.25 both utilize bit-oriented link protocols and window acknowledgement procedures. The SNA SDLC system is actually a subclass of the more general OSI and ANSI protocols that are primarily designed for networks organized in a hierarchical "master/slave" arrangement, referred to as a "tree" network. The X.25 system is another subset of HDLC called LAP (Link Access Procedure) or LAP-B in its "balanced" version. Both these procedures are more efficient when the stations communicating have a "peer" relationship rather than a master/slave one, and LAP-B is the preferred form because it has the simplest of all structures when applied to such relationships of equal stations. Users of SNA cannot select a form of link protocol, because SNA specifies only one form of SDLC at this time. However, X.25 users should utilize the LAP-B form of HDLC because the current CCITT standards specify that LAP-B will be "the only one available in all networks."

Figure 8 shows a data string of the characters "ABC" as it appears in SDLC/HDLC, DDCMP, and IBM 3270 BSC.

■ THE NETWORK LAYER

Open Systems Interconnection (OSI) specifications call for the level 3 network layer to provide a path for higher-level data free

The OSI Reference Model—A Blueprint For Network Communication



SDLC/HDLC FIGURE SHOWS AN I-FRAME, NOT AN SNA OR X.25 "PACKET"

Figure 8 • illustrative level 2 formats.

from any network dependencies on routing, connection request procedures, and so on. This normally provides a unique "call" identification so that users can share a level 2 link without interference and means to request and clear user "calls," to notify a user of incoming calls, and to control flow of data across a point-to-point connection so that intermediate network elements are not "swamped" with data if a high-speed user sends data to a slow user.

All implementations of level 3 define a form of logical path between a data sender and a data receiver. The characteristics of this path and the way it is maintained differentiate the various architectures at level 3. One path form is called a "virtual circuit" because it simulates a physical connection between users. Data introduced by the sender emerges from the other end at the receiver in the same sequence; the "pipe" between is established and maintained by the network and users need identify only the "pipe" or circuit and need not continually provide a long formal network address of the other party. The second popular path is not a fixed path at all, but a "promise" of eventual delivery. Called a "Datagram," it provides for routing data messages individually using an integral address. Without a fixed path, there is no fixed order of arrival and the receiver or an agent in the receiver system must assemble the data in proper order.

X.25 provides an implementation of level 3 that conforms to the OSI model, defining this layer as the "packet" layer. Each link, controlled by level 2, is divided into a series of "logical channels" numbered from 1 to 4095. Any unused channel can carry the request for a connection through a special control packet called a "call request"; it contains the "addresses" of both the called and calling parties. Once a call is set up, the data associated with it carries the "logical channel number" of the call as a part of its level 3 header, separating it from other data on the link, allowing it to be individually routed to its final destination user. Either the called or calling parties can break the connection with a "clear" packet. If a station receives a call, the network assigns an unused channel and presents the user the call request packet with that channel's number assigned. The user can then "call accept" or "clear" the call as desired, corresponding to picking up the phone or hanging up.

Certain types of data exchanges can be supported without call setup. Where two users have a continual need to exchange data, a "Permanent Virtual Circuit" (PVC) can be established as a kind of dedicated route between the users. Although the route is permanent, the resources of the lines and nodes are dedicated to the users' conversation only when it is taking place. The datagram, as discussed earlier, is useful for very short messages. It is a data packet with the source and destination user addresses included. This packet is routed as though it were a call request using one of several logical channels set aside for that purpose. Because all of the data is included in the datagram, there is no need to set up a connection. The datagram is delivered to its destination based on the address within it, and no clearing of a path after delivery is required. While PVCs are considered to be a required feature of X.25 networks and most host computer support packages, datagram support is rare. This fact should be considered when application designs that could utilize datagrams are reviewed.

Flow control in X.25 is done with a sliding window mechanism just like the one used in level 2, except that the ability to retransmit packets in error is not normally implemented. A station is assigned a "window size," the maximum number of packets which can be sent without an acknowledgement, based on a series of network standards or by "negotiation" between the two users. No station can send more than this number of packets without a "flow control acknowledgement" from the other station or the network node to which the user is connected, thus large amounts of data cannot accumulate in the network when the ability of a user to send is not matched by the destination's or network's ability to accept or deliver data.

IBM's SNA calls level 3 "Path Control," and its function is similar to that defined in the OSI model. Implementation, however, is very different. X.25 defines only the connection between

The OSI Reference Model—A Blueprint For Network Communication

networks and users, and by extension between two network elements. Issues of how connections are requested are covered in X.25, but not the way in which the network actually makes them. SNA, on the other hand, specifies the entire operation of the Network layer, including routing and connection mechanisms. Its users are therefore provided a framework within which to build routing tables and to evaluate data movement while X.25 users must gain access to such information through the suppliers of the network and accept multiple methodologies and possibly little opportunity for control.

Because SNA is a hierarchical network, the relationship between network elements is largely defined by the structure of the network. Each element in the network is "activated" by a central control element called the "System Services Control Point" or SSCP. The activated elements have a series of paths connecting them called "virtual routes" defined in routing tables contained in network communication controllers such as 3725s. There is no connection setup within SNA at level 3 in the way supported in X.25; the routes and therefore the connections exist at system activation time. SNA messages will carry addresses in each message rather than use a logical channel number, but the routes are virtual circuit paths by the previous definition; they are established by table and will not permit data to become unsequenced.

SNA formats for level 3 data vary depending on the type of SNA devices involved. Simple terminal-to-controller interactions require no real level 3 addressing; the connection uniquely identifies the device. When this same message is sent from the cluster controller to the communication controller, a more complete address is required and is provided by the cluster controller. This hierarchy makes the design of terminals easier because complex address structures for these devices are unnecessary in most cases. The type of "transmission header," and therefore address, is defined by a "format ID" (FID) code. FID-2 is used for most interactions, while FID-3 is used for interaction between a terminal and its cluster controller.

Flow control in SNA is achieved through a technique called "pacing." The pacing technique uses "tokens" to control message flow. A receiver sends transmitting stations a "token" which is "good for" some number of messages. The transmitter can then send until the token is "spent"; the receiver can send additional tokens to maintain the flow of data.

Digital Equipment's original DECnet was a pure datagram network. Each message has a destination address for delivery. A DDCMP packet can loop in the network for a while, be duplicated and delivered twice, or discarded. It is the responsibility of a higher layer of DECnet to prevent out-of-sequence delivery and to recognize duplicates or missing packets. Digital now uses X.25 in its Phase IV DECnets; DDCMP is used in DECnets up to and including Phase III.

Routing and flow control parameters are generally under some user control, and improper selection of the values can degrade network performance. Routing decisions should be made with the expected volumes of traffic for each connection in mind to prevent congestion and associated delays in delivery. Flow control or pacing values should be chosen so that the throughput of the connection can be maintained without requiring extensive resources for storing data in the network. With X.25, flow control window size can be negotiated at call setup within boundaries set by both the user and the network. SNA will dynamically pace on its trunk paths, but will allow the user to specify a "pacing limit" on certain types of lower level paths.

■ THE TRANSPORT LAYER

One of the most significant layers of the OSI model is the level 4 transport layer. Unlike lower levels, level 4 is present only at the two end-points of a connection and thus represents the user's gateway to the communication facility. The purpose of level 4 is to provide a reliable "transport path" to support user communication. To do so, level 4 calls on level 3 to establish a network route and on level 2 to provide error-free exchanges on all links in the route.

Transport functionality is end-to-end; consequently, it is generally part of the communication software in a host computer and not a part of a network; thus, standardization of it has been more difficult. X.25 does not include an implementation standard for level 4. A draft standard for level 4 was approved by the OSI committee at its plenary session at Ottawa in October 1983.

A transport path can be considered the equivalent of the point-to-point connection between users which most applications assume. Thus, a transport path must possess the characteristics of a direct connection so far as the users are concerned. It must provide user-to-user error detection and correction, a path immune to failure resulting from network element failures. It must reblock user data presented for transmission for optimum utilization of the level 3 path, and control the flow of data between users, in the same manner as level 3 controls the flow of data between user and network.

SNA is the dominant implementation of level 4 functionality. Unfortunately, there are differences between "transmission control," the SNA title for level 4, and the OSI model. OSI separates the functions of level 4 from those of levels 3 and 5 to the same extent as the functions of level 3 are separated from its surrounding levels. SNA, however, groups the functions of levels 4, 5, and 6 into a single structure called a "half-session," for which a single header is established. Within SNA, a "transport path" does not exist independent of higher level connections, because all high-level functions are equalized into a single control structure—the request/response header. Transmission control provides flow control of information end-to-end in much the same way as the path control pacing concept applies flow control to an entire "path," but the end-to-end error control assigned to level 4 by the OSI model is applied at level 3 in SNA. The path failure immunity supplied by level 4 of the OSI model is a level 5 function in SNA because, as mentioned earlier, the transport path has no separate identity in SNA and cannot be reestablished separate from the higher levels.

The OSI standard for level 4 provides a much more structured approach. A "transport user" can select various types of service: normal data transmission or "expedited flow" of data. The non-connection oriented data, such as datagrams, is optional. Users can also select a basic or extended class of transport service to match the reliability requirements of the user and the characteristics of the network. Transport paths are established to support normal or expedited data; failure recovery of a level 3 circuit supporting a transport path can be achieved without loss of user data by establishing a new path. End-to-end flow control and error recovery are both supported within level 4. Like SNA, the standard provides for reblocking user data where message sizes are too large to be efficient as packets through the network. The transport level provides the user with an immunity from the failure of a network element so long as any path between users can be defined. This is achieved through a sliding window protocol similar to that used at levels 2 and 3. Without this immunity, user application programs receiving data from remote network users must include procedures to handle loss of contact during the application of these remote transactions—such as "backing out" of partial batches of changes and a new request for data from the remote.

DECnet provides a transport facility very similar in function to that defined by the OSI model, except that the additional function of datagram handling is introduced. Since the lower levels of DECnet operate as a pure datagram service, the transport level provides the facility of detecting and handling the arrival of a packet out of sequence, the arrival of more than one copy of a packet, and the loss of a packet within the network.

Out of sequence arrivals are handled by waiting a predefined time limit for the missing packet(s) and reporting a loss if the packets do not arrive within the limit. The opposite station must then retransmit the lost data. Duplicate arrivals are handled by discarding any packet whose number is not greater than the last "accepted" (meaning consecutively received and therefore deliverable) packet.

■ THE SESSION LAYER

The functions of level 5, the session layer, are probably the most difficult to relate of any in the OSI model, not to a small extent that no accepted implementation for session services exists. SNA deviates from the model in levels 3 and 4, and continues to do so

The OSI Reference Model—A Blueprint For Network Communication

in level 5. The OSI plenary session in Ottawa in October 1983 also adopted a draft standard for level 5 protocols.

If a level 4 transport path is assumed to exist between user systems, individual user applications can establish sessions, or user connections, through that transport path. The transport path is a "raw transmission facility" similar to that of a line. Session services add some user features to it, such as the ability to logically identify network users. A 14-digit international format address is a reasonable way to request a connection at the network level, but probably is not so at the level of a user. An extension of the logical address is the satisfaction of local procedures for accessing the user. If an application has a specific security access procedure, it may be undesirable for a user to carry it out directly, especially if that user has many such applications with which to interact. Session services can allow the user to select the destination logically and carry out the required procedures to access it via an available transport path. It also provides the user with the ability to send data via the normal or expedited transport paths.

SNA data flow control provides for the functions of level 5, but most of the functions are in other layers. The establishment of the session in SNA is a cooperative task between the SSCP and the NAU (Network Addressable Unit) services manager (level 6). Once a session is established, SNA level 5 (data flow control) does provide for some of the message-control aspects of the OSI model, but SNA does not provide for the use of expedited flow at a user level, and the access control aspects of OSI session services are handled in SNA at a higher level.

Most other networks do not provide a separate session layer.

■ THE PRESENTATION LAYER

Although SNA calls level 6 "function management," the OSI model and SNA converge again here. The purpose of level 6 is to provide a means for parties to communicate in spite of differences in the format of the data, a concept sometimes called a "virtual terminal" interface. The concept of "virtual terminals" dates back to the early experimental work done for the Department of Defense with its network for the Advanced Research Projects Administration, ARPANET.

In a virtual terminal interface, all users of the network exchange data in a single logical form that identifies the manner in which the data should be "displayed" rather than providing specific display format commands for any particular terminal. For example, a virtual terminal interface might specify that a field was to appear in the middle of the first line of a display, leaving it to the particular level 6 process for a user to issue the commands to accomplish that on the user's terminal.

SNA provides two widely used standards for end-user data streams which provide a form of virtual terminal interface. One is the popular 3270 format of data, and the other the SNA character string.

3270 Data Stream Compatibility (DSC) provides the same cursor control, display control, and modification control of data to SNA users as provided by the 3270 BSC terminals. This supports upward compatibility for 3270 BSC application programs, which can issue the same terminal commands to position a prompt on the screen as were used in older BSC applications. These same streams can be sent to application programs for processing if the programs will interpret the 3270 commands and field attributes properly. Since many minicomputers and microcomputers utilize 3270 emulation as a means of connecting to IBM or other mainframes, this DSC string capability will allow them to operate if they have the ability to receive SNA at lower levels.

SNA Character String Compatibility (SCS) is a device-independent data string in that no actual IBM hardware utilizes it. SCS codes are used to build what IBM calls a "presentation surface," or a dummy internal display form, which is filled in by data sent in the SCS format. When the surface is filled, it can be "mapped" to a particular device. The same surface could, for example, be mapped to a CRT display or to a printed form so long as both media had sufficient display space for the data. A minimum form of the SCS called the Basic Data Stream Profile includes only a form feed (or clear screen) and new-line function. All SNA destinations are expected to provide presentation support for this Basic DSP, providing the users with a universal, if trivial, display form.

Data stream support causes some difficulties with users of so-called "SNA-compatible" devices. Although a device that provides the services of SNA at the communication levels can technically support connection to an SNA network, lack of support for the particular type of data stream employed by the user may make the device useless.

There are no ISO or CCITT standards or draft standards for level 6, but the American National Standards Institute (ANSI) has a draft called X12 that provides prototype presentation services. A major corporation is using X12 as an experimental presentation model. The OSI presentation level draft proposed is expected midyear 1984.

■ THE APPLICATION LAYER

This layer provides the user interface to the network. The ISO committee is working on a number of standards for this layer: file transfer, access, and management (FTAM); virtual terminal, job transfer, and manipulation; systems management; and layer management standards. Other groups are working on standards for specific industries, such as banking. IBM has developed a Document Interchange Architecture (DCA) standard for its application layer of SNA for its office automation products. A similar standard may someday be adopted for the OSI model. These standards will generally be optional for the applications layer because they tend to be application oriented.

■ USING THE OSI MODEL

The primary value of the OSI model to a data communication user lies in its ability to serve as a benchmark of facilities against which protocols can be judged. An ISO committee is working on ways to specify the standard in a precise mathematical format. Another committee is working on ways to verify that a network conforms to the OSI model and at what level.

The developers of the model did not invent communication functions, but defined and separated them in a logical manner. A user can therefore expect that each function of the model must in some way be represented in each use of data communication, even if the "representation" is an implicit decision to ignore it. The popular X.25 and SNA protocols have been used as examples of features of the various levels, but it is also possible to evaluate them globally against the model. Figure 9 shows the model compared with both IBM SNA and DEC DNA.

At the OSI Plenary session in Ottawa, CCITT agreed to adopt the OSI model. The only difference in the ISO document and CCITT document defining the standard will be in the introduction.

X.25 defines three levels of the OSI model: the transport, session, and presentation services will be the same as the OSI model.

SNA evaluation against the model does not produce questions of missing functionality because SNA is complete through level 7, if somewhat differently organized. This difference in organization does present its own questions, however. For example, IBM support of X.25 on the IBM 3705 communication controller can be evaluated in terms of the model. Because X.25 has three levels and SNA has 7, one could question whether the "gateway" provided by X.25 support will have any functional restrictions. What will happen to the higher 4 levels of SNA? The answer is that they are still present and their control header will be carried into X.25 as data. This means that the X.25 output of the 3705 will still contain SNA control information which is expected to be received and processed by an SNA level at the opposite end of the X.25 connection. Because X.25 has no such layers, the other user will be presented with this header information and expected to respond to it as an SNA device would. The "gateway" is thus limited to service between two SNA devices fitted with X.25 adapters and not a generalized X.25 gateway at all. In subsequent implementations, IBM will fit the gateway with higher SNA levels to intercept this control information and leave the X.25 data stream free of SNA-dependent information, thus making it

The OSI Reference Model—A Blueprint For Network Communication

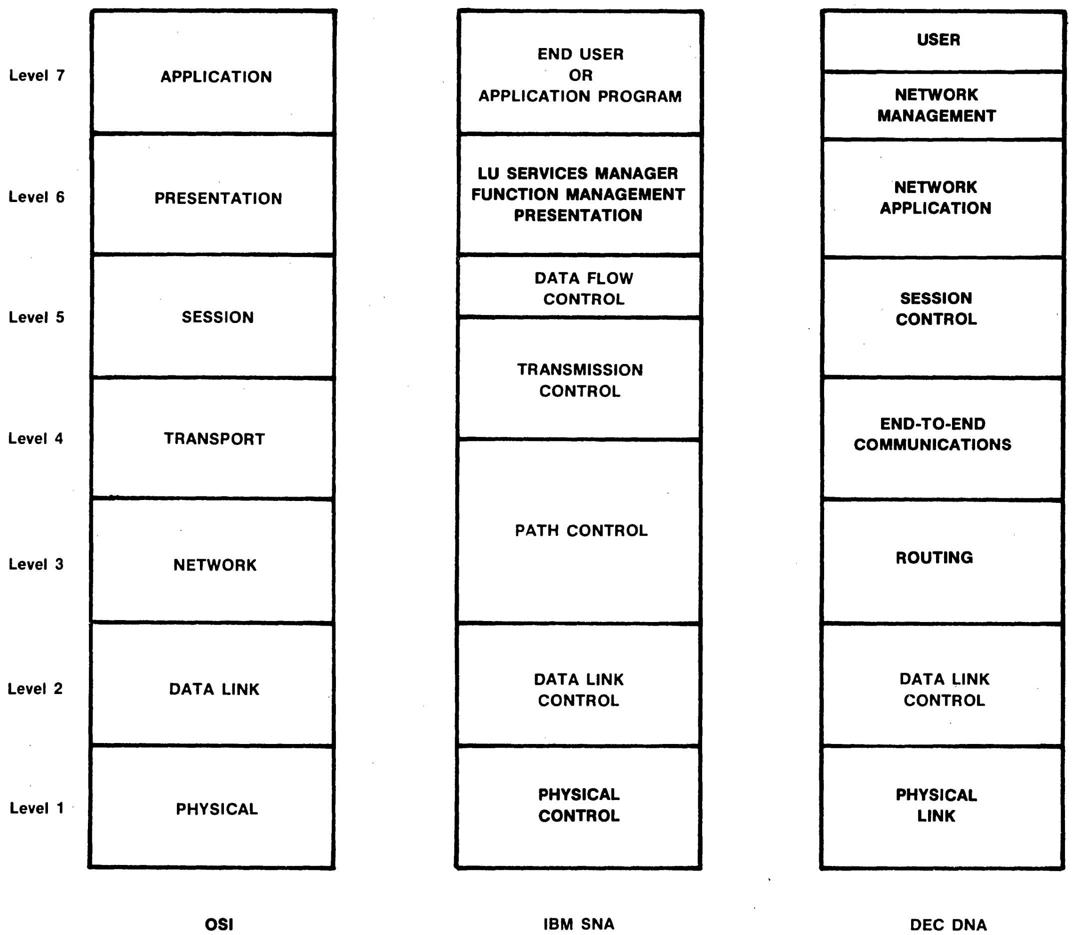


Figure 9 • level comparison.

suitable for connecting SNA devices to non-SNA devices via X.25. Figure 10 shows a model of the current and proposed gateways between SNA and X.25. Now that CCITT and ISO have agreed on the OSI model, future networks should all be compatible. The problem will be interconnecting old networks to the new networks. Vendors will need to provide gateways from one to the other. Most network vendors are committed to supporting the OSI model and to building OSI products directly.

■ THE FUTURE OF THE OSI MODEL

The OSI reference model has been accepted internationally as the basis for future developments in data communication protocols. The year 1983 was a banner one for standards development. Cooperation was the order of most standards meetings. CCITT and ISO agreed on the OSI model, using exactly the same wording for each organization's standard, except for the introduction. Voting by each organization's membership is considered routine and should be completed by midyear 1984. The two organizations also agreed on the underlying standards to implement the Session and Transport layers.

The draft proposal for the Presentation Layer is expected by July 1984. Work is proceeding on various services that will be offered at the Applications Layer, such as file transfer, virtual terminal, systems management, and layer management. The draft proposal for virtual terminal is scheduled for mid-1985. Subcommittees are working on standards to incorporate local area networks into the OSI model.

The OSI model is not a static vehicle. It is a flexible overall framework that can accommodate many services at each layer. Many services will be provided in the Applications Layer. The services will be optional depending on the applications users will perform on the network. The framework will support all kinds of media for transmitting data, and is meant to support all the standards developed by such organizations as CCITT and EIA. Optional standards will continue to be added.

Eventually, the OSI model will be specified in a precise mathematical language so it can be interpreted exactly. Also, verification procedures will be published so vendors can be sure their networks meet the specification. Because some services will be optional and the model will change over time, the vendor will specify whether its network conforms to the OSI model version **x** level **y**, where **x** will be a time variable and **y** an option variable.

The data communication industry is developing so rapidly and in so many directions, the OSI model will change over time to remain current. The basic structure will remain the same, but optional services will be added at various levels to respond to changes. The model was designed to be flexible and adaptable. That was the original reason for the layered architecture, to accommodate changes within a layer without affecting the other layers. The Physical and Applications Layers will have the most implementation options because users will perform a diversity of

The OSI Reference Model—A Blueprint For Network Communication

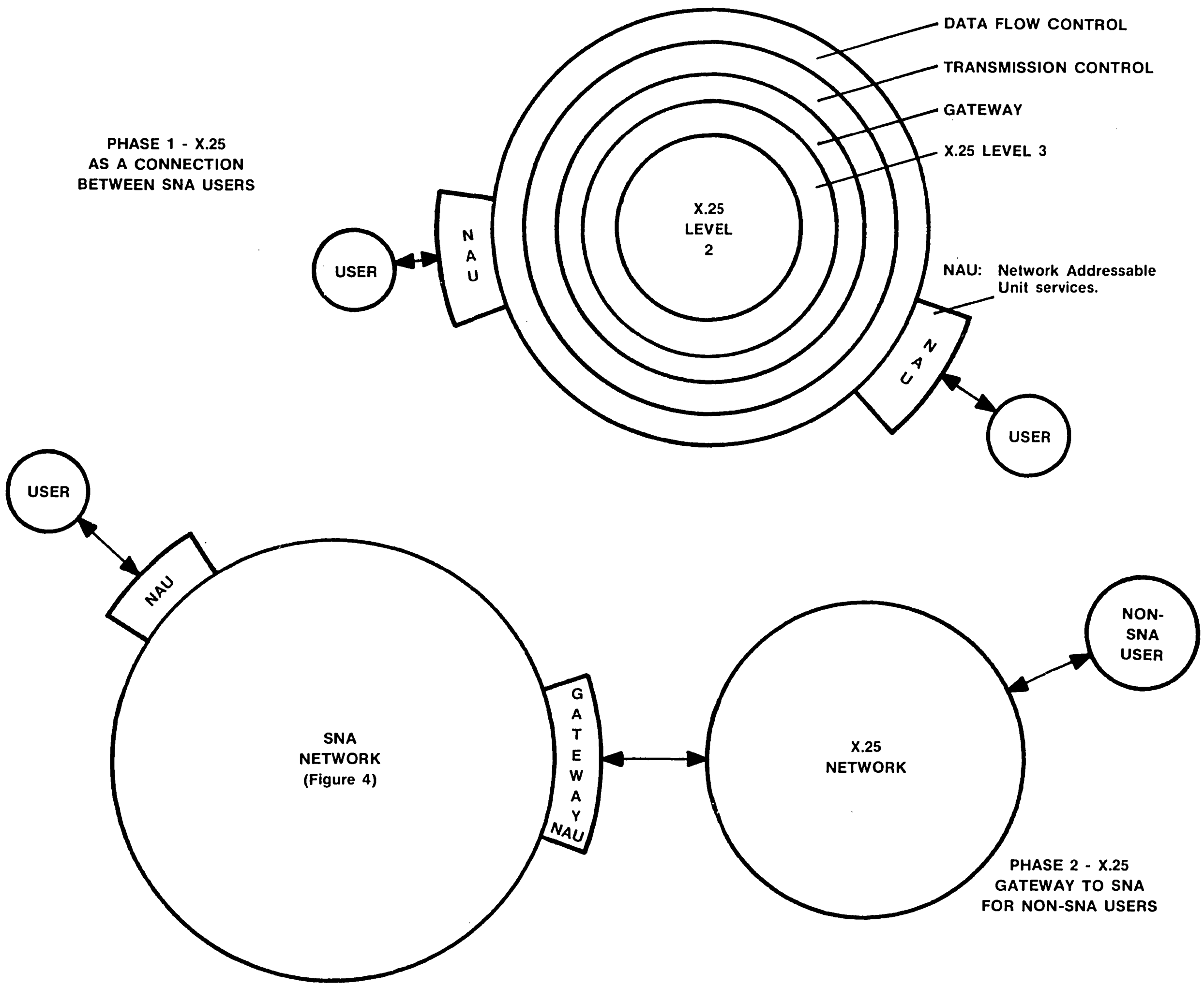


Figure 10 • IBM X.25 support models.

applications and the communications industry will provide a variety of media for communication links.

The OSI model is now a reality. Most standards will be in place by 1985. The first demonstration of an OSI network will be presented at the 1984 National Computer Conference (NCC). It will be somewhat restricted, but it will be the first one ever demonstrated. The OSI model is here to stay for the projected future.

• END

PC-to-Mainframe Communication

■ INTRODUCTION

The increase in use of personal computers in corporations with existing data processing facilities has already raised questions on the coexistence of distributed processing power in PC form and central data centers. The existence of local PC data collection naturally opens questions on the ability to capture PC data directly for data center applications, or even to use PCs in conjunction with traditional DP equipment for efficient data collection. But there are major questions of data processing policy to be resolved in PC use. One of the most critical issues is the maintenance of data control in such a mixed environment. Many data processing experts fear that PCs may weaken central management information systems by diverting data from the corporate database, or that dissemination of information to so many separate processing points will result in local distortions or inconsistencies.

Providing a link between personal computers and the corporate data center can be a significant step in optimizing PC usage and in controlling the integrity of the total corporate information base. To perform effectively, however, the communication link must bridge the very different environments of personal computer and mini-computer or mainframe. This report discusses the problems in connecting personal computers to mainframes, alternative connection technologies, and factors which affect the selection of an ideal link between PC and data center.

■ PERSONAL COMPUTERS AND DATA DISPERSION

Today's business environment is increasingly dependent on the productivity and strategies of its professional, technical, and managerial personnel. While automation in support of these individuals is not a recent development, the advent of personal computer systems has made it possible to provide key people with a locally controlled level of computational power which equals that of an entire data center of twenty years ago. The explosive growth of personal computer usage by key corporate personnel can be attributed to corporate recognition that such power, used in support of planning and operational decisions, can reduce costs, increase profits, and free decision-makers to apply themselves to other projects in support of corporate goals.

The microcomputer has also altered the price/performance relationships on which the decisions for clerical automation have been based. Data entry productivity can be increased manifold by providing local screen formatting and data entry; and local systems for such functions as document retrieval can often reduce department clerical effort significantly. Personal computers based on current microprocessor technologies are often less expensive than terminals into data center mainframes, and package software which supports the users' needs can be purchased from a retail store rather than developed at considerable time and expense by corporate data processing organizations.

There is a price for the productivity gains of personal computers, however, and it 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. Personal computers can "shortstop" vital information before it reaches the data center, making it unavailable to corporate MIS. A department may be able to maintain its own budgets more efficiently on a local personal computer, but is that gain in efficiency worth the loss of the information to the business as a whole?

Even local departments may suffer from some problems associated with local personal computer data storage. User organizations normally lack the DP department's familiarity with the needs for proper editing of information and controls to insure reliability and consistency of data. It may be perfectly obvious to a local department supervisor that an expense report date cannot be in the future, but unless that fact is related to the program gathering expense information, that restriction cannot be assured in data being entered on the file. Reports which display expense data in current-month, last-month form may fail to show the item at all, and it will almost certainly be in the wrong place when it eventually does appear. Lack of a central collection point for data may result in multiple "versions" of information in the personal computer files of several departments, or even in several systems in the same department. These inconsistencies and editing problems can be disastrous if attempts are made to collect and use the information at a later date.

■ USING DATA COMMUNICATION TO HALT THE "DATA STAMPEDE"

A logical answer to many of the data integrity and consistency problems, and to the need for central information management and control, would appear to be connection of the personal computers in use within a corporation to that company's data center through some form of communication link. The connection could then be used to gain access to primary data bases in the data center, eliminating the need for local storage of information. A file could be loaded to the personal computer when needed and returned when updated, or a direct dialog between data center file manager and personal computer user could be accommodated for item-by-item access and update.

Personal computers and the data center can relate to one-another in several basic ways, as shown in Figure 1. First, they can link together using a "batch" protocol when required to exchange files. This process substitutes communication paths for a form of compatible storage media such as disk or magnetic tape. If a file could be carried from a PC to the data center on a floppy disk, file

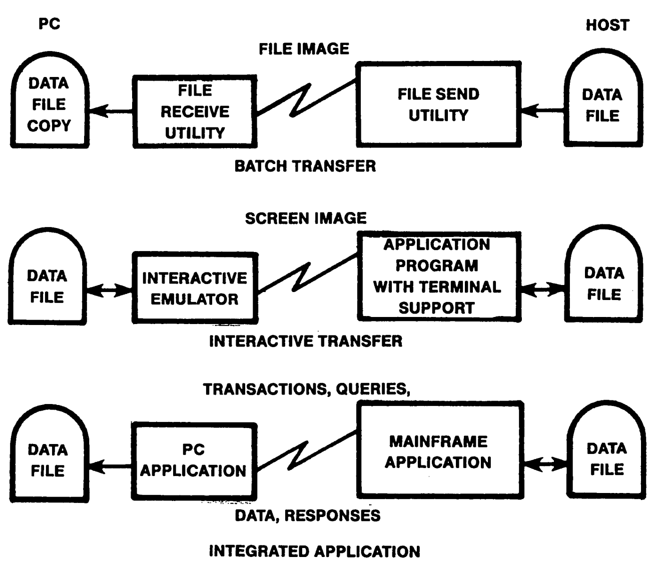


Figure 1 • PC/ mainframe relationships.

PC-to-Mainframe Communication

exchange would actually be easier than this form of communication. A second option is to attach the PC by emulating an interactive terminal on the data center system. This gives the local PC operator a means of interacting as a terminal user, and may be extended to provide programs on the PC with the ability to interface through the "pseudo-terminal" as well. File transfer may still be accomplished through terminal emulation if the data center software runs a form of terminal-to-disk copy program. The final option for attachment of a personal computer involves software on the host which works with software running on the PC to provide a pathway between systems. Data and even requests for processing can pass over the pathway, and the PC/data center form a single unit co-operatively working to satisfy the user's needs.

Ideally, corporations would like microcomputers and data center equipment to relate to each other in a way which 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. If this is to be made possible, there are some basic rules about the data link between the systems which must be enforced:

- The communication technique used must be compatible with the support currently available at the data center, or at least must be supportable with upgrades to that equipment.
- The personal computer must emulate a form of data interaction which is acceptable to the communication software in use or available. A data center which supports only interactive terminals, for example, may have difficulties supporting personal computers operating in batch mode even if the hardware at the data center would permit the connection.
- The information exchanged must be in a form accessible for processing at the destination with a minimum of additional processing. This may mean code set translation since most personal computers use the ASCII character set and IBM mainframes use EBCDIC.
- The operation of the communication link must be as transparent to the non-technical PC user as possible, and must disrupt the normal flow of the PC application as little as possible to preserve the local ease of use which probably justified the use of PCs in the first place.

The best approach to providing a facility such as a PC/data center link is to examine the functional requirements and select a method which satisfies them within reasonable economic restraints. This approach may be excessively idealistic for the micro/mainframe communication problem because the data center equipment tends to be a "given" because it is already in place, and the capabilities of microcomputers in communication may be very limited. Given these restrictions, it is probably best for the user to evaluate the characteristics of the data center environment and the microcomputer communication environment to establish the framework within which a flexible communication facility can be developed.

■ THE MAINFRAME COMMUNICATION ENVIRONMENT

Data center equipment, being more expensive and supporting the mainstream of corporate data processing, normally establishes the basic communication environment. While some accommodation to personal computer communication can often be made, the cost of major changes to the data center communication facilities can often exceed the cost of personal computers. If PCs are to successfully communicate with the data center, the first step is to determine the communication parameters which the PCs must satisfy to do so.

Data communication support in existing facilities can be classified according to the type of exchange of information supported. Most modern data centers which provide any form of communication support do so for INTERACTIVE TERMINALS such as CRTs. These devices provide the means for direct user entry of data into the central systems, a substitute for keypunch and other means of off-line data collection. Some data centers may also support BATCH transmission of information, either because off-line data collection is still in use or in order to support direct file transfers from computer to computer. Often batch and interactive communication use different communication protocols, but even where the protocols used are compatible, batch and interactive communication may involve different communication software in the host system. Application software will almost certainly be different. A major bank which uses CRTs for data entry can interface very easily to a personal computer which can appear to the data center as a CRT because the software used for the application is designed to deliver data summaries in response to a terminal inquiry. The same system, presented with a batch of inquiries in a single transmission, lacks the software to produce a batch of status reports as a single-transmission response. MOST SUCCESSFUL PC-TO-DATA-CENTER CONNECTIONS WILL EMULATE APPLICATION COMMUNICATION DIALOGS WHICH ARE ALREADY SUPPORTED BY THE DATA CENTER. In all cases, it is important to know the communication software and application level support available at the data center before considering any particular method of PC communication.

Given the goal of emulating some form of existing application dialog, the setup of a PC to data center link must consider the following characteristics of such a dialog:

- **Protocol.** Most minicomputers and mainframes have a preferred communication protocol—IBM for example normally uses its System Network Architecture (SNA). The PCs communicating to mainframes MUST use a protocol which is supported by the mainframe, but SHOULD use the particular protocol already in use at the data center. Some systems, particularly the larger mainframes, are somewhat terminal independent within the group of terminal types supported. These systems may even permit the substitution of terminals of one protocol for those of another within the APPLICATION-LEVEL dialog. For these systems, protocol can be a secondary factor in selection of a communication technique, as long as the protocols being considered can all be supported in the same way at the application level.
- **Attachment mode.** Conventional point-to-point RS-232 cables or phone circuits and modems are most often used to attach PCs to data center computers. Some types of equipment may be usable in multipoint mode on circuits already supporting native terminal devices. Other more exotic methods of connection are sometimes available. IBM systems may support the attachment of PCs to 3270 controller devices directly via coaxial cable, and DEC computers may be able to accept Ethernet attachment of some personal computers. Most of these alternative attachment methods offer significant advantages, either in data rates or economy, and should be seriously considered where available.
- **Facility layout.** Personal computers will rarely be used in close proximity to the data center, so consideration must be given for the impact of the relative locations of the equipment. This is particularly important where attachment other than via conventional phone/modem means is contemplated. RS-232 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 may offer greater distances, these may not be supported either at the PC or on the data center equipment.
- **Speed.** The data rate of the data center communication equipment may be set by the requirements of other terminal devices or connected computer systems. If the data rate is high (above 1200 pbs) it may restrict some personal computer attachments because popular PC modems and other interface devices normally operate at 1200 bits per second or less.
- **CPU port capacity.** Most data centers have a few spare hardware ports for the attachment of new communication devices, but there may be several types of ports, each with their own restrictions and characteristics. The cost of adding new ports to a CPU or data communication controller may be quite high, especially in the case of protocols such as SNA or IBM's older binary synchronous protocol. You may want to review the CAPACITY of the data center system for communication ports of each type, as well as for the number of currently available spares. It would probably not be wise

PC-to-Mainframe Communication

to dedicate the last SNA port in an operation heavily dependent on SNA communications to a personal computer!

While a data center environment may be restrictive in terms of its ability to support PC communication, it is at least normally designed for SOME form of communication support. Data center computers are usually designed to support multiple users and to share system resources among a large number of separate "tasks". Adding personal computer connections to such an environment is largely a matter of matching communication capabilities on both ends and of allocating the required resources. If the economics of the PC link application justify it, the data center can shift policy somewhat to accommodate a reasonable PC interface method. The key question is often whether the PC can supply such a method.

■ MICROCOMPUTERS AND DATA CENTER COMMUNICATION

There are three primary factors to consider in personal computer architecture when data connections to a host computer are required:

- The ability of the personal computer to support communication functions at the same time as it supports local use.
- The support for communication protocols and terminal functions which match the expectations of the data center environment.
- The level of integration between PC application software and data center software to support co-operative tasks and to permit transparent file sharing.

Data communication is a function which is often called "real-time," meaning that communication activities are almost like human activities; they tend to occur in response to outside influences which cannot easily be scheduled or predicted, and once they occur they must be handled promptly. A microcomputer system which serves a local user may or may not have the ability to concurrently serve a communication link to a host. Microcomputers of the type normally used as personal computers or executive workstations were designed primarily as standalone systems. The operating systems used on the systems probably do not support multiple users or multiple "tasks," actions which are logically separate and could take place concurrently but are associated with a single user. The difference between single-task and multitask systems can be seen in Figure 2. A single-task computer can undertake only one function at a time—word processing, spreadsheet, database, or communication. The user must decide when to invoke each of these functions, based on current needs. This type of serial scheduling of work is normally suitable for human interactions; people are also typically single-thread in work habits. But even with non-communication environments there are times when you probably could have used a quick access to a spreadsheet program in the middle of writing a large report on a word processor. To get it, you must probably exit the word processor, saving your work, and enter the spreadsheet program. Communication interactions are quite different. Communicating between micro and data center is not an end to itself, but a step in the handling of information which has company-wide meaning. As such, it may occur as a PART of any other job and not likely as a job in itself. Putting it another way, communicating between personal computer and data center is naturally an INTERACTIVE kind of task. Single-task personal computers, on the other hand, are BATCH oriented. Requests for communication service are normally made outside the application, leaving the user with the problem of integrating the data received with local work or of extracting and formatting data to send.

Multiple 1=1 task systems, as shown in Figure 2, address this problem by making the communication program coexist with the application. Services can be requested at any time directly from the user's application program. Since this permits a program to request communication directly, it supports the transparent use of communication links to the data center as a part of any user application, should that application be written to use such a link. Multiple-task communication may be provided on personal computers which are designed for multiple-task operation, or vendor software may establish a separate "communication" task in an

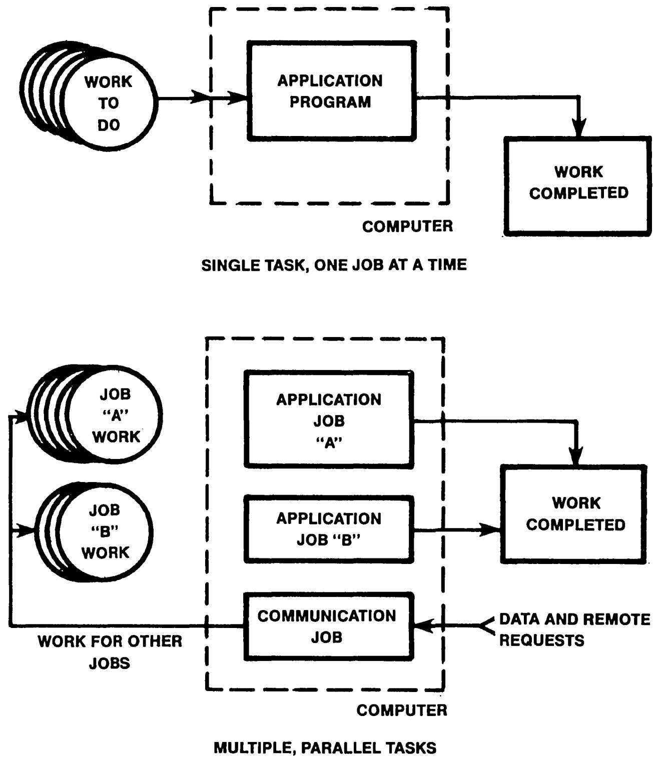


Figure 2 • work flow in single and multitask systems.

otherwise single-task environment. Figure 3 shows the structure of a popular 3270 terminal emulator system available for the IBM Personal Computer. The system permits the user to "install" a copy of the emulator software in the personal computer memory ALONG WITH THE USER PROGRAM BEING RUN. This resident copy can then be used by programming languages such as BASIC to access data center information, or the operator can key a special sequence of characters to switch between the emulator and the user program.

Both multitask personal computer systems which permit a continuously active communication function and "co-resident" communication programs, may permit the operator to switch keying and display between local programs and emulation of a data center terminal. This permits the operator to interlace local and remote functions, but only by acting as the common element in two dialogs—one with the local application on the PC and the other with the data center application. If PROGRAM-LEVEL interchange between data center and PC is desired, not only must the communication support on the PC provide facilities for it, the application programs must use those facilities. It is this factor which makes multitask operating systems or other forms of full-time communication support from the PC vendor superior to any form of "add-on" communication support. Software developers will not undertake an incremental expense to support the new XYZ Corporation communication board and software unless they believe that a significant number of personal computer users to whom they target their product will buy and use XYZ Corp's board. The PC vendor's standard features have a good chance of gaining widespread support, and optional features which prove popular have some chance. It is very unlikely that a significant number of PC application programs will support communication capabilities offered by a third-party vendor.

TECHNICAL FACTORS OF THE COMMUNICATION ENVIRONMENT may have a significant impact on PC connections to the data center. This is particularly important for users of non-IBM, non-DEC systems, since most popular "terminal emulation" systems

PC-to-Mainframe Communication

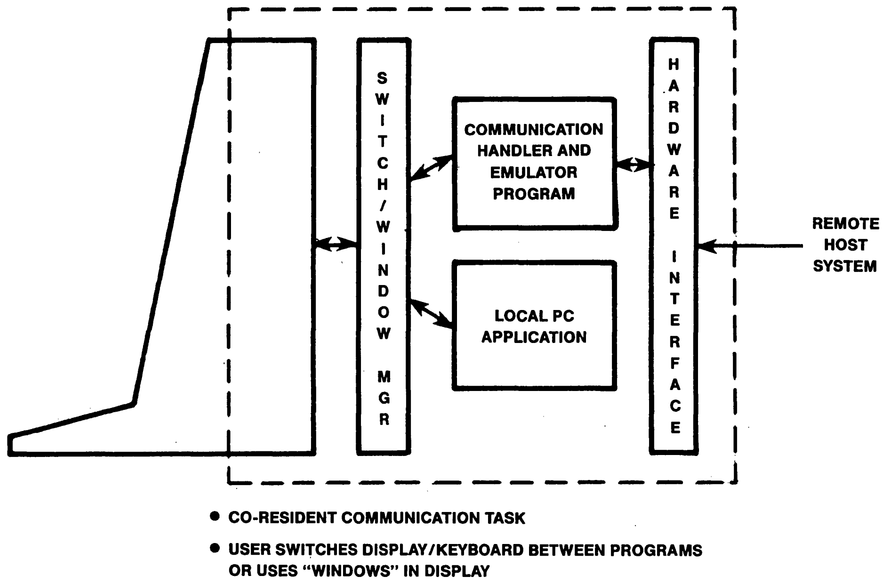


Figure 3 • a popular terminal emulator structure.

tend to emulate IBM or DEC terminals. The physical support for communication on a micro must suit the data center to which connection is to be made. That support is often envisioned as a "board"—hardware additions to the personal computer with the proper interface for the communication system involved. This is rarely true. ALL PRODUCTS WHICH SUPPORT CONNECTION OF PCs TO DATA CENTER COMPUTERS WILL REQUIRE A SOFTWARE COMPONENT. Selection of a package for communication support must be made by evaluating both the hardware and the software elements of the package. A package which consists only of hardware is logically incomplete, since SOMEONE must provide program support for the actual communication. If the vendor does not, you are elected. Packages which consist of only software are more common, and may be useful if the hardware on which they depend is either standard equipment on the personal computer being used or a commonly-used option.

Hardware options for PC attachment to the data center are shown in Figure 4. First, the communication hardware provides the means of physically attaching the link to the host computer and controlling the signals on that link. This attachment may take several forms:

- A standard communication interface such as RS-232C, used to attach the PC to a host system directly via cable or through a modem.
- A coaxial cable or other special interface to a COMMUNICATION CONTROLLER, which allows the PC to attach as part of a cluster of terminals. The IBM 3270 display system attachment of CRTs to a cluster controller is an example of this type of connection.
- An interface to a vendor-proprietary or industry-standard bus or local network architecture. Examples of this type of interface are the various Ethernet connections or the IEEE-488/Hewlett-Packard Instrumentation Bus.

Selection of a personal computer attachment hardware architecture is normally dictated by the capabilities of the data center. Where several options exist, the flexibility of the approach (communication interface via modem is suitable for many environments while bus connection is a local option only) versus the performance and convenience of the interface (coaxial or bus attachment may support much higher data rates and require no additional CPU communication ports or software support) must be evaluated.

OPERATIONAL ISSUES IN PC LINKS TO MAINFRAMES may outweigh the technical ones. A personal computer acting as a mainframe terminal is normally designed to emulate the functions of the terminal as well as the technical characteristics. This may mean that the PC user will have a set of operating rules for the stand-alone environment, established by the local software which is normally used, and one or more additional sets of rules associated with each mainframe and perhaps each application interaction on that mainframe. A New York bank PC operator, having spent the morning switching between 3270 emulation to an IBM host and a local spreadsheet package was heard to remark, upon sitting down at the computer after a break, "Let's see—who am I now?"

■ MICRO-MAINFRAME LINKS—WHAT ARE MY CHOICES?

A user faced with the need to select a means of linking personal computers 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 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 equipment and software needed can be acquired from a single source or multiple sources, and those sources can be computer vendors, modem vendors, third-party hardware or software suppliers, or system integrators.
- **Information management and control.** Figure 5 shows some of the forms of PC data interaction which can take place, and the impact on data center files. PC data can be maintained entirely on data center files and updated only through data center applications. This restricts the PC to a data manipulation and reporting role. That role can be expanded by allowing local data entry and processing to take place, with reconciliation of information with the common data bank taking place either interactively or

PC-to-Mainframe Communication

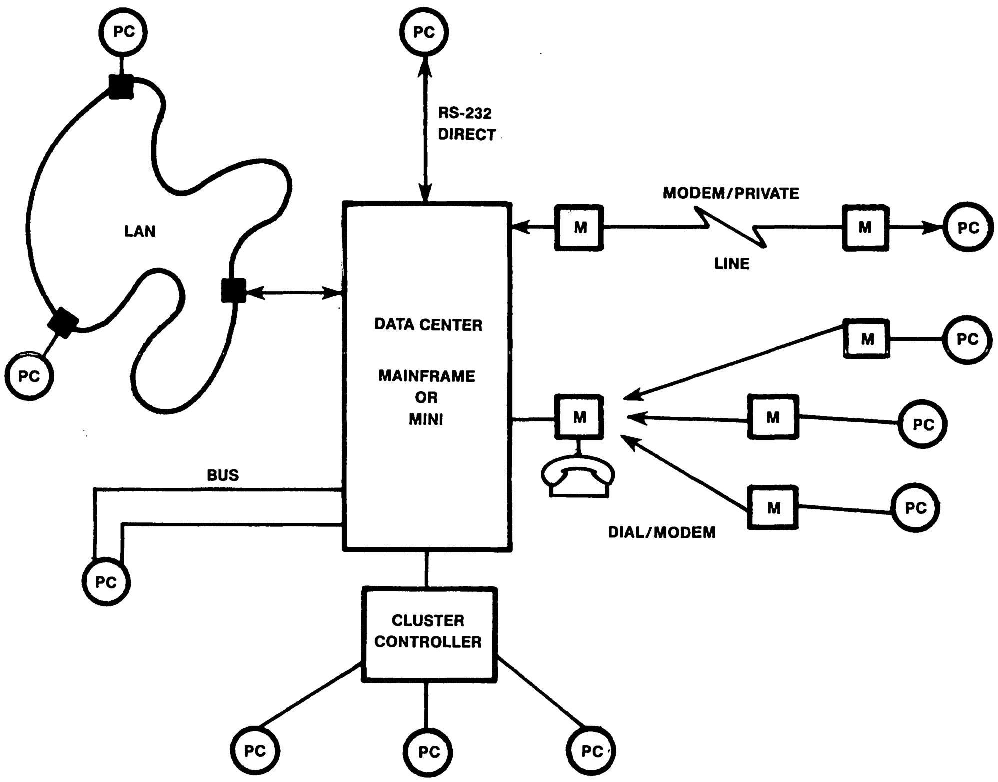


Figure 4 • hardware connection options.

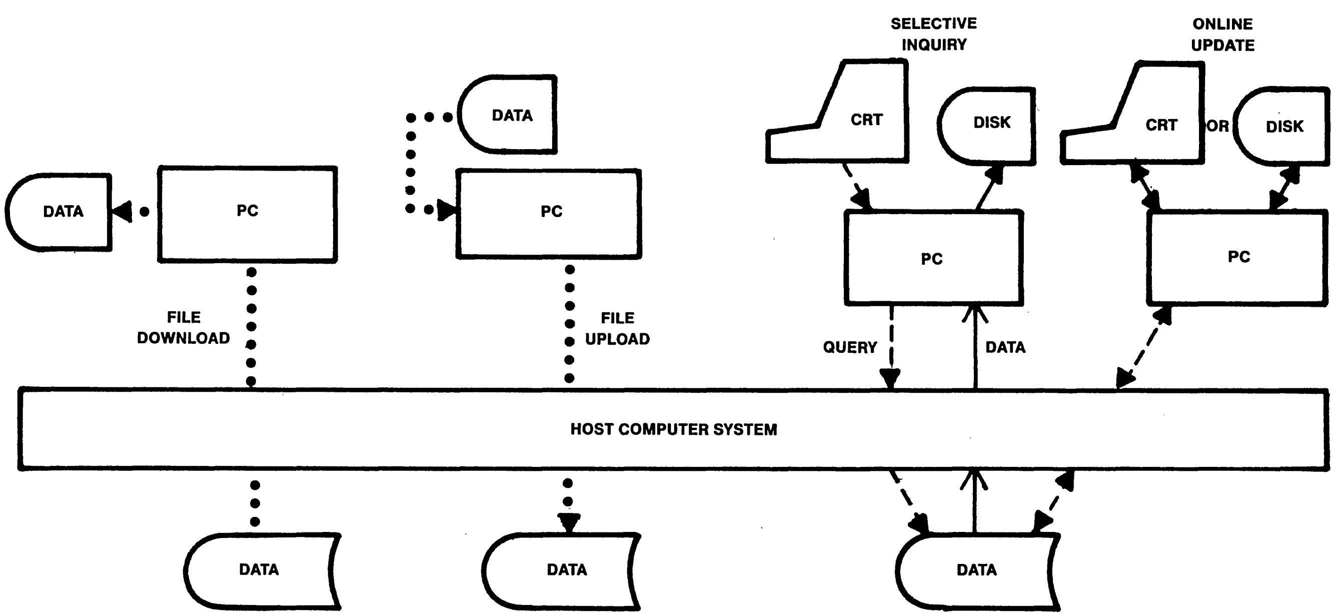


Figure 5 • file interactions with remote PCs.

when the data is ultimately loaded into the data center files. If local update of existing data center files through an upload-modify-download process is also to be permitted, some form of protection must be applied to prevent concurrent updates of the same data and to warn other users of information that a copy is outstanding for update purposes.

PC-to-Mainframe Communication

■ TERMINAL EMULATION SYSTEM

Most users will probably find that connection between their microcomputers and the data center will be supported by having the microcomputer emulate a terminal which is supported by the data center. The key questions in such applications are the type of terminal to emulate and the way in which local applications can utilize the emulator interface.

Most users have some form of CRT terminal on the data center equipment already, supporting online applications or programming. Emulating one of the existing terminal types would permit attachment with a minimum number of changes to the data center communication software, a task which often makes PC considerations pale into insignificance. IBM computer systems will normally be best served by the emulation of interactive display terminals of the 3270 family or the 5250 family. DEC systems operate best with terminals emulating the VT52 or VT100 terminals. Some IBM users may also find products which emulate the batch workstations such as the 2780 or 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 applied. If the PC and the mainframe are relatively close together, a simple interface cable can be used to connect them. BE CAREFUL ON DIRECT CONNECTION WHERE THE EMULATED TERMINAL HAS A SYNCHRONOUS PROTOCOL. SYNCHRONOUS LINKS REQUIRE A CLOCK SIGNAL, AND NOT ALL PC TERMINAL EMULATORS CAN SUPPLY A CLOCK. You may need a modem eliminator on such connection. IBM 3270, 2780/3780, and 5250 devices are ALL synchronous.

Modem connections will normally work between the PC and the mainframe, but the type of modem will depend on the distance. Limited-distance or short-haul modems, sometimes called line drivers, will serve best for connections within a building. Your existing telephone wiring may be able to handle everything, or new wires may be needed. The type of modem and its speed must match the terminal being emulated. Try to supply the highest speed supported by both PC and mainframe—nobody ever complains that response time is too fast!

IBM 3270 emulators may directly connect to the host computer using either the binary synchronous or SDLC/SNA protocols, or they may attach via coaxial cable to one of the 3270 family's cluster controllers. The coaxial attachment method may reduce your installation costs considerably if you have a cluster controller in the area of the PC. Check with the PC and cluster controller specifications to be sure that the cable run is not too long, that the terminal type emulated by the PC package is supported, and that the host software has been modified to add the new terminal address if required. You may want to consider buying a new cluster controller just to attach PCs if there are a large number expected from a single area of the building. No modem is required to attach a PC to a cluster controller via coaxial cable.

DEC users may find that Ethernet attachment to the DEC host is an attractive option, providing that DECnet support for Ethernet is available in the data center. In cases where Ethernet is already used or where a large number of PCs will be emulating DEC terminals, Ethernet may be an economical connection option. Several vendors supply personal computer links to Ethernet, and the data rates supported by these products are high. Since Ethernet attachment is not as common as the more traditional RS-232 or modem attachment, the ability to select products based on features will be restricted.

Terminal emulator products will rarely emulate EVERY aspect of the target terminal. Keyboard layout is seldom flexible on PCs, so emulators are rarely able to do more than approximate the actual device. This may result only in small adjustment factors for operators, but sometimes terminals have a bank of function keys which cannot all be duplicated on the PC. Loss of some of these keys may affect the operation of some applications. Screen display characteristics, particularly in graphic mode, may also differ, making the emulator unsuitable for some specialized functions. If the target terminal has a status line, be sure that it is emulated in an understandable form. It is important to review OPERATOR factors and APPLICATION factors as well as technical factors in selecting a terminal emulator product.

The most significant problem in the use of terminal emulators may be in the area of data control. When a PC is used to capture a segment of a mainframe data base and move it to local storage, that data becomes subject to manipulation OUTSIDE THE APPLICATION FRAMEWORK WHICH WAS INTENDED TO VALIDATE IT. In many cases, the movement of data to a PC, its editing, and the return to the host will bypass any host editing or validation. 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 polluted with questionable information.

The problem is magnified if several PCs are allowed access to the information. A major insurance company had an application where file segments were sent to personal computers via a batch terminal emulator and processed locally. There was no protection against several PC users calling for the SAME FILE SEGMENT, and when the data was returned to the file only the last user's changes were reflected.

There are some steps which can be taken to reduce the risk of corrupting data center files through the use of personal computers emulating terminals.

- Limit "out-of-data-center" updates. If personal computer copies of files are used only for access, there is no danger of poor edit control producing bad data.
- Require that update applications be reviewed by the data processing organization to certify that the data quality assurance measures taken by the data center system are also applied at the PC.
- Provide a "staging" area where PC files which 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 which can be loaded into PCs to prevent multiple users from requesting the same data elements at the same time.

SELECTION OF A TERMINAL EMULATOR should consider the following points:

- Does the product emulate a terminal type and protocol which are already supported at the data center, or which can be supported easily?
- 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?
- Is the product compatible with the PC or PCs used and with any special hardware or software which is already in use on the systems?
- Is the mode of operation of the product suitable for the level of operator on the PC and for the type of work which is expected to be done with the product?
- Are the operating procedures so different from those associated with the PC's local operation that they will be difficult to complete properly?
- Can the combined environment created by the product between the PC and the host be controlled from the viewpoint of access security, data integrity, and information resource management?

IF SEVERAL PRODUCTS ARE GENERALLY SUITABLE FOR USE, the following guidelines may help narrow the choice:

- If the PC vendor supplies a package which meets your needs, 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.
- If the communication hardware supplied by the PC vendor is supported by a software-only product supplied by a third party, give that combination preference over another

PC-to-Mainframe Communication

product where both hardware and software are third-party products.

- Avoid products which rely on the combination of third-party software from one vendor and hardware from another third-party vendor. How many support people do you really want to talk to?
- Select a product with an interface to an external modem over one with similar features but using an integral modem. Modem technology is changing very rapidly, and the non-modem product will track those changes if you desire.
- If a product contains an interface board which plugs into the PC directly, buy from a company which supplies many types of such boards over one who supplies only the board used in the product. Broad familiarity with the PC's bus architecture will translate into fewer potential problems in coexisting with other board-level products.
- Select products which ENHANCE the capabilities of the device emulated without requiring host support for those enhancements. Integral printing support for terminals which cannot normally drive a satellite printer is an example of this. Beware of buying on gimmicks, however.

■ THE INTEGRATED APPLICATION SOLUTION

Terminal emulators have several significant disadvantages. Their basic design causes them to emulate a device which possesses no local processing or storage power, and therefore the interactions which the host supports with them will not easily take advantage of such capabilities. Terminals also have logical design restrictions, limitations in functionality which can be safely accepted since the human operator's range of supported functions is also limited. The limitations may be unacceptable when a PC is substituted for the terminal. For example, a PC emulating a 3270 terminal device probably cannot transfer binary data files because binary data is not presentable to an operator and is not supported as input to or output from most interactive terminals. Extension of the power of the PC through such devices will normally require special programs and operational controls in the data center to permit free data exchange and preserve data integrity.

The problems with projecting PC power can be solved by applying some of the principles of distributed processing to the PC environment. If software in a personal computer can link to similar software on a mainframe, the resulting INTEGRATED package can work to satisfy user needs somewhat independent of the exact location of the resources which must be applied. Figure 6 shows an example of the structure of an integrated software program package designed to permit personal computer users to interact with their data center mainframe. In contrast to the terminal emulators, which place the human operator of the PC into communication with the host application, the integrated applications provide an "application manager" element at both ends of the connection. This element-pair manages the communication path on behalf of the two systems and provides the services of data exchange and process requests to either system. The complexity of such an environment is directly proportional to the complexity of the applications managed, so cost constraints act to limit the scope of integrated software systems. MOST INTEGRATED SOFTWARE PACKAGES ARE LIMITED TO ONE OR A FEW APPLICATIONS, AND THE MORE SOPHISTICATED A SINGLE APPLICATION SUPPORT FACILITY IS, THE FEWER SUCH FACILITIES ARE LIKELY TO BE PRESENT.

The operation of the integrated application can be understood by reference to Figure 6:

- The operator makes a request at the PC keyboard, which is received by the local application manager. The manager determines that the request is one for a local PROCESS and directs it to the proper software element in the PC.
- The PROCESS ELEMENT determines that it requires data for the request and that need is communicated back to the application manager, who determines that the data is not local but resident in the mainframe.
- The application manager requests its opposite number running in the mainframe to provide the data. The mainframe application manager makes its own local PROCESS ELEMENT request to do so, and sends the data back over the communication link.
- The PC application manager delivers the data to the PC PROCESS ELEMENT, and that data is used to satisfy the user request.

Remote service requests can likewise be transferred over the communication path to the host system, and the data which was supplied by the host can be stored locally for use later in the same session. If the data is updated rather than just accessed, the PC

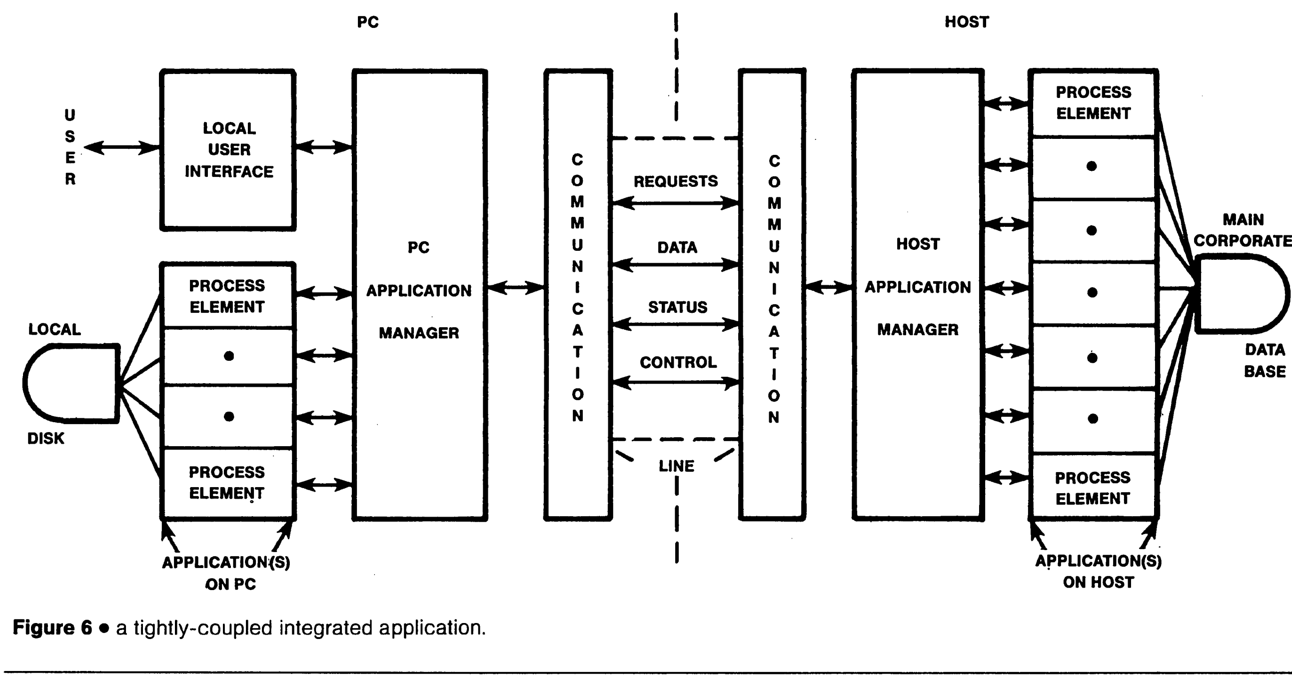


Figure 6 • a tightly-coupled integrated application.

PC-to-Mainframe Communication

application manager can route the update through to the mainframe, where all the proper controls for changing the database can be applied.

In some systems the structure of the software and the flow of information is much as described above. Other systems use a "loose" structure, where an application manager at each end of the connection acts as a "data filter" between the PC and the mainframe. Figure 7 shows this type of system. In it, a PC user makes a data request, often as a "query" in a database language. This query results in a request for data, which is satisfied by downloading information from the host to a PC file. That file is then processed locally on the PC using standard programs and uploaded when update of the mainframe data base is desired. This style of integration is more easily implemented, but often offers the user little more capabilities than a terminal emulator. Since data is NOT exchanged interactively, the problems of data integrity and loss of operator attention due to changes in job context are likely to remain.

Most users find this concept of application integration so attractive that they wonder why anyone would ever elect to use another method of PC communication with data center systems. The answer is that integrated applications are extremely rare, a fact that relates to the differences between such applications and the normal PC and mainframe software.

Figures 3, 6, and 7 which show the comparative structure of a terminal emulator application and an integrated application illustrates these differences. While the PROCESS 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 which are generated internally by the software and not through the action of a terminal operator. Each PROCESS ELEMENT in an integrated application must be very modular in structure, because it may perform only a small part of a total task which may span several computer systems. It is more difficult to design and program such structures, and therefore more expensive. Personal computer software has developed in a simple, single-user environment, and most of the popular software available is not easily adapted to integration with mainframe software, a fault which is equally applicable to the mainframe products. Some progressive vendors are already producing integrated systems, and others are working to reconcile the two computer environments with modified or totally new software. The market trend seems to be clearly taking the direction of the integrated application.

TECHNICAL COMMUNICATION SUPPORT SELECTION with integrated application software is usually a matter of following the requirements stated by the supplier. Most of these packages will NOT support wide varieties of communication 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 would be advised to prioritize their needs for such software and watch for incompatibilities in supporting hardware. In general, the systems and hardware with the best chance of having a wide range of compatible integrated applications are those where the PC vendor itself has provided proper technical facilities for mainframe communication. Software vendors, given a consistant and preferred hardware environment, will tend to build to it. The IBM PC is an example of this; IBM's own host communication support is the preferred attachment method for most of the PC's integrated application software.

DATA VALIDATION AND CONTROL in an integrated environment may be totally within the control of the package, making data handling at the PC as safe 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 fulfilled in practice. As the level of integration decreases, from the "ideal" tight coupling shown in Figure 6 to the looser structure shown in Figure 7, the ability of the application to manage the data flow and to insure that proper validation rules are applied at each update decreases. 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 communication 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 which provide almost transparent host access and full distribution of files and computer resources are likely to be restricted to a single application, but 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 corruption of data through poor control on the PC is significantly higher. Users must evaluate the benefits of wide application flexibility against the potential loss of control.

EVALUATING AN INTEGRATED APPLICATION can be

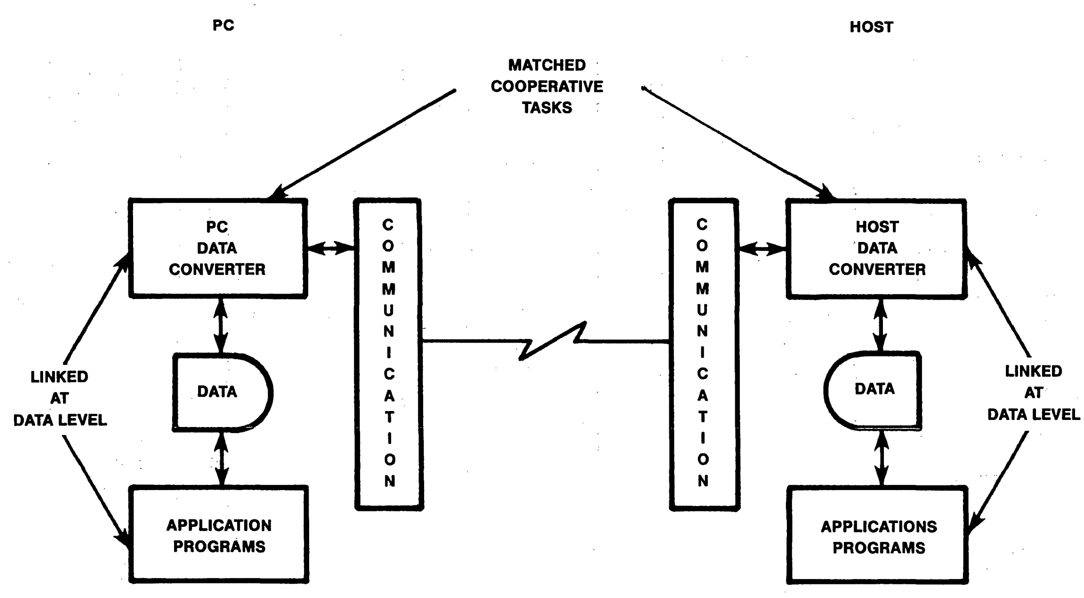


Figure 7 • a loosely coupled integrated application.

PC-to-Mainframe Communication

approached in phases. 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. 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 the user with an idea of the current environment for the task and the investment in it, and of the way in which the task should operate, economic constraints aside.

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. When a list has been compiled, each product should be subject to an examination based on the following considerations:

- **Host hardware and software requirements.** The corporate investment in any mainframe packages already in place in the application area is likely to be impressive, and the value of archival data accumulated through the life of the present system may be very high. Integrated software which requires a major change in the host environment, while sometimes justified, is usually just too expensive an option to consider.

- **PC requirements.** The penetration of PCs into the corporate environment has proceeded faster in some organizations than others. Companies and divisions with a large PC investment will have to consider the impact of each potential integrated application on their PC users. These impacts may range from fundamental issues of compatibility (it won't work with the existing computer base, or it will) through configuration questions such as the addition of memory or disk storage to each system, to operational questions relating to the software already used or the way in which the systems are run. Each package considered should be rated in its impact on the PC environment in the key areas of basic compatibility, hardware configuration requirements, adaptability to currently used software, and operational considerations.

- **Level of integration.** Packages will vary according to the degree to which the PC and host software is tied together. In a tightly integrated package the PC user will make requests without knowing where the data or resources needed to carry them out actually reside. This type of application will limit the ability of the PC user to operate with non-integrated software, and may in fact prevent use of the integrated files with any other software package. Loosely integrated applications will generally provide only a facility to download data from the host, translate it into a usable PC format for local processing with many different software packages, then permit its uploading to the host to replace or update the database. Highly integrated packages are easier for the users to handle because they present a consistent, request-oriented operating environment, but they 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 which is associated with the package.

- **Data security and information resource control.** This is perhaps the most important aspect of the evaluation process, and the most difficult to apply. DP management should normally be asked to provide input on the level of security provided with each application.

■ MAKING A FULL EVALUATION OF MAINFRAME COMMUNICATION ALTERNATIVES

The first step in any selection of mainframe communication package is to GET THE DATA CENTER MANAGEMENT INVOLVED. They should provide the information on the current data center configuration, hardware, software support, application programs, and data security measures.

Once the proper specialists have been assembled, the application should be evaluated in BOTH terminal emulation and integrated application terms. A good way to do this is to "walk through" the interactions associated with the application to get an idea of what the PC user will experience. You will probably 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. You may want to do three "walk-throughs"; one for interactive terminal emulation, one for batch terminal emulation, and one for integrated applications. Don't try to get too specific to any package here; just get an idea of whether the GENERAL interaction is acceptable.

The next step is to evaluate the data security and information resource management implications of the PC link. Ask the following questions:

1. Are requests for data made from the PC subject to satisfactory levels of access control, such as password control? The controls over PC access should be at least as good as those controlling the access of terminals which could normally be expected to query the data.

2. Can similar control be provided at the PC for operator access to data once the data is downloaded? It is undesirable to protect data in the data center and leave it available to all the world once it is loaded onto a PC.

3. Is it possible for several copies of the data to 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 divergence of information. If you cannot prevent a file from being accessed 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. Integrated applications may offer a solution to problems which come to light during discussions on this issue.

4. Can the return of data to the mainframe result in the loss of any update controls? These may arise from a lack of control in the PC 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, a relaxation of the validation rules may result in a lowering of data quality standards, and this should be SPECIFICALLY UNDERSTOOD AND APPROVED BY MANAGEMENT.

5. 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 impacts of "out-of-file" data on key reports which are run periodically for corporate management or external distribution, and conflicts may indicate that PC production schedules will need adjustment for the period near the report cycle.

The walk-through and the security and control evaluation should provide a list of requirements for the mainframe communication link, which can be added to the technical features of the PC and the data center equipment for the purpose of completing an evaluation of existing packages. The most suitable can be examined in detail or even benchmarked for final selection.

PC/MAINFRAME INTEGRATION must eventually be addressed if personal computers are to be properly used in a corporate environment. The potentials offered by widespread data exchange between the small computers and the data center in terms of productivity must be balanced against the potential effects of such an exchange on the quality of the corporate information base; a resource which is increasingly recognized as one of the principal assets of any business.

• END

Local Area Networks (LANs)—A Solution to Resource Sharing & A PBX Adjunct or Alternative

An Examination of LAN & PBX Solutions to Local Communication Issues, An Introduction to LAN Technology & Guidelines for Evaluating Your Own System

■ INTRODUCTION

Developments in the use of coaxial cable to connect terminals or computers has stirred user interest in one of the oldest problems in data communication, the connection of local devices and local computers together. While interest has been generated by cable technology, cable networks are by no means the only, and may not even be the best, solution. This report examines the problems of local communication and the technologies to provide solutions, including areas that impact the user more than the often-discussed problems of transmission methodology.

■ LOCAL DATA COMMUNICATION—THE ISSUES

Studies have shown that the majority of data communication traffic is between devices that are separated by distances **less than 10 miles**. Experts feel this figure is understated because users tend to view locally-connected devices as part of the computer facility and not as data communication equipment. As computer use has grown, local user devices have multiplied to the point where they pose problems for managers and planners of data communication facilities. Many of these problems have no parallel in conventional, long-haul, data communication.

The first local communication problem is the **number of connections**. Many relatively small corporations have local terminal populations close to the 100 mark, and some have several hundred. This user population creates a predictable load on the computer facility itself, but it also creates problems with the installation, operation, and maintenance of the terminals:

- Attaching many devices to a computer requires additional ports, likely to be expensive and unavailable in the desired quantities.
- Restrictions on cable length between terminal and computer may force an undesirable "clustering" of devices, creating an operational environment that reduces productivity.
- Calling requirements may cause problems with cable routing, local building codes, and installation costs; the cost of running a cable 100 feet in an office building may exceed the cost of the terminal being connected.

A second local communication problem is the **data rate** of the devices involved. Most locally-connected devices are operated at relatively high data rates, often 9600 bits per second (bps) or greater. These speeds are common for long-distance communication between computers or with high-performance, high-cost terminal systems, but local systems use high data rates for terminals costing less than $1,000. This is substantially less than the cost of a pair of high-speed modems, and conventional solutions for providing such speeds are unreasonably expensive.

The third local communication problem is **connectivity**. Users of the public dial telephone system or value-added networks, such as TELENET, TYMNET, or UNINET, are accustomed to selecting a destination for a data call at the terminal through either a data entry or secondary handset. Locally connected devices cabled directly into the computer ports have no such selection capability unless the computer itself acts as a data switch between devices. Users who require access to multiple computer systems within the same building may need 2 different terminals or a manual cable switch.

The requirements for length and number of cables, data rates, and connections can be used to define the characteristics a local communication facility must possess:

- connect several hundred users.
- support device-to-device distances of thousands of feet.
- attach devices to a host computer without using a host port for each terminal.
- reduce the number of individual cables required for connection.
- handle high data rates for synchronous or asynchronous devices.
- provide ability to request a destination for a connection rather than use a fixed route.

Traditionally, LANs have been associated primarily with switching and resource sharing technologies. Many of the elements in these local networks are **dumb** devices such as terminals, thus no network strategy that places routing responsibilities on these devices can be supported. Configurations that consist of connected nodes such as those shown in **Figure 1a** are reasonable for a long-haul peer-device network environment but not feasible for LANs. The only structures that can be supported are those where a node can only accept or pass on data. These configurations, shown in **Figure 1 b, c, and d** are popularly called **star**, **bus**, or **ring** structures. In these network topologies, a terminal communicates with another device by simply sending data and no element (except the hub of the star which is assumed to be a computer) selects alternate paths to pass data on to another device.

■ DIRECT CONNECTION OF LOCAL DEVICES—THE EASIEST SOLUTION?

LAN discussions do not normally include the option to connect the local devices directly to the computer, but such configurations represent the **star** structure shown in **Figure 1 b**. In the "star," the computer operates as a message switch so that devices can interconnect. Some users find this solution satisfactory; for a small number of connections, it is cost-effective. If a device is also included to perform port-selection in direct-connect configurations, the combination serves many functions of an LAN at a reasonable cost.

Direct local connection through the normal 25-pin RS-232C interface works for distances up to 200 feet if the cable is specially selected, and both terminal and computer use the interface optimally. The RS-232C standard normally limits a connection to 50 feet, and suppliers may not support an extended cable run even though they are in technical compliance with the RS-232C standard. Such direct-connect cables combined with port-selection and/or port contention devices at the computer site can give users access to multiple computers or allow a large terminal universe to compete for fewer computer ports.

For greater distances, the user can substitute the RS-449 interface standard. RS-449 allows data exchange at distances up to 4,000 feet. Tests show even longer distances can be traversed using special cable. RS-449 is not commonly provided on terminals and may not be available on all computer ports. Adapter units can convert the RS-232C interface to RS-449 and they may be an economical solution for users with local network problems primarily centered around connection lengths. Like RS-232C, RS-449 can be combined with port selection/contention devices to provide switching and reduce the number of CPU ports needed to support the terminals.

Data rates on local connections are limited by the interface

Local Area Networks (LANs)—A Solution to Resource Sharing & A PBX Adjunct or Alternative

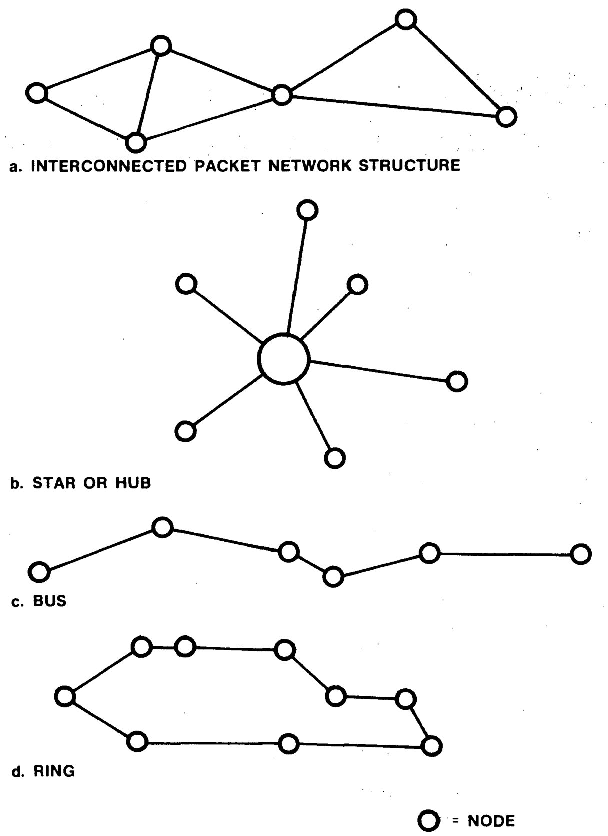


Figure 1 • network topologies.

standard. RS-232C provides service to 19.2K bps; RS-449 can operate at rates up to 10M bps. Actual data rate depends on the distance, cable characteristics, and host and terminal interface implementations. RS-449 has RS-422 and RS-423 sub-standards to define signal types on critical interface circuits such as those actually used to send or receive data. RS-422 and RS-423 provide a compromise between a high-performance form of the interface and compatibility with the older RS-232C standard. RS-422 is a high-speed differential-drive connection. RS-423 is easily transformed to RS-232C for compatibility with existing equipment.

Many computer vendors have proprietary channel interfaces to attach local devices, providing high data rates over reasonable distances. Such interfaces normally operate only with the vendor's own equipment; use of switching or contention devices and equipment of other vendors may not be permitted.

Direct connection can provide an acceptable solution to local data exchange, but it has the following disadvantages:

- each device must be cabled directly to either the CPU or a port selection/contention device • cable routing and cost are potential problems.
- switching between the connected devices themselves is rarely possible.
- incremental cost for installation of new devices may be very high, especially in high-rise office buildings.

■ LOCAL DATA SWITCHES

Another form of the **star** structure is the local data switch or data PBX. In the configuration shown in **Figure 2a**, the host computer is replaced at the hub of the star by a switching device that serves the terminal population and multiple host computers. Switching interaction is outside the host, and therefore, it does not put a load on host facilities. The data switch provides terminal users with many of the same capabilities that the phone system gives to voice subscribers. Some data switches have capacities of thousands of lines and support speeds as high as 56K bps. Most operate with several hundred lines, however, and limit speed to 19.2K bps, not a serious restriction for most local data users.

In its simplest form, a local data switch operates as a combination port selector and port contention device. Each terminal has an individual line to the switch, and each CPU has a line for each available port. A terminal user requests connection to a specific CPU (port selection). If all of the ports for that CPU are in use (port contention), the terminal user can either try another system or elect to **camp on** the busy CPU and wait for a free port. Users can request connection to other users, subject to compatibility of the terminal equipment.

Data switches can provide **call accounting**, another feature often found in voice PBX systems. The switch is an element in each connection, thus it can record the destination and duration of each data call. This capability may be vital when the cost of the switching/connection facility is allocated according to its use. It can also provide valuable information on the loading of the computers connected to it, an aid in management planning for facilities expansion.

Some modern data switches consist of individual switching elements connected by some form of high-speed local communication path. These switching elements can be separated by several thousand feet, and each element can serve as a focal point for terminals in that area. Connections between two users served by the same **switching node** can be made without access to the high-speed path between nodes; thus a properly structured **node switch** can support users with shorter cable lengths and reduced cost without impacting data performance.

Another form of data switch is the **switching multiplexer**. These devices combine the data concentration services of a multiplexer

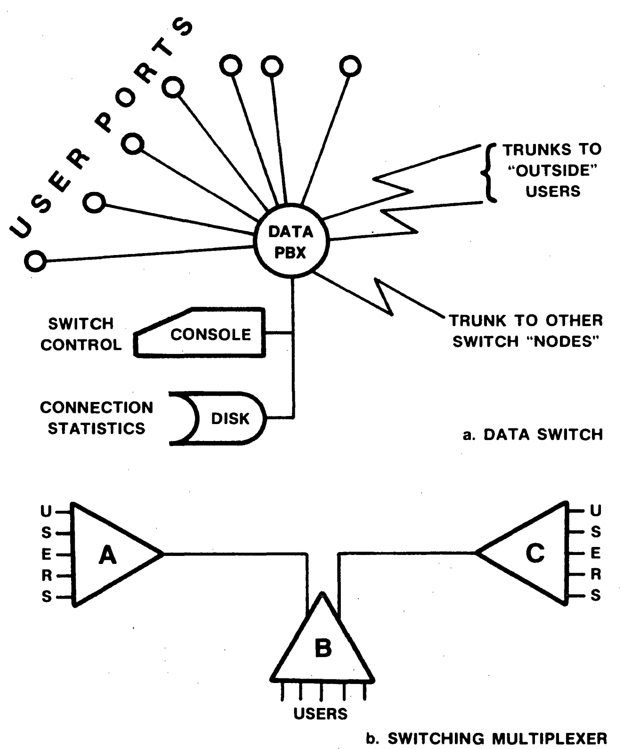


Figure 2 • data switches and multiplexers.

Local Area Networks (LANs)—A Solution to Resource Sharing & A PBX Adjunct or Alternative

with port contention and switching. As shown in **Figure 2b**, switching multiplexers look like nodal data PBXs, but the data links between nodes can be much longer with multiplexers because they use standard dedicated communication lines operating at up to 56K bps. Some switching multiplexers can be linked into multinodal networks with each node operating as a switching element, like a distributed data switch. On the other hand, multiplexers normally introduce more delay into the data exchange than data switches, and they cannot support as many lines per switch or handle as many simultaneous connections. The primary advantage of these systems is the distribution of switching functions over a wide geographical area such as nodes in New York, Chicago, and Los Angeles.

■ VOICE/DATA SWITCHES

Combination voice/data switches are a new solution to data switching or local-area communication problems, but an expensive one. A voice/data switch typically costs more than double that of data switches. Combination voice/data switches still have considerable advantages for some users.

Voice/data switches are sometimes called **digital PBX** or **third-generation** switches because they use a digital pulse-code technique for sending voice information, making the paths readily adaptable to data exchanges. The exact connection method varies from switch to switch, and the maximum number of simultaneous connections, called **non-blocking** maximum connections, the switch will support is not necessarily the same as the number of user ports supported. This may not restrict a user, however, since very few PBX systems will ever need to support as many connections as there are potential user pairs. A more serious restriction imposed by some switches limits the number of data calls or allows data calls to absorb the resources of multiple voice calls, reducing the voice capacity of the switch.

The primary advantage of the voice/data switch is the combination of voice and data itself. Some users find a single office position may originate a combination of voice and data calls throughout the workday to internal or external destinations. Using a single switch and single line to handle both types of calls may save considerable equipment and cable cost.

A second benefit of the voice/data switch is the high data rates it supports. Many such switches allow user communication at rates of 56K or 64K bps, as compared to 19.2K bps for most data-only switches. This high speed is not always available to the user, however. Common asynchronous terminals are not capable of operating at such speeds. The more sophisticated synchronous devices often use a protocol that makes switched connections impractical or even impossible. A host is unlikely to respond properly to a terminal that suddenly "appears" on a polled line. It is unlikely that the average communication user will be able to justify a voice/data switch on the basis of data rates.

The number of ports supported by a combined voice/data switch is another benefit. Some permit tens of thousands of user connections, as compared to about 5,000 maximum for data switches. Like data rate, this will not be a significant benefit to most users because only the largest corporations can expect a user population in this range.

The final potential benefit is integrated call accounting. Most voice/data switches have complete call accounting packages to provide excellent charge-back information for the liquidation of switch expenses and the allocation of out-of-office telephone charges. The combined switch can give a user a single charge sheet for both voice and data connections. Furthermore, voice/data switches often include a feature to select the most economical circuit for a given call, taking advantage of tie lines, WATS, or other special-rate service. This feature can reduce long-distance charges when the callers are not necessarily aware of the best method of transmission available at the time of the call.

Like data switches, some voice/data switches have a nodal structure. Each node can handle local switching requirements for its connected users and, in addition, connect users between nodes by routing calls over the nodal trunks. Trunk lines can be coaxial cable, parallel twisted-pair, or digital high-speed links such as T1 carrier (1.544M bps). Most systems restrict the node separation to a few thousand feet, but some support communication paths such as the AT&T T1 carrier for transcontinental separations. High-speed digital trunks have a limited value for voice connections or high-speed pulse-coded digital data because each connection requires 56K to 64K bps limiting a 1.544M-bps trunk to only about 30 calls.

■ LANS—AN ALTERNATIVE SOLUTION TO LOCAL DATA COMMUNICATION

In terms of user interest, cable technology has dominated the means of providing local-area communication services. Although some of the original popularity of the cable systems was undoubtedly due to the novelty of the concept and subsequent interest of the trade press, cable architectures provide real advantages.

Local networks, based on a switch or **star** architecture where terminals link to a hub, require a dedicated communication line for each terminal. Local exchange, or switching, is done by the hub switch connecting the paths of the user terminals involved. The line savings are limited to the dedicated lines between a given user terminal and all other user terminals with which it communicates in favor of a single link to the hub per terminal. Cable networks replace this hub structure with a shared cable to which all users connect. The shared cable can be a single line or **bus**, having two end-points with user terminals connected between them or a closed **ring or loop** path. The cable can be a coaxial or CATV cable, a twisted pair of wires, a parallel-connection cable or ribbon, or even fiber optics cable.

■ CLASSIFYING LAN ARCHITECTURES

Some of the alternative concepts of cable networks have already been mentioned, but a more complete list of classification points is useful in evaluating cable network design alternatives.

Topology—Networks can use a ring, bus, star, or some hybrid structure such as a center-connected ring. The bus and ring are the typical cable forms.

Transmission—Data can be sent directly on the cable (baseband) or used to modulate a radio or light-frequency carrier in a manner similar to a TV channel (broadband). Broadband networks can assign fixed frequency slots to communicating parties, use special modems called **frequency-agile modems** to permit users to switch slot frequencies to connect to other stations on the cable, or use frequency modulation to define channels shared by many users.

Control—Stations on the network can contend at random for use of the cable, use a form of arbitrated transfer such as station polling in the same manner as used in long-distance protocols like IBM bisynchronous, or pass "tokens" granting transmit permission from station to station.

Station Linking—Broadcast networks attach all stations to a single cable and data sent from one is received by all. Point-to-point or repeater networks form the ring or bus with a series of station-to-station connections. Each station in such a string receives only from the station "above" and sends only to the station "below."

Medium of Exchange—Cable networks can use transmission-style coaxial cable, CATV cable, twisted pair or sets of twisted pairs of wire, fiber optic cable, or nearly any other form of electrical or optical medium.

Speed—All cable systems have a maximum internal data rate that limits the total capacity of the network. Many also have a loading limit associated with the design restricting the actual operating range to a value less than this maximum.

Data Format—Most baseband networks rely on time-division use of the cable and assemble data into packets much like a public data network. This affects network performance if the user data is not already packetized in some way, or if user messages are very short. Ethernet, for example, requires that at least 46 bytes of data be included in each **packet**. This means that very short messages must be padded by the Media Access Unit (MAU) out to 46 bytes, a waste of cable transmission throughput. Broadband frequency-division networks, on the other hand, are normally (but not always) transparent to the data structure of the users in the same way as a data PBX.

Local Area Networks (LANs)—A Solution to Resource Sharing & A PBX Adjunct or Alternative

■ RINGS & BUSSES

Cable topology, or the way in which the network elements are connected, is a basic decision for cable networks; it almost dictates other design alternatives. Alternative topologies are the **ring**, where user devices tap onto a closed loop of cable, or the **bus**, where they attach to a single run with unconnected ends. **Figure 3** shows the ring and bus cable network structures.

Ring networks are actually constructed as a series of point-to-point paths formed into a loop. Each ring tap breaks the cable, receiving data from the tap "higher" on the cable and regenerating it for the "lower" tap. A bypass connection is needed on the ring if a station fails and is unable to regenerate the message. Without the bypass, any tap failure **breaks** the ring and makes communication between some stations impossible. Rings must close on themselves; thus they require more cable than a bus system for the same connection geography. All ring systems regenerate the signal for each user-to-user hop and can support greater station-to-station distances than the bus systems that use broadcasting. Because a station on a ring must send to only one other and receive from only one other station, the matching of new taps to the network is less complex and the tolerances in the receivers and transmitters are less critical. On the other hand, each new tap breaks the ring, introducing the chance to create a cable fault during the process. Primenet, a proprietary offering of Prime Computers, is a typical example of a ring network with bypass facilities to prevent a single station failure from breaking the ring. **IBM's loop networks** are examples of rings with a single controller.

Bus networks can be constructed of individual paths in the same way as the ring systems, but this form of bus requires a repeater at the "end" of the bus to send the data backup line to stations attached above the sender. This downline/upline path requires either a broadband system that uses 2 different frequencies for the 2 directions or a dual-cable system. Either arrangement is more expensive than broadcasting, thus the most common practice is to broadcast data onto the cable from a station; all other stations on the cable listen for their messages. The bus must be terminated at the end-points to prevent the data from **echoing** from the open end of the line back into the network. Broadcast taps do not break the cable each time a tap is added. A simple tool can be used to attach a new cable interface unit without interrupting the cable. Some forms of continuity and quality testing can be conducted on this continuous cable to verify its operation without involving the interface units themselves. A failure of a single unit **cannot** cause the entire cable to fail because the interface unit serves only its own devices.

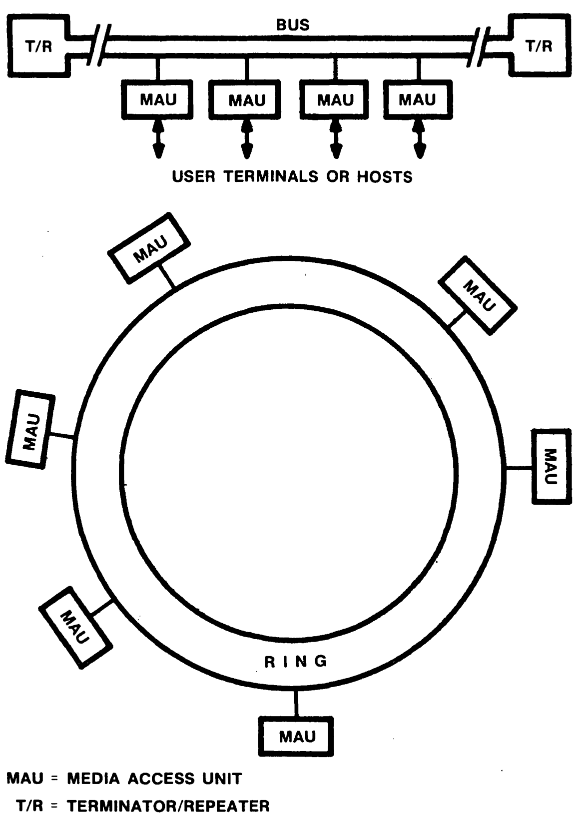


Figure 3 • ring and bus cable networks.

A special problem in bus networks that use broadcasting is the **propagation delay** of the message. Even though the data moves through a cable at the speed of an electrical impulse, it takes a measurable time to move from one end of a bus to the other, on the order of four to five microseconds. This interval is long enough to move a complete message if cable data rates are high enough, in the million-bits-per-second range. The effect of the delay depends on the method used to control access to the cable, and its impact on **Carrier Sense Multiple Access (CSMA)** strategies is covered in the following section.

There are many variations on the basic ring, star, and bus structures. Daisy-chaining alternate stations together with a secondary link can be used on ring networks to eliminate the problems associated with loss of ring continuity due to station failure. All of these structures, however, are special-purpose networks for use under conditions that justify almost customized equipment, software, and installation. They are unlikely to be offered commercially at a price that can be justified by most user needs.

■ BROADBAND & BASEBAND LANS

Most early cable networks were designed to send the data directly on the connecting medium in a digital form, called **baseband transmission**. Baseband systems can be compared with AT&T Communication's digital service, in that they do not require modems for the communicating parties. Although even baseband systems use a form of **modulation**, or a technique to convert the digital data bits into cable electrical impulses, baseband systems do not impress the data signal on a radio-frequency carrier. The interest in baseband concepts in the early networks grew out of the desire to save modem costs, at that time considerable. The cable television industry changed the economic picture for broadband networks by making devices for the handling of radio-frequency carriers both widely available and inexpensive. Broadband systems based on CATV components compete favorably with the baseband systems in cost, and they offer other significant advantages. **Figure 4** shows a typical example of frequency usage in broadband and baseband cable systems.

Baseband systems use special modulation techniques to convert digital bits into electrical signals on the cable, a concept much like the coding of a television picture into impulses as done by the TV camera. The signal can be converted back to its digital form at the other end, but two stations converting data to this baseband form create signals that interfere with each other if sent at the same time. Baseband systems, therefore, must provide means for passing control of the cable to a station so that data can be sent without the interference of other stations. Several schemes have been used, beginning with simple time-division multiplexing or **slot** concepts such as the **Cambridge ring**. Most modern baseband networks use either a contention system, **Carrier Sense Multiple Access (CSMA)**, or a rotating-control method called **token passing**. CSMA networks rely on listening strategies to avoid collisions of data, and token networks circulate a **permission to send token** that a station captures to use the cable and regenerates when the message is complete.

In broadband systems, the digital signal is modulated onto a carrier frequency in the same way as a video signal is modulated onto a single TV channel frequency. The resulting combination of data and carrier can be received by another station by **tuning** to the carrier frequency. Broadband systems can use modems that can change frequencies at will to pair sender and receiver on an

Local Area Networks (LANs)—A Solution to Resource Sharing & A PBX Adjunct or Alternative

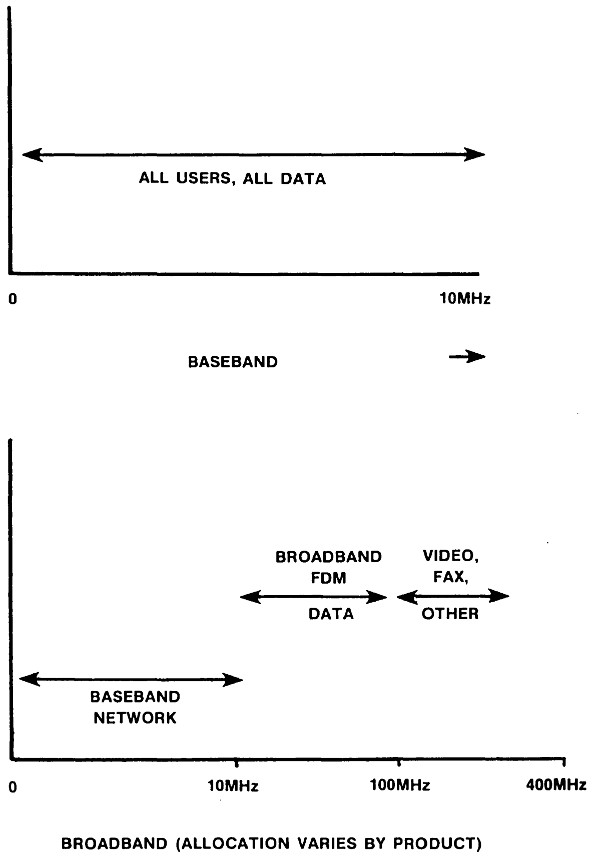


Figure 4 • broadband/baseband frequency uses.

unused **channel** for communication. Because the sender and receiver use a special frequency, there is no need for time-division strategies such as CSMA. Some broadband networks use frequency division exclusively, while others such as WANGNET divide the cable into **bands**; some bands use fixed-frequency pairings, some use frequency-agile modems for switched frequency-division multiplex communication, and some use CSMA techniques to share a frequency for low-volume exchanges across large user populations.

Broadband systems not only allow users to separate from one another by using different frequencies, but allow a network to share its cable with television, facsimile, or other uses so long as the "channel" used for data transmission is not shared (see **Figure 4**). Media Access Units (MAUs) for broadband systems can be designed to operate at a fixed frequency within the broadband network's allowed range or to move about within that range. Fixed-frequency modems are less expensive but require that the sending and receiving stations have modems of the **same frequency**, essentially establishing a non-switched environment. Frequency-agile modems can **switch** frequencies to connect between any two parties on whatever portion of the bandwidth that is available. Frequency-agile modems are more expensive than the fixed-frequency type, and the cost difference is one of the reasons that quoted costs for broadband access taps vary so widely. Frequency-agile systems have the potential for incredible amounts of traffic movement. A single channel slot on a CATV cable provides 6 million cycles per second of bandwidth, enough for about a hundred channels at the highest data rate currently in common use, and all at the same time. Switching time for frequency agile modems, however, is slow, in the milliseconds to seconds range, thus frequency switching will **not be practical** when high data rates are required.

The potential traffic handling of broadband systems is considerably greater than baseband, and the broadband concept has greater growth potential than baseband.

■ CONTENTION VERSUS TOKEN PASSING—WHO CONTROLS THE FLOW?

Like the conventional long-haul networks, LANs need a way to control user access to the shared facilities. Several hundred technical papers have been written describing alternative strategies for access control, but the process of commercial selection seems to have produced 2 principal alternatives, **contention and token passing**.

Early cable systems used time division multiplexing (TDM) to separate users and prevent collisions. Time slots could be visualized as "passing around" a ring network and being filled with data by the appropriate station. Such systems have much the same disadvantage as time division multiplexers; a station has time allocated whether it is needed or not, and the capacity of a network is limited by the number of slots available regardless of how much they are used.

Token passing (**Figure 5**) is a scheme similar to the circulating time slot. A station with data to send must first acquire a **token** that grants permission to transmit. Having done so, it removes the token and substitutes its data, regenerating the token at the end of its message. The token is produced by a station when the network is initialized and continues to circulate as long as the network operates. Each user regenerates the token in some way or another, and the network should, in theory, operate forever with the single token circulating. Malfunctions in the stations on the network or in the cable medium can cause the token to be damaged or destroyed. This condition causes the network to stop traffic because no station acquires permission to use the cable. Token networks have strategies to detect the loss of the token and force a reinitialization of the network to regenerate it.

Contention systems offer a simple alternative to token passing as a control mechanism; it is popularly called **carrier sense multiple access (CSMA)** (**Figure 6**). CSMA was popularized by its use in the University of Hawaii "ALOHA" packet radio network. In this type of access control, a station simply sends data when it detects that the line is free. People who have tried this strategy at a crowded dinner party have already deduced its deficiencies—as the number of stations increase the chance of getting a message across reduces to zero because another station jumps in at the same moment and the data collide. The problem is increased by the propagation delay mentioned earlier. Both ends of the cable are some millionths of a second apart, and a collision can occur simply because one station's message has not reached the other end of the cable. Message collision is therefore **inevitable** and must be dealt with in CSMA strategies.

One way to deal with it is to ignore it on the assumption that control over the number of users and messages can reduce the probability to an acceptable level. A corollary to this strategy is that stations involved in the collision can try their messages again later. The **backoff** mechanisms vary. Some systems have stations wait a random time interval before trying again. The randomness prevents synchronization of the competing attempts to send, resulting in deadlocking the network. Another method couples random number generation with timing measurements based on the time interval required for a message to travel the entire length of the cable. This system, called **persistance**, requires the waiting station to "roll the dice" each time the propagation interval passes during an idle channel period. If the correct "number" or "persistance" number comes up, the station sends its message. Otherwise, the station waits another cable delay interval and tries again. Networks that allow collision and depend on backoff mechanisms only are called **listen-before-talking networks**.

An obvious problem with the "listen-before-talking" system is that time is wasted sending a message into an already-existing collision. That time can be partly recaptured by having the stations listen for another message while sending their own. Actually, the sender must listen only for a time interval at the start of the transmission equal to the propagation delay of the cable because collisions can only occur during the time it takes for the message to reach the furthest station. After this, the normal "listen first" strategy prevents collisions. "Listen-while-talking" strategies

Local Area Networks (LANs)—A Solution to Resource Sharing & A PBX Adjunct or Alternative

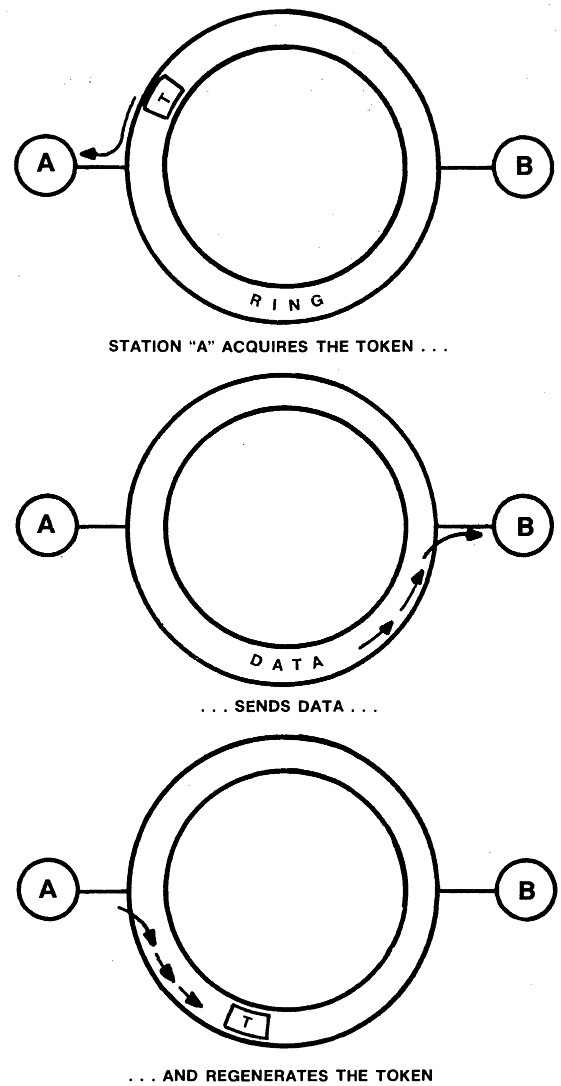


Figure 5 • an example of token passing.

allow a network to detect another station's colliding message, so these networks are called **CSMA/CD** to indicate the addition of **collision detection (CD)**. The CD systems are more complex than CSMA systems. The hardware required to detect another station's message under the station's own transmission is not trivial, particularly when all CSMA systems are broadcast systems and must also contend with differences in signal levels from nearby and distant stations on the cable.

The propagation delay of the cable also enforces a minimum message size on CSMA networks. If the time it takes to send a message is less than the cable delay, there is a possibility that the message will be completed before the sender "hears" a collision, making the error undetectable at the cable level. CSMA/CD networks normally require a **minimum message size**—longer than the delay interval on the cable. If this minimum is not imposed, a higher-level message acknowledgement protocol within the network must check for errors due to undetected collisions and to request retransmission.

At low levels of data traffic, CSMA and token systems have little difference in performance. When data traffic increases, the CSMA systems experience more collisions and less free cable time, and they **degrade in performance**. Token systems wait longer for tokens and performance degrades as well. The difference in the 2 systems is that the time a station must wait to capture the circulating token can be calculated based on the maximum message length and the number of users on the cable. This precision measurement of maximum delay has led specialists to describe token systems as **deterministic**, meaning that performance can be determined mathematically. CSMA systems, on the other hand, are **probabilistic**. This means the average performance of the network can be predicted, but the specific performance at any point in time can be stated only in terms of probabilities. The deterministic/probabilistic difference has caused concern among users who are reluctant to rely on a network scheme with a chance (to be sure, very, very small) that a station might wait hours before it could gain access to the cable. In fact, the risk of long access delays in CSMA networks is quite small unless the cable utilization is **much higher** than the recommended levels—the risk little more than the probability of a lost token or a failure of the cable itself. Most users find the performance of either network acceptable.

Network Systems, with its HYPERchannel and HYPERbus, uses a CSMA/CA (Collision Avoidance) scheme which combines features of token passing with CSMA to prevent collisions between 2 nodes that listen then start transmitting simultaneously. Once the network becomes idle, Network Systems allots a time slot to each active node on the bus to begin transmitting. With this scheme, the first/next node on the bus with data to send can capture the bus when its time slot rolls around. This arrangement is comparable to token passing for bus access, and it provides for deterministic bus performance.

Ethernet is the dominant form of CSMA network. It utilizes collision detection (CD), imposes a 46-byte minimum message size, and implements a binary exponential backoff strategy—each

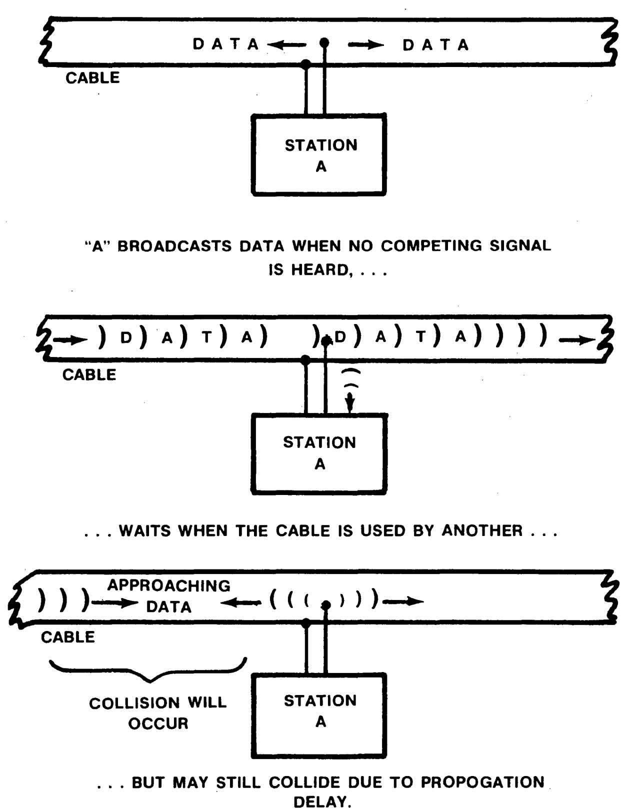


Figure 6 • CSMA.

Local Area Networks (LANs)—A Solution to Resource Sharing & A PBX Adjunct or Alternative

time a collision is detected the backoff time is doubled. Performance testing and mathematical modeling have shown this combination is **efficient** for normal medium-scale cable networks. IBM's loop products are the most common form of token ring network.

■ CONNECTING THE NETWORK TO THE USER

The network potential of hundreds or thousands of possible connections to the cable is not always easily available to the user. Cable networks, like public data networks, provide a set of services to their user community. This service set, not the potential of the architecture, must be the basis for selecting a network vendor. The Media Access Unit (MAU) structure shown in **Figure 7** represents the hardware attachment to the cable; it does not show the logical features provided by the software needed for cable use.

Some communication environments are made up of a large number of fixed user pairings, such as those between a timesharing terminal and a host. Cable networks in this mode primarily offer installation cost and cable routing advantages; switching or full connectivity is **not** a user requirement. For networks of this type, the user wants the connection map set up once and regenerated each time the network is restarted. It is undesirable to make each user set up the normal connection. Such **fixed circuits** are not inherent to any particular cable network type and must be supported by the vendor in some way. In CSMA networks, fixed-connection support consists of loading each MAU with the address of each of its users' partners. In fixed-frequency-pairing broadband systems, it consists of selecting modem frequencies properly so that paired parties operate on the same channel. Some networks downline load each fixed-connection MAU with the connection parameters at initialization time.

Where fixed pairings are not desired, the user must determine whether the exchanges are long (**circuit-level support**) or very short (**datagram support**). In either type of connection, the user must identify the called party to the network. This requires a service protocol similar to that required for a public network or a dial telephone. If a user wants to connect to a fixed destination on becoming active, some networks allow contention for that destination's addresses and report busy or call-waiting if all the addresses are in use. Where true switching control is required, a terminal or host may use a service protocol to request connections.

Even if switching on user command is available, different levels of support can be expected. Some vendors allow users to connect only through the cable's own address scheme—normally a numeric code and not easily associated with either specific users or application programs. Other vendors provide a logical address system that allows users to know others by more descriptive names such as "IBM4" or "PAYRLCLK."

Another issue is protocol conversion. Networks normally perform speed matching (connecting a 300-bps terminal to a 9600-bps host) but **rarely** convert asynchronous-to-synchronous communication or ASCII to EBCDIC code.

These "extra" functions in a cable network are almost always supported by microprocessor intelligence in the MAU and often by having a **control node** to maintain user lists, assign addresses, and load station parameters. Many users find these services necessary and select such a network, and yet do not consider that in doing so some of the disadvantages of cental data switches have been transported into a cable environment. As the level of expected network services increases, the cable networks begin to operate like packet switch networks and must be evaluated in those terms. Such networks are likely to be expensive and can be inferior in service, performance, and reliability to large data switches.

■ STANDARDS IN LOCAL NETWORKS—HOW MUCH, HOW SOON?

Cable technology has exploded in the last several years, resulting in several dozen vendors competing for what is expected to be over 1 billion dollars by 1990. It is likely that many of the offerings of today will not survive in their present form to the end of the decade, thus **users must be wary** of customized implementations and proprietary designs. The alternative is to adopt a standard—a definition of local network operation and equipment to do for LANs what X.25 did for public packet networks.

International standards do not happen overnight. In fact, it is possible that full international agreement on an LAN standard will take several years. The user can hardly wait that long. In fact, a wait can even impact the market for the LAN product. Everyone in the LAN marketplace may find an alternative solution. Committee 802 of the IEEE was formed in 1980 to standardize LANs. It immediately became a battleground of ring versus bus, broadband versus baseband, and CSMA versus token passing. The committee decided to address LANs with respect to the OSI

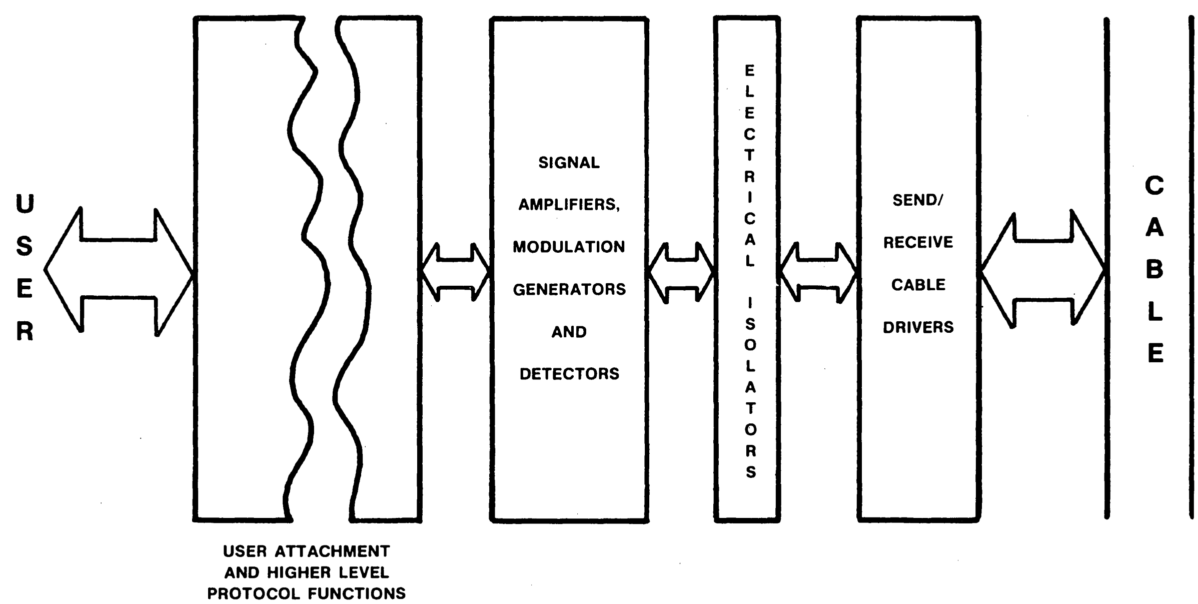


Figure 7 • a media access unit (MAU).

Local Area Networks (LANs)—A Solution to Resource Sharing & A PBX Adjunct or Alternative

Reference model—limiting committee efforts to the physical layer (level 1) and the data link layer (level 2); leaving higher level standards to normal standards organizations. **Figure 8** shows the OSI model compared to the ETHERNET model and 3Com's model of higher level protocols for an ETHERNET-type network.

The 802 standards committee was unable to define a single architecture for cable networks, for which it has been unjustly criticized. There is no more hope of a single standard for local networks than there is for a single standard for all data communication. Without trying to impose a single standard and render the effort valueless, 802 has proposed standards for the principle variant architectures. This recognizes the fact that no single cablé concept will be **ideal in all** communication environments. The current form of the draft offers the following alternatives:

- datagram or circuit service.
- CSMA/CD control using baseband or broadband technology.
- token ring or token bus topology.

The committee has adopted a CSMA/CD bus standard and proposed a token bus standard, a token ring standard, and a metropolitan area network access method. The CSMA/CD standard differed in one particular from the ETHERNET description initially but the ETHERNET specification has been changed to agree with the 802.3 committee standard.

Other organizations are interested in LANs. The European Computer Manufacturers Association (ECMA) provided input to the 802 committee, and some changes were made in the March 1982 802 draft as a result of the ECMA suggestions. ANSI and ISO appear ready to include the 802 standard in any future LAN standard it produces. The ISO, through its committee on the OSI reference model, is working on incorporating "**Connectionless mode (LANs)** communication in the OSI model. This group is working with the IEEE 802 committee to fit LANs into the model.

■ USER SELECTION STRATEGIES—APPLYING CAPABILITIES TO NEEDS

No network selection task can be undertaken without a clear set of user requirements. This is particularly important with LAN selection; such networks almost always carry key information traffic for a business and disruption of the LAN is certain to be a disaster. The first step in selecting a local network is to study the traffic the network will carry and answer the following questions:

- How many users will be connected? Count both the number of terminal devices and the number of host computer ports. Also project reasonable growth over the next 3 to 5 years.
- What data rates will be supported? If terminals and computers are running at less than optimum speed because of current network restrictions, don't assume this will continue.
- What is the average number of characters exchanged per hour, and what is the peak volume? Add both terminal input and host response. This is especially important on CSMA networks, so if they are being considered, spend extra time here to get it right.
- What type of connections must be supported? Most users require some permanent connections for host-to-printer paths. Switched connections where the host computer "dials" the number may require software changes in the host, and terminal users must be able to understand and complete any terminal switching procedure. Also, large terminal populations in a freely switched environment almost force logical address schemes. Otherwise, all users will require "phone books" and spend excessive time looking up numbers or recovering from bad connections.
- Are there any incompatible connections possible among the connected users? If both asynchronous and synchronous terminals can be connected to the same network, a misdial can cause a connection between incompatible devices. This might cause a **hung port** or force manual recovery procedures, certain to be time-consuming and resented by the operating personnel. Switched systems with incompatible devices attached should have network protection against illegal connections.
- Are there any security problems with any of the users on the network? Providing free user access to the payroll system will be appreciated by some, but not likely by management.
- What are the local building codes in the area of cable routing and use? Baseband systems sometimes cause a problem with grounding because the cable is continuous. This is particularly true if the network spans several buildings. In all cases, some buildings are difficult to route new cables through, and some local codes will not allow certain types of cables. Checking installation costs is also recommended; a New York cable user found that routing a cable from one floor to another in a high-rise office building costs more than a small computer system.
- Are there shared uses of the cable? CATV or other video applications can share a broadband cable, saving considerable installation cost over separate facilities.

Once the network needs are determined, they can be matched against the capabilities of the vendors. Selecting an architecture before looking at vendors can cut down on the number to be evaluated, but should be done only if a need clearly requires it. For example, a requirement to share the cable with surveillance TV is a legitimate reason to require broadband systems, but the assertion that the "need for datagram support requires a baseband CSMA system" is a prejudgement to be avoided.

In reviewing vendors, look for a reasonable number of installed systems, but do not expect to find a 5-year user base—the architecture is not old enough. Be wary of very small firms with no capital base, no user base, and a proprietary technology. If possible, select a vendor that either implements a standard concept or shares an architecture with other vendors.

Finally, users with no data communication experience should carefully study the use of the network concept of each vendor considered from an operations point of view. Any network will necessarily have some performance and operational differences from a direct-connect environment. Users with no other communication experience may need to visit another installation to visualize the effects of the new system. Do so if you have any doubts.

■ CABLE & SWITCH STRATEGIES COMPARED—A LAST LOOK

There are many more switch-based networks installed than cable networks, and even the optimistic estimates of cable growth are

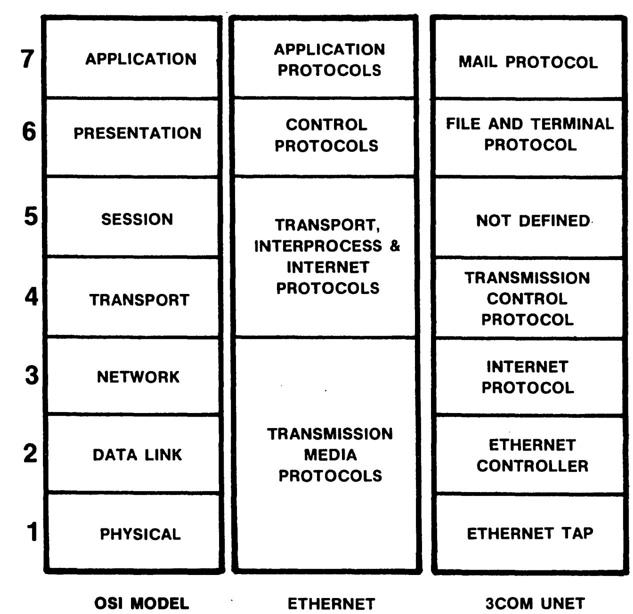


Figure 8 • protocol layering in LANs.

Local Area Networks (LANs)—A Solution to Resource Sharing & A PBX Adjunct or Alternative

unlikely to change this pattern for 3 to 5 years. Potential cable users must evaluate whether cable itself is justified in their application before they select a specific cable architecture.

The star-structured switch networks have the following advantages over cable systems:

• Very large user populations can be supported without loss of performance.

• Cost of the switch itself is normally lower than cable alternatives for networks with a large number of ports.

• Integration of a switch into an existing directly-cabled environment is easier than is cable.

• Central control of the switch tasks allows implementation of password and other security mechanisms.

• Central accounting/billing functions are easily added and normally available from the vendor as an option.

• Network features such as logical user addresses, password protection, and closed user groups are more likely to be available.

The primary disadvantage of the switch networks is the reliance on a single central element for exchange of all user data. Failure of the switch causes all information flow to cease unless **redundant facilities** are provided at additional expense. A second disadvantage is the cost of cabling each device to the switch, a significant factor if the users are widely separated or if building codes and structures require unusual installation practices.

Cable networks have the following advantages:

• The cost of a small network is much lower than with data switch networks.

• Cabling cost is almost certain to be lower than with data switches except where all users are located in the same area.

• Data transfer rates supported by cable are much higher than those available with all but the most expensive switches.

• Many cable systems are designed so that the failure of a component other than the cable itself will affect only a small number of users.

• Some cable networks permit sharing of the cable with CATV, facsimile, surveillence video, or even digital voice applications.

The primary disadvantage of cable systems was the **lack of standards**, but this disadvantage is slowly evaporating. Users can expect to purchase all their media interface units from one or a small group of suppliers; some concepts and features that become available after installation of a network may not be available for retrofitting to an existing network. Further, the **lack of central billing and accounting information** may make cable systems unattractive to users who bill for the communication service directly or through internal accounting procedures. Some sophisticated LAN systems such as Sytek's baseband/broadband Local Net have an optional network control center that can provide such services.

Economic issues are often paramount in local network selection, because the prime goal of such networks is usually cost reduction. The cost of a network can be divided into **per-port costs and shared facility costs**. For data switches, the per-port cost is normally low because the interface to the shared facility is designed to be simple, but the cost of the shared facility (the actual switch) is high. In contrast, the cable network port costs are often higher than data switch port costs, but the central costs (the cable) are quite low.

In general, switch networks are preferred when the primary task of the network is to permit **free connection** among a relatively **large population** of devices located within several thousand yards of one another and where the devices are operating at normal communication speeds of 19.2K bps or less. Cable networks are preferred where the primary goal of the network is to **reduce cable cost**, the data rates are expected to be **above 19.2K bps**, or the same transmission medium can be used for other purposes.

■ TRENDS IN LOCAL COMMUNICATION

Over half of the data exchanged travels less than 500 yards. The market for data switches as local network devices is not easily separated from other uses of such devices, but most industry experts feel that the market for local networks could reach $2 billion by 1990. Cable networks appear to have captured less than half of this market at this time, and are not expected to improve their share dramatically.

One major problem with cable systems is their lack of maturity. The physical mechanisms for exchanging data on a cable are still hotly debated, and almost no progress has been made in defining higher level protocols for cable systems except at a vendor level. High-level protocols are as vital for the effective use of cable networks as they are for the use of public packet networks. Lack of higher-level protocols will continue to force users with a need for network features into the PBX camp. This issue will disappear once LAN standards are fitted into the OSI model.

In the long term, most experts agree that broadband systems offer higher data movement potential and greater flexibility. Projections of the data capacity of baseband cable over the next 10 years indicate 15% to 20% improvements are possible. Broadband cable capacity will improve nearly 70% and fiber optic cable can more than double its potential capacity. High-performance systems such as Network Systems' HyperChannel have already been proven effective as an extension of the computer bus, and reduction in the cost of such systems will make them a reasonable base for connecting the multiple computers, devices, and storage media needed for automated office/distributed processing and other applications seen as critical for business growth through the 1980s.

• END

Protocol Converters—Bridge To The Future

An Introduction to Protocol Conversion as a Solution to Communication Compatibility

■ INTRODUCTION

From the early days of data communication, users have faced the problem of rapidly changing technology. The 1980s have seen a dramatic acceleration in this rate of change, coupled with an economic climate that requires careful analysis of capital investment. Many companies with older terminals have hesitated to move to newer network architectures because of the high cost of replacing their terminal populations. A significant barrier in the use of older devices with new networks is incompatibility between communication protocols, an issue with little relevance to the user's application but a potentially insurmountable technical barrier.

Protocol conversion is an obvious solution to this problem. Many vendors supply equipment that allows a user to migrate to new network or transmission systems without replacing the terminal base. But these devices have restrictions which users must evaluate before committing their network expansion policies to protocol conversion. This article explores the techniques, advantages, faults, and applications of protocol converters.

■ PROTOCOLS—WHAT & WHY

Users of data communication equipment are seeking a way to exchange information, and the process is similar in many ways to a human conversation. For example, parties in the conversation must agree to a set of rule or limits in language and flow of speech to assure some level of mutual understanding. The collection of these rules is called a **protocol**.

The complexity of a protocol is related to the complexity of the communication environment. Simple applications of text exchange between human operators of terminals can function with a very simple protocol because human judgement can be applied at both the sender and the receiver. More complex protocols must be used where one of the users is a computer or where the speed and volume of information exchanged makes human, character-by-character review impractical. This dependence of protocols on application has resulted in an evolution toward more complex protocols as the importance of the computer in business has increased and the number of intelligent workstations has grown.

Manufacturers of computers have typically viewed protocols in much the same way as other computers or products, and have introduced their "own model" rather than adopt that of a competitor. This has the same effect on communication users as the multiplicity of host computers has had on data processing users; problems with compatibility when equipment must be upgraded for any reason. Although international standards organizations such as the CCITT have presented protocols that are nearly immune from arbitrary variations due to vendor incompatabilities, such protocols are relatively new and not yet widely used. The majority of terminals in operation today are using a protocol for which no industry-wide standard exists. Because of this, many terminals that perform the same functions cannot be used together because they do not employ the same protocols, and many new network architectures cannot be fully utilized on the older models of terminal.

What is in a protocol? Generally, a protocol contains agreement in three major areas:

• The method in which data is to be represented or encoded, called the **code set**. Most data communication today uses either the American Standard Code for Information Interchange, or **ASCII**, or IBM's Extended Binary Coded Decimal Interchange Code, **EBCDIC**.

• The method in which the codes are transmitted and received; by sending them as available at essentially random intervals and with start/stop markers (called "asynchronous") or by sending them as a continuous string of codes that must be broken into characters by counting the number of pulses that make up a character (called "synchronous").

• The **non-data exchanges** of information by which the two users control the conversation, detect failures or errors, and initiate corrective action such as retransmission. These exchanges are normally given acronyms by the developers, and readers of protocol descriptions find them sprinkled with such terms as "ACK," "SABM," "BIND," or "POLL." While these sequences are not user data, they establish the context in which data can be exchanged.

Protocols vary in structure according to their use. The most simple text applications use asynchronous data, the ASCII code set for character representation, and rely on the operators to detect and correct problems. Point-to-point or multipoint lines connecting high-speed terminals to a computer typically use a synchronous protocol with either the ASCII or a vendor-defined code set, and a proprietary set of control procedures. IBM's Binary Synchronous protocol, called **bisync** for short, is the most common example of this type of protocol, but other forms are found with the CDC UT-200, Honeywell GRTS/VIP, and Sperry UNISCOPE. Finally, there are protocols designed to control networks of users or to attach users to such networks. Examples of protocols of this type are IBM's System Network Architecture (**SNA**), Digital Equipment's **DDCMP**, or Honeywell's Distributed System Architecture (**DSA**), and the international standard **CCITT X.25**.

Often grouped with communication protocols is another class of agreements more properly called **device control protocols**. Most users of minicomputers who have purchased asynchronous terminals from several vendors have discovered that even though the code set, speed, and transmission method may be the same, communication with different terminals from the same computer port may not be possible. This is because the terminal has a set of **commands** or sequences of special characters that it recognizes and uses to perform functions such as cursor positioning and screen editing. Terminals of different manufacturers do not typically execute the same commands.

Taken in combination, communication and device protocol compatability problems are sufficient to keep most users to a relatively restricted portion of the universe of available terminals, and many would like to stay there so long as business growth does not force change. Unfortunately, there is increasing pressure for change, and dramatic change.

■ CONVERTING PROTOCOLS—THE NEED

User networks are affected not only by the requirements of their business but by the current set of technology/cost tradeoffs. When the cost of transmission facilities changes, business must evaluate the possibility of changing equipment if such a change would allow use of another transmission medium which is now more economical. When a new type of computer with more powerful support for business applications such as MIS, the advantages of moving to it must be weighed against the cost of replacing communication facilities that may be incompatible with the new system. Users must also monitor the costs of maintaining old equipment, and be prepared to move to newer models if the cost of maintenance is too high or the uptime obtainable is unacceptable.

The 1980s have seen a rapid rise in the cost of conventional communication services and an increasing number of alternative transmission methods, most notably satellite communication and public packet networks. Users of the older style of terminals, particularly of IBM bisync devices, have found that the long

Protocol Converters—Bridge To The Future

propagation delays of satellite links or packet networks reduces the effective speed of the lines because of the method of error control used by bisync, called "block acknowledgement." Figure 1 shows the way block acknowledgement works. When a user sends a block of data, the data characters are processed through a mathematical formula to calculate a **check sequence** of one or more characters. The receiver of a block, applying the same formula to the same data, should arrive at the same check sequence value. User "A," sending the block to User "B," includes the transmitter-generated sequence which User "B," compares to the one generated at the receiver. Finding no error, "B" sends a short acknowledgement message called an "ACK" to "A," giving "A" permission to send the next block rather than repeat the one just sent. Where the delay in moving the data from "A" to "B" and the ACK back to "A" is small, the effective rate of data exchange is limited by the speed of the line, often 9600 bits per second or even more. But where a satellite path or packet network introduces a delay in the path, the effect is that of lowering the line speed, because "A" can send a second block only after "B" has received and acknowledged the first, and the ACK has reached "A." If the delay of the path is a half-second, "A" will delay one second between blocks of data. More modern protocols, such as X.25 or SNA, can transit satellite paths or networks without such losses of line capacity because they have the capability of sending multiple messages without acknowledgement, but users of bisync must either avoid such circuits or convert to another protocol. In many of the areas of the United States satellite links may be chosen randomly for dial-up connections, so bisync users must accept the delays, change equipment, or change their protocol.

Application program considerations may also dictate communication changes. As companies move toward a unified MIS environment, they must cut the ties between specific terminals and specific applications, treating their networks as resources for the support of the entire company. Network architectures such as IBM SNA or DECnet have the capacity to share communication devices among a series of application programs, but older asynchronous or bisync devices may have severe restrictions in interacting with the newer network architectures. Users with multivendor computer centers may find that sharing a single terminal between two host computers is expensive or impossible because of protocol differences. Few asynchronous terminals are used in an SNA environment, for example, and fewer SNA terminals are used with minicomputer hosts made by DEC, Data General, Prime, or other vendors.

A final, but crucial, issue in protocol conversion is cost. The costs of changing an entire terminal base to adapt to a new transmission technique have already been mentioned, but there is also a cost of extending existing facilities in an environment where the supporting protocol is already known to be obsolete. Should the operator of a large network of IBM 3270 bisync terminals add additional clusters of 3270s when plans to upgrade to SNA are already being discussed? Or should a minicomputer user who owns a hundred or more asynchronous terminals purchase more when the application is already a target for migration to a new host?

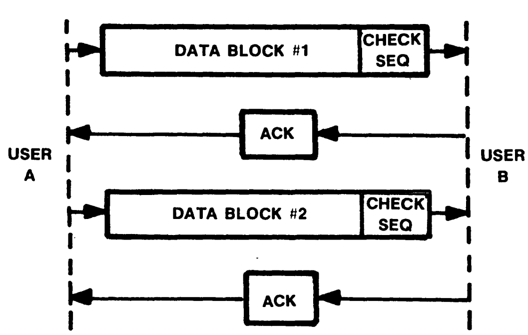


Figure 1 • block acknowledgement.

Protocol conversion, or protocol emulation, is an obvious technical answer to these problems. If a device could connect between users and somehow make their communication and device control protocols compatible, many of the existing devices could be utilized in newer networks and over modern transmission facilities without loss of capacity. Users could buy converters for simple terminals that would allow them to emulate more costly devices, and change the converters when the networks or applications required it without changing the terminals as well. Existing protocols would be converted to more efficient forms for satellite communication. If fully realized, protocol conversion could change the way in which communication devices are purchased and communication networks are designed, but it is safe to say that protocol converters have not yet reached this state. The task of converting one protocol to another is very complex, sometimes impossible, and always beset with seemingly small restrictions that must be examined in detail for their effect on user communication and use of communication devices.

■ CONVERTING PROTOCOLS—THE MECHANICS

The earlier comparison of data communication and human conversation is useful in understanding the structure of protocol converters. If two people speaking different languages wish to communicate, they may employ a translator. The translator talks to each in the correct language and internally "repackages" the ideas for presentation in correct form to the other party. A protocol converter performs a similar task, sitting on the communication path between two users and simulating another user of compatible protocol to each. As **Figure 2** shows, this gives protocol converters a distinctive "double-ended" structure. At each end of the converter is a local protocol handler that acts as a communicating party in the protocol required by the circuit attached. Connecting these handlers is a **gateway** task which provides for the movement of user data between the handlers. If all communication protocols were structured in accordance with the OSI Reference Model, the converter would be a set of seven-level OSI protocols joined by a common "user," the gateway task. Because the central task of a fully structured OSI protocol is the isolation of the user from the communication environment, a protocol converter dealing exclusively with full OSI-model protocols would be fairly simple to develop and would operate with few restrictions. With non-OSI protocols, such as those commonly in use in today's networks, the task of conversion may be complicated by the following issues:

• The format of the user data itself. If the data is easily separated from communication and device control protocols, it is more easily transferred to another environment. Use of special features such as data compression will normally complicate protocol conversion because such facilities do not necessarily exist in the other protocol.

• The degree of layering in the protocols. Even though full correspondence with the OSI model is unlikely in protocols being considered for conversion today, any amount of OSI layer structure in the protocols will aid in the separation of useful data from control information which must not be introduced into the other environment.

• The availability of common functions in the protocols involved. Data exchange between the users requires a degree of synchronization between the two foreign protocols. For example, most older protocols operate **half-duplex** in that only one station at a time may send. It is necessary for converters operating between half-duplex protocols to insure that both stations are not given "permission" to send at the same moment, since neither could receive under those circumstances.

Where a protocol converter is used to allow a terminal of one type to simulate the operation of another type device, some form of device control protocol translation may be needed. IBM's popular 3270 series of terminals is often emulated using lower-cost asynchronous devices, but the 3270 has special features such as the ability to return only modified fields to the host computer. This ability must be emulated within the protocol converter, making converters of this type look almost like a small computer system. **Figure 3** shows the structure of a terminal emulator protocol converter.

Protocol Converters—Bridge To The Future

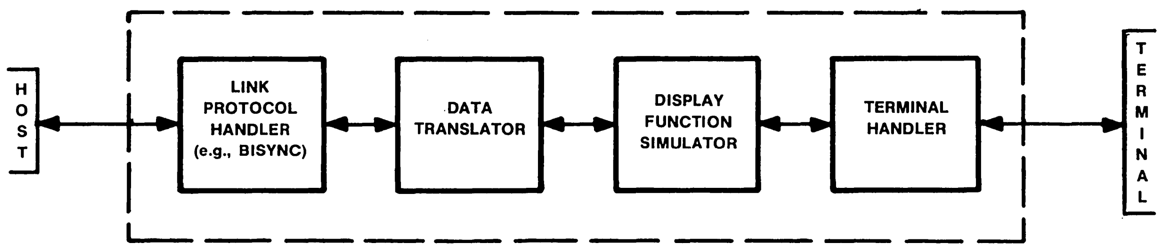


Figure 2 • protocol converter structure.

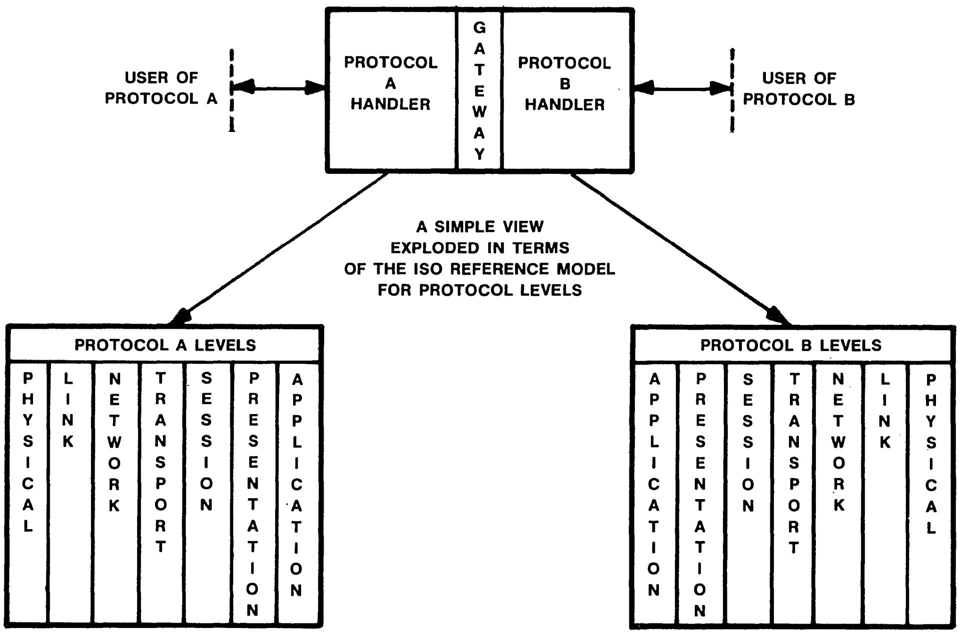


Figure 3 • inside a terminal emulator.

■ KEY TASKS IN PROTOCOL CONVERTER OPERATION

The operation of protocol conversion is best understood by dividing the functions into the following three key tasks:

• **Data Isolation.** This is the task of extracting the actual user data from the protocol-oriented structure of one user and correctly presenting it in the structure of the other. Data isolation must also handle conversion of the code set (from ASCII to EBCDIC, for example) and any changes in the form of the data itself, such as the introduction or elimination of data compression.

• **Control.** The protocol converter must act as a compatible user in both of the protocols involved. In some cases, the converter may emulate a terminal, while in others it must appear as a host computer. The correct control information must be delivered to the user at all times, and data must be introduced only in the proper context of control exchanges. Where a protocol requires the start and end of data to be delimited, special codes such as STX and ETX may be inserted. Control information must also extend to the proper handling of the control signals on the physical interface, such as RS-232C, since the two users are no longer directly attached and performing exactly the same functions at the same time.

• **Status Translation.** Data exchanges are not trouble free, and terminal devices may also malfunction or simply need attention for tasks such as loading paper. The status message delivered by a terminal, such as an IBM 3270, cannot be expected to be meaningful to a host computer which believes itself to be connected to an asynchronous device such as a DEC VT-100. The protocol converter must recognize status conditions and relate them correctly in the context of the destination protocol, or handle them within the converter where no equivalent status exists.

Figure 4 shows an example of data isolation. Here we assume that a protocol converter is in use to allow an asynchronous terminal to access a computer using IBM bisync. The host computer **selects** the terminal, sends data, and finally indicates that the exchange is complete. Only the data is meaningful to the asynchronous terminal, and that data must be removed from the **frame** of control characters introduced by IBM bisync. Since most IBM bisync communication lines use EBCDIC as the code set and most asynchronous terminals use ASCII, the converter must also convert the data from one to the other. For example, the letter "A" is a binary value 01000001 in ASCII and a value 11000001 in EBCDIC. Data isolation must also repackage data if the amount of information received from one user is too large to deliver as a single message to the other.

Figure 4 also shows the importance of control handling in the task of protocol conversion. When the host computer sends a bisync **select** command, it expects the selected terminal to send an "ACK" or acknowledgement of receipt and readiness to

Protocol Converters—Bridge To The Future

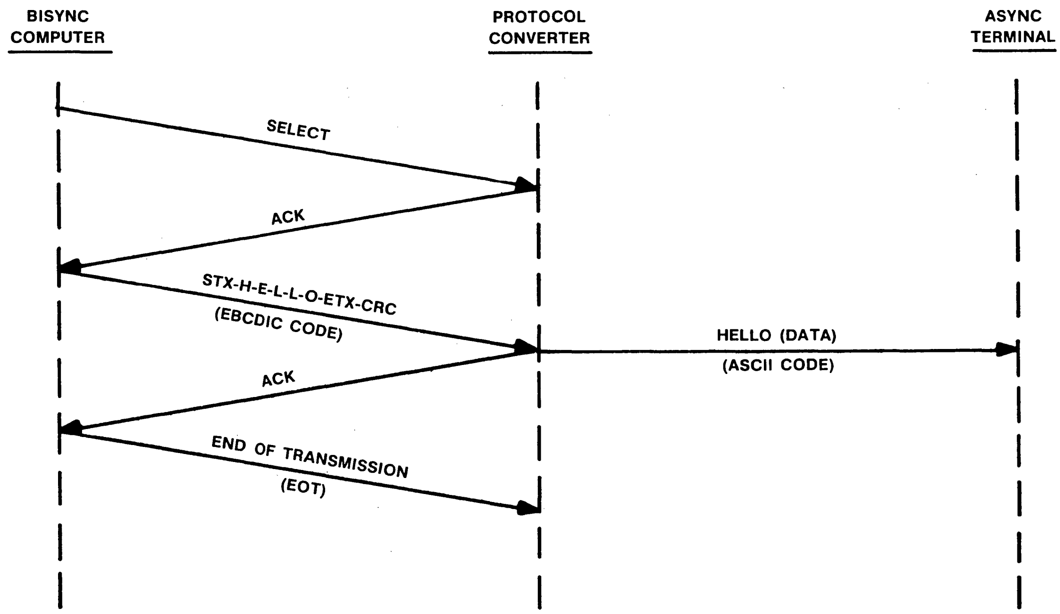


Figure 4 • data isolation in protocol conversion.

receive. If this "ACK" is not received, the host will not send the data, and yet the actual asynchronous terminal will not supply bisync control information. The protocol converter must stand in for the true destination device and send the "ACK," and must also acknowledge any data blocks at the expected points. Furthermore, the protocol converter must control the RS-232C interface itself, raising "Request-To-Send" when it is ready to send data to the asynchronous terminal or to send data or control sequences to the opposite station. This control signal handling is needed because the number and sequence of transmissions from one protocol to the other may be different, even though the data exchanged is the same.

Figure 5 shows an example of status translation. The asynchronous terminal has been selected by the bisync host through the protocol converter and data is being exchanged. The asynchronous terminal becomes overloaded and sends a message to the protocol converter asking for "flow control" of the incoming information, a sequence called an X-OFF. The protocol converter must now either hold the data sent by the host within itself until the asynchronous terminal becomes ready, or signal the host to wait as well. Since an X-OFF has no meaning in bisync, the protocol must **translate** it to the equivalent bisync function, called a Wait-Acknowledgement or WACK. Had there been no equivalent function, the protocol converter would have had to buffer the data internally, a requirement certain to add to the cost of the unit.

A complete example of protocol conversion is shown in **Figure 6**. Here we will assume that conversion is between the 3270 terminal form of IBM bisync and the so-called **batch** form used by the 2780/3780 remote job entry terminals. The host computer is using the 3270 form of bisync, and sends a "select" to the terminal through the protocol converter. Since this form of command is not valid for 3780 devices, the converter translates it to a "bid" for the line. The converter must properly acknowledge the "select," and must wait for the 3780 to acknowledge the **bid**. Data may then be moved from one device to the other, taking into account the differences in the device control protocols. If the 3780 detects an error, the protocol converter must retransmit the data since the block involved has already been acknowledged to the host computer by the converter. If the 3780 reports a status of "out-of-paper" at some point in the exchange, the protocol converter must translate it to the 3270 form of the indication and deliver it at the first opportunity, saving any information which it has waiting for the 3780 lest the data be lost.

■ APPLICATIONS OF PROTOCOL CONVERSION

The primary area of user interest in protocol conversion has been that of 3270 terminal emulation using asynchronous devices. This is partly because of the significant cost differential between 3270 and asynchronous terminals, and partly due to the desire for users to upgrade to a more modern protocol without replacing existing asynchronous terminals. These 3270 emulators were initially converters to IBM bisync but have recently become available and very popular as converters to IBM SNA. Buyers of the 3270 emulators may also use a form of switching to allow several of

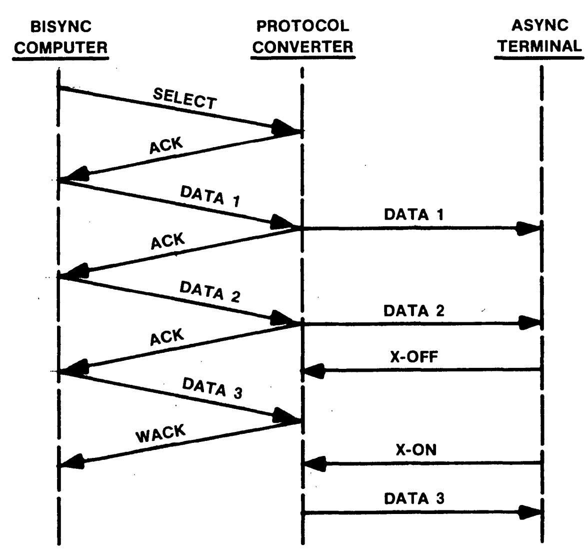


Figure 5 • status translation in protocol conversion.

Protocol Converters—Bridge To The Future

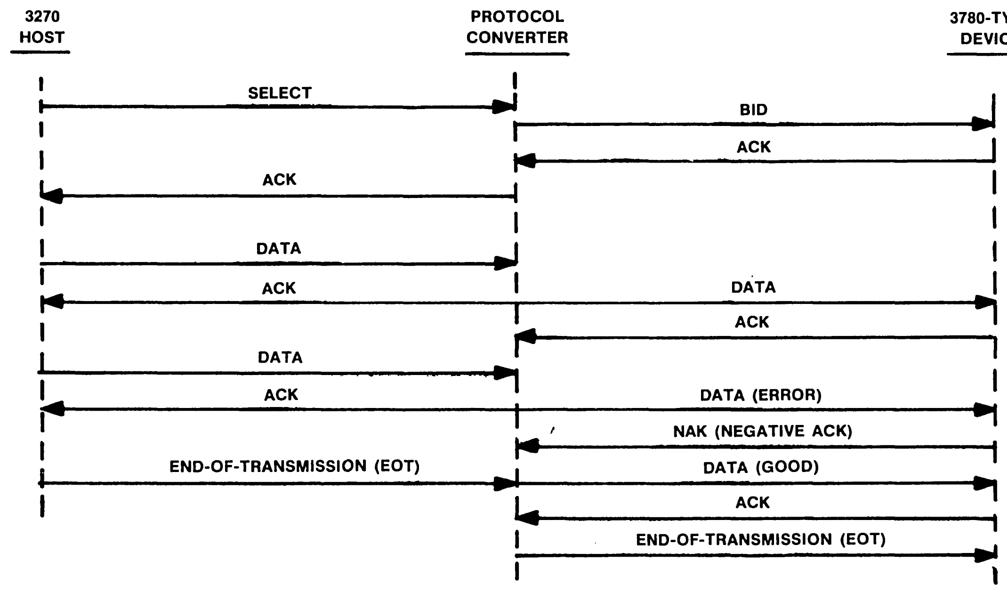


Figure 6 • a complete sequence in protocol conversion.

their asynchronous terminals to act either as normal asynchronous devices into one host computer or as 3270-emulator devices into another, as shown in **Figure 7**. This sharing of terminals between applications and even host computers is of increasing value where business finds the need to provide the ability to interact with several different programs from a single workstation, since it eliminates the need for multiple terminals and the cost and training problems that could arise.

A variation on this application of protocol converters is their use to supply an inexpensive means of extending an existing network in cases where an upgrade in the communication facilities are planned for the future. Rather than purchase an older device which may be obsolete with the new network, the user can acquire a protocol converter and a less expensive set of asynchronous terminals, replacing only the protocol converter when the network change is finally made.

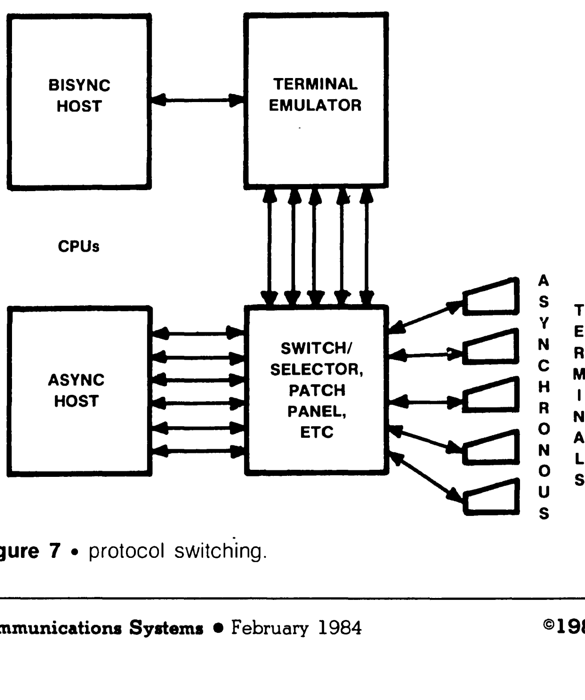


Figure 7 • protocol switching.

Another use of protocol converters which is increasing in importance is that of satellite or network delay compensation. As an earlier figure showed, protocols such as IBM bisync will not operate efficiently over circuits with long propagation delays because the sending station can transmit only as fast as it can receive the acknowledgement of correct reception. As **Figure 8** shows, a protocol converter can be used at each end of the satellite path to convert between bisync and another protocol with better immunity from delay effects. Since a pair of converters is used, the identity of the internal **hop** protocol is not important. It may be a network protocol such as SNA or X.25, or a unique proprietary protocol designed especially for satellite applications. Some satellite network operators offer such **satellite delay compensators** as a part of their service, since studies have shown that performance may fall to as little as 10 percent of terrestrial-line values if a satellite path is taken.

A more general, but also more complex, use of protocol conversion is to support the connection of incompatible devices where changes in company structure of business needs forces a tighter integration of computers and applications. Many larger companies have purchased computers of various types and makes to support what once were relatively small and separate application programs, only to find that pressure for an integrated management information database requires some means of collecting the data at a single point. Banks in particular have found that the direct connection of their various minicomputers and mainframes into a real-time network allows a more timely and accurate picture of their relationship with their clients, a decided asset when capital costs are high. The multivendor nature of many data centers makes the connection of all of the computers and terminals together a major task in protocol conversion, since there are no truly universal protocols. Even in cases where two vendors promise communication support for IBM 3270 bisync, the user often finds that the support is limited to the ability to talk to a 3270 terminal, useless for interchange between computers unless one can also emulate such a terminal to the other.

Another application for protocol conversion is increasing in popularity, particularly among large companies; internetwork gateways. Most corporate communication networks have a single internal protocol, and often serve a single or tightly coupled group of applications. When several such networks exist, there may be a limited need to pass information from one to the other which is not significant enough to give each user a connection into both. These limited needs can be solved through the use of a

Protocol Converters—Bridge To The Future

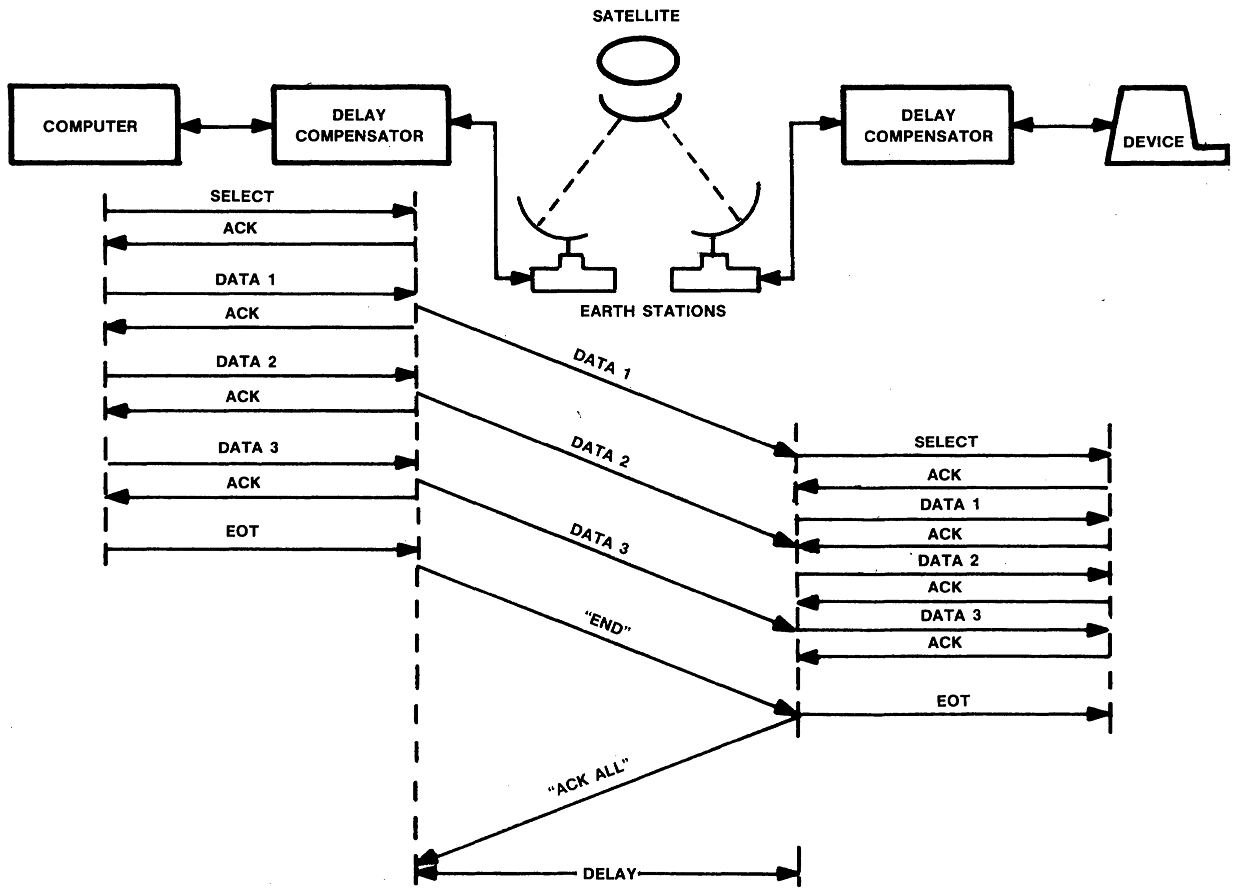


Figure 8 • delay compensation.

"gateway" between networks which serves the same function as the gateways between international packet networks—supporting limited cross-connects. Such a gateway must be a protocol converter that supports the protocols of the two networks so connected, often as a part of a communication processor or front-end processor in the network.

A final application of protocol conversion is the use of personal computers to communicate with host mainframes. Here, the personal computer **off-loads** some processing from the mainframe. This is attractive for two reasons: first, it reduces the load on the mainframe and communication network; and second, it provides a degree of processing autonomy to the user. The actual exchange between the mainframe and personal computer might involve file transfers, database management/retrieval, number "crunching" services, or a combination of all.

The role of the protocol converter in this case is to make the personal computer look like an acceptable terminal device to the mainframe. In the case of IBM, the most common technique is to make the personal computer appear to be a 3274 controller with attached terminals or a 3276 control unit display station. The protocol converter also must compensate for any differences between the personal computer controls and those associated with the attached terminals. Such protocol converters are available in standalone units to which the personal computers plug-in, or by a combination of firmware-software incorporated within the personal computer itself.

Protocol converters are also used to allow personal computers to attach directly to IBM 3274 or 3276 controllers. Here, the protocol converter is usually implemented by a combination of firmware/software contained on a circuit card which fits into the personal computer.

■ TYPES OF PROTOCOL CONVERTERS

The variation in the application of protocol converters is reflected in the variety of such units available. These can be classified by the following groups:

• **Terminal emulators** provide both protocol and device control conversion so that inexpensive terminals can be used to substitute for more expensive types. Most such units allow asynchronous terminals to be used in place of IBM 3270 display system devices, either in a bisync or an SNA environment.

• **In-line protocol converters** connect two devices of incompatible protocol. These are often used to connect two computers together. Although line protocol converters may appear similar to terminal emulators, they often lack the ability to perform device control translation, making them restrictive or useless in applications where the devices cannot be programmed to eliminate such device control.

• **Paired converters**" such as satellite delay compensators are used in pairs so that devices of like protocol can communicate via an internal and intermediate protocol used by neither but more suitable for the characteristics of the transmission path. IBM's current support of X.25 on the 3705 falls into this category, since a pair of devices is used to allow SNA networks to use an X.25 packet network as a delivery service.

• **Gateway converters** provide a bridge, often capable of being used by multiple conversations, between different networks.

Protocol Converters—Bridge To The Future

Examples of this type of converter are devices that provide bridges between local-area Ethernet networks and conventional network architectures such as DECnet or X.25. These differ from other types of converters in that they must typically simulate a network element serving many users rather than a single user.

Figure 3 shows the configuration of a typical terminal emulator/protocol converter. Note that the protocol converter is a combination of communication device and small computer, simulating a cluster of 3270 terminals, typically in SDLC/SNA mode. Most of these units appear to SNA as a Physical Unit Type 2 (cluster controller) and require little or no changes in the SNA Network Control Program in order to operate. The protocol converter carries on the communication dialog with the host computer and maintains a memory portion for each terminal attached which represents the 3270 display memory. When the operator keys data into a field, the protocol converter must **mark** that data as modified so that it can be delivered to the host computer. Since field numbers and attributes such as "protected field," characteristic of the 3270 display system, have not been implemented in most asynchronous terminals, these functions must be simulated as well. The number of types of asynchronous terminals supported by these devices varies from vendor to vendor, as do the restrictions for having more than one type of asynchronous terminal per converter. Most of these devices will not emulate all of the special features of the 3270 Display System, particularly programmable function keys. Costs of these units vary, but often it is possible to obtain a converter capable of supporting a dozen existing asynchronous devices in 3270 mode for the cost of a single 3270 terminal.

Figure 9 shows a line protocol converter being used to connect an asynchronous host computer to a synchronous host operating with the "batch" or IBM 2780/3780 form of bisync. Because both users are computers, device control information such as might normally be required for the operation of the 3780 can be eliminated from the data stream, making more complex device emulation unnecessary. The line converter can also be used to change the speed of the data in case one host operates at a fixed speed unsupported by the other. Applying this type of conversion will almost always require some change in the programming of the computers involved, but the concept can allow a word processor to communicate with an IBM host computer that has only synchronous capability.

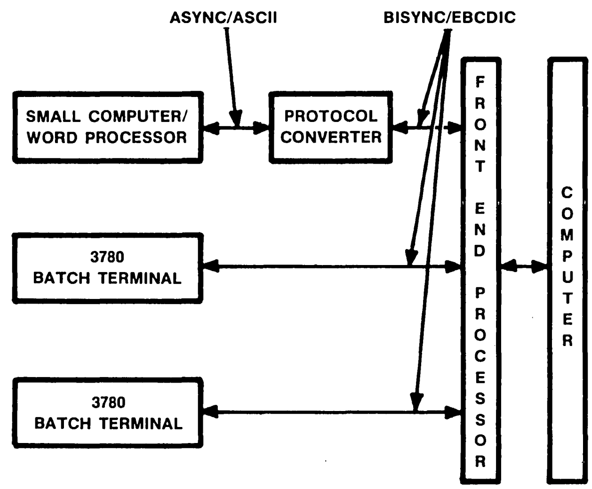


Figure 9 • in-line conversion.

Figure 10 shows a pair of converters used to allow IBM bisync to pass efficiently through a public packet network. Some packet networks have round-trip delays in excess of a second, making the transmission of batches of bisync data nearly impossible. The converters here change the bisync protocol to X.25 for passage

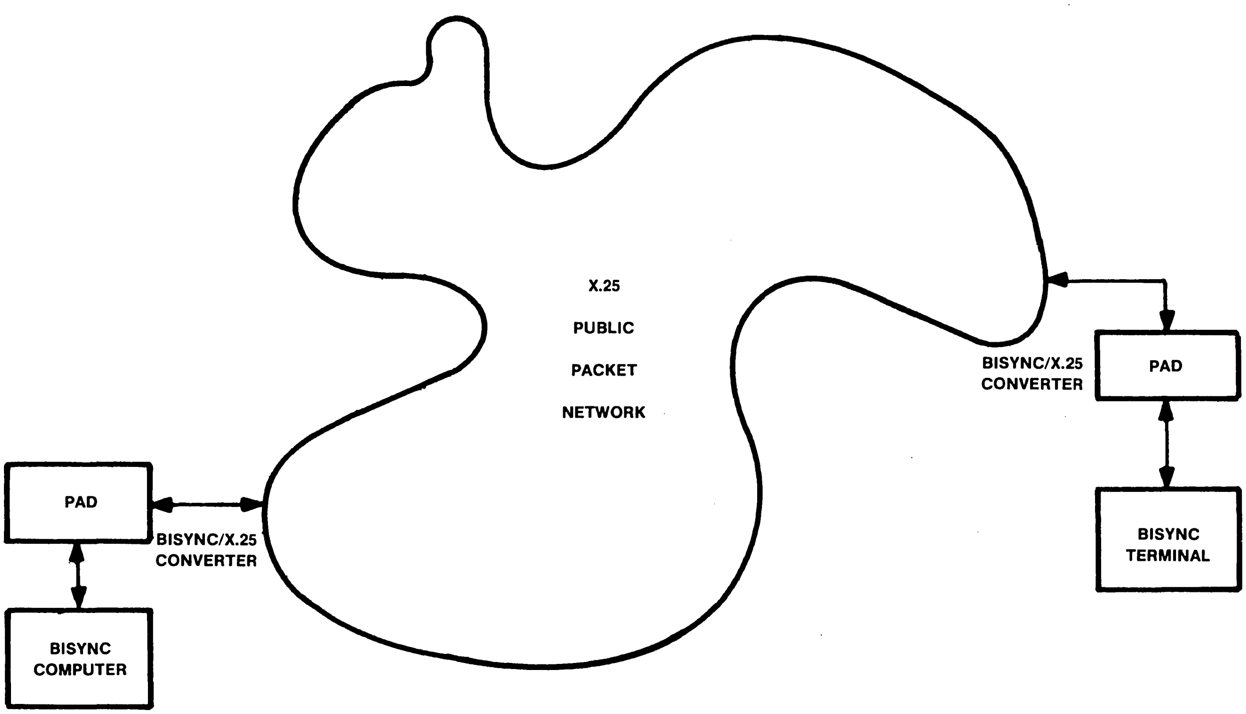


Figure 10 • bisync/X.25 pads in a public network.

Protocol Converters—Bridge To The Future

through the network and in addition provide local acknowledgement of data blocks, eliminating the effects of the round-trip delay on circuit performance. In some cases, the protocol used between the pair of converters may be specially designed for long path delays. Most units of this type are primarily designed for use as bisync packet assembly/disassembly devices for public networks rather than generally for delay compensation, but will operate satisfactorily directly over satellite paths. X.25 also multiplexes users on a single link, so that this form of protocol conversion may also reduce the number of lines required, thus providing additional savings.

Figure 11 shows the second phase of IBM's planned X.25 interface with SNA. Unlike the other forms of conversion, this gateway converter appears to both X.25 and SNA as a network element serving multiple users, a **packet node** to X.25 and a communication controller node to SNA. This allows users in one network to address those in others as though the second network was a real extension of their own. It is not necessary that every user in one of the networks be reflected in the other, a feature that may be useful for connection security or to reduce the chances of an incompatible connection.

■ POTENTIAL PROBLEMS IN THE USE OF PROTOCOL CONVERTERS

Some of the previous examples of protocol conversion give an insight into the complexity of the issue. Since no protocol currently gives **total isolation** between user and communication facility, some communication issues cannot be easily separated from data and thus interfere with the flow of information between users.

All protocol converters except terminal emulators create a **time lag** between the users who are communicating, especially if delay compensation or gateway systems are involved. This time lag has an effect similar to the effect of the delay in radio signaling between earth and moon had on the conversation with the astronauts. Since each word took about one and one-half seconds to transit the path, one party knew only of the status of the other a second and a half ago. One consequence of this delay is called **loss of context** and occurs if one party asks a question which requires a yes/no answer, then immediately asks another without hearing a response. When the answer comes back, it is not clear which question it is associated with. In communication terms, this can occur if a station "listens" to see if the line is free then begins to send, unaware that the other station has done likewise, and the delay in the protocol converter has prevented each from hearing the other. Loss of context is a possibility when important changes in status occur within a protocol, and can be minimized by reducing the number of these changes. Sending many small messages, for example, will expose a user to the "send collision" situation just described much more often than sending a few long messages, since once transmitter and receiver have agreed on the direction of data flow, few important status changes are expected.

Another significant problem in protocol conversion is that of **data overrun**. This can be caused by any one of the following reasons:

• Problems with data delivery from the converter to the destination, causing data to **back up** in the converter.

• Differences in the speed of the two users. For example, using 1200-bps terminals on a terminal emulator converter operating at a line speed of 9600 bps obviously causes data to become clumped in the converter if the host computer sends a long message to the terminal.

• Slight differences in clock speed on the lines to each user, or the timing of the control signals, delays delivery of data to one user.

• Differences in the control sequences required for data delivery at one end causes the converter to back up while satisfying that protocol's requirements.

The effect of data overrun varies. If the protocols involved time responses to messages, a **timeout** can result. If the data buffers in the protocol converter are small, data can be lost. Such losses of data can leave the receiving station missing some vital sequence, causing a **lockup** of the device.

Lockups can also be caused when status conditions occur at one user device that cannot be related in the protocol of the other. Most synchronous devices like the 3270 have a rich set of error

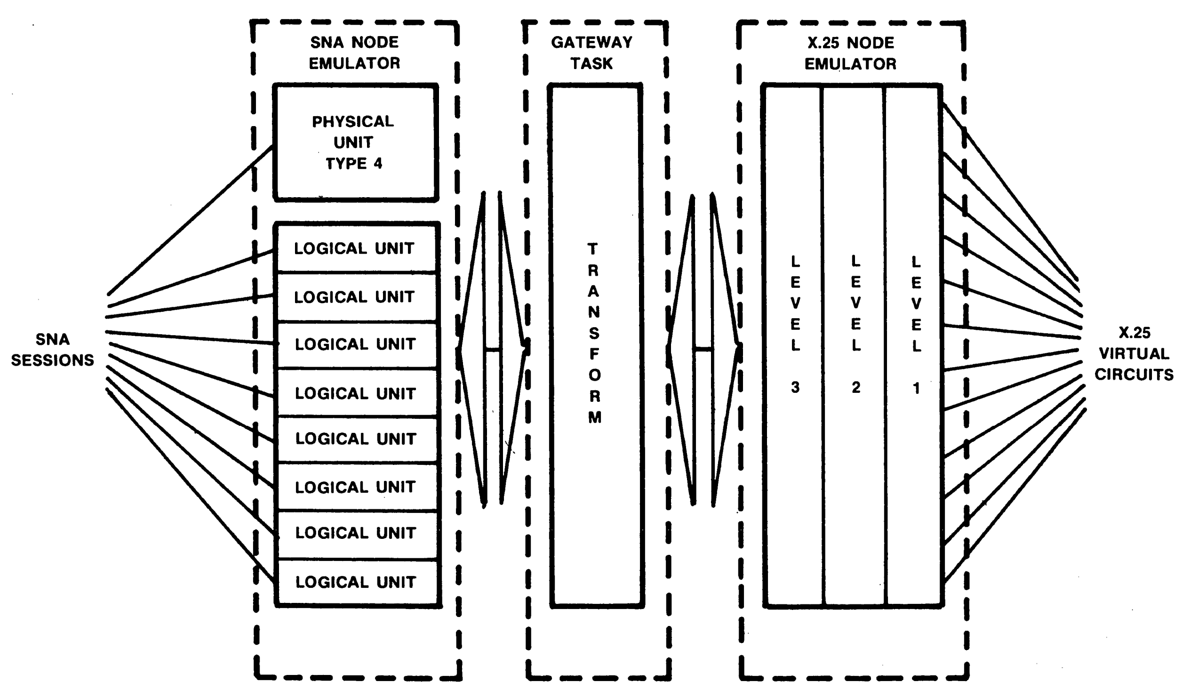


Figure 11 • IBM SNA to CCITT X.25 gateway.

Protocol Converters—Bridge To The Future

status codes and most 3270 control programs use them for error reporting and recovery. Asynchronous devices have typically no ability to report status beyond the limited set of control signal indications on the RS-232C interface. An asynchronous host controlling a 3270 terminal through a converter could, therefore, be expected to have almost no ability to learn the status of the device, leaving its control entirely in the hands of the protocol converter. This lack of precision in status reporting, coupled with the delay introduced by the protocol converter, can cause a host computer to fail to recognize an incomplete transaction, a potential disaster in applications like electronic funds transfer.

Both status problems and context problems can combine to cause the host program to lose control of the terminal in such vital areas as line error monitoring. If the protocol connecting a user to converter is itself not capable of proper error detection and correction, the facilities are lost to the other protocol as well, insofar as that line is concerned. For this reason, protocol converters that support both an error-recovery and a non-recovery protocol should always be placed so that the connection to the non-error controlled device is local and as short as possible, while that of the controlled device transits the actual transmission facility.

A final area of concern is that of unusual or even illegal programming practices. Most protocol converters will operate with the normal communication sequences of the protocols involved, and even with common error recovery procedures, but many will fail when a special command or unusual timing condition occurs. A prime example of this is the use of protocol converter/3270 terminal emulators directly on any IBM host channel rather than over a communication line. There are many commands used in local mode that are invalid or not recommended over a link, and link-oriented terminal emulators will probably not execute them properly.

■ EVALUATING PROTOCOL CONVERTER APPLICATIONS & PRODUCTS

Not all communication users will benefit from protocol conversion, so the first step in evaluating the technology for a business is a survey of the possible points of application:

• 3270 users, whether SNA or bisync, who are planning to buy additional 3270 clusters or terminals may save money by using 3270 terminal emulators instead, especially if they already have asynchronous terminals available.

• Bisync users considering public networks or satellite networks can save time and network charges by converting their bisync to another protocol such as X.25 prior to network entry. Batch transmissions such as print runs or remote job entry are most affected by **propagation delay** because many blocks of data are sent in sequence and are prime candidates for **delay compensation.** This may also save money in network attachment charges, since most networks charge more to pass bisync than X.25. Be sure that your network or satellite carrier doesn't already provide it!

• Multivendor communication users who connect their computers together should **evaluate the protocol** chosen for the connection. It is the best choice, or one dictated because it was the only protocol supported by both systems? Protocol converters could allow a better protocol choice, possibly increasing performance and reducing error susceptibility.

• Users with very old equipment that cannot be replaced and is expensive to maintain and operate may be able to replace such a unit with a protocol converter and an inexpensive modern device.

• Users adding new computers or packages that normally require special terminal equipment may be able to use existing terminals with the aid of a protocol converter.

• Plans for giving a single workstation or office two or more different types of terminals to attach to different applications can often be altered to use a single terminal, a data line switch, and one or more protocol converters. This is normally feasible only if one or more of the applications requires an expensive terminal.

• A complex company-wide network to replace several overlapping existing networks can often be deferred through the use of **gateways** between the networks supported by protocol converters. This should normally not be considered except as an intermediate stage in a master plan to phase in global, application-independent, communication facilities since nearly all communication experts recommend the isolation of communication resources from individual applications.

Once potential application areas are identified, users should review them for possible problem spots:

• Communication networks or circuits supported by custom programming rather than by product-line communication packages should be reviewed carefully for unusual practices, preferably via actual testing. This is true particularly of front-end processors that have been programmed in-house, since expertise at this type of programming is often scarce and the programmers may have had little experience with the unit involved.

• Any devices that have been hardware modified by the user, via vendor RPQ, or by a third party should be tried in the new environment or benchmarked against an unmodified device to assure no changes affected the communication interface.

• Use of protocol converters on lines known to have poor error characteristics should be studied carefully unless the use of the converter will provide better error handling in its protocol.

• Use of protocol converters in conjunction with network or transmission facilities that can introduce propagation delay must be evaluated to eliminate the changes of timeout or poor performance. This includes multiplexers, concentrators, satellite paths, or public networks. Some protocol converters will improve delay immunity, especially those installed in pairs as delay compensators.

• Terminal emulators can create serious operational and training problems if operators are already used to the 3270 terminal. This is particularly true if the asynchronous terminal used with the emulator has few programmable function keys. A pilot program to judge the level of operational impact should be considered.

• Gateway converters can introduce **network security problems** by allowing users of one network to gain access to another. This is particularly true if one network is connected to or is itself a public data network. Addressing and security limitations of gateway converters must be evaluated.

• Some terminal emulators will emulate **only one** of a family of communication devices. Be sure that the converter will emulate a device compatible with the existing software and hardware, and/or other projected purchases.

• Mixing protocol converters with native devices on multidrop lines must be **carefully evaluated.** A test should be considered before committing to a purchase.

When the potential problems with each projected application of protocol conversion have been explored, users should contact several vendors who supply the type of protocol conversion equipment their application requires and ask for literature, a list of users, and a demonstration of the equipment. It is important not to minimize efforts at this phase since the protocol converter will have a major impact on the communication network, probably on the productivity of its users. Some of the specific questions that should be explored with the vendor are:

• Are there any features in the operation of the protocols being converted that are NOT available with the converter attached? Be particularly careful of multidrop or dial-up lines.

• Can the equipment be rented or leased for a period to assure its operation in the user network? If not, can a visit be made to a user with similar equipment and applications?

• If the protocol converter is the terminal emulator type, will it operate at the same range of data speeds as the device being emulated? If not, is there a chance that the application may require the unsupported speeds later?

• Are 3270 terminal emulators that operate in bisync protocol field-upgradable to SDLC form? If not, another converter purchase may be needed if migration to SNA is planned.

• What type of diagnostic support is available from the converter? Will it accept host test messages? Will it report failures of problems with front-panel lights or log them internally for later review from an attached terminal? Remember, a protocol

Protocol Converters—Bridge To The Future

converter may "hide" errors from a computer or front-end processor by correcting (or attempting to correct) them itself. If line errors are not reported in some way, for example, a degrading error pattern on a line can go undetected until it fails completely.

• Exactly what type of device is being emulated? Any protocol converter, even a line or delay model, will **look** like a device in the protocol being converted. For example, a satellite delay compensator may appear as a set of multidropped 3270 cluster controllers. Which type? There are communication and device control differences between models of many terminal cluster controllers, and some may be incompatible with the host programming.

• Are all communication options on the protocol converter compatible with the modems/equipment on both sides? This must include the physical attachment interface (RS-232C, V.35, X.21, etc), the modem/line type (half- or full-duplex), and the control signal uses and timings. Most users have already experienced the problems caused by devices that expect the transmit and receive data lines of the interface have been **crossed** or that Request-To-Send (RTS) has been **looped back** to Clear-To-Send (CTS). More subtle problems may occur where **fast poll** modems with very short transitions in control signals are used—the converter may not respond fast enough.

■ PREPARING TO USE PROTOCOL CONVERTERS

Once a protocol converter has been selected, the installation and cutover tasks must be considered. Some users think that because a converter is supposed to work "just like" another device which is perhaps already in use, it will introduce no additional problems. Actually, the very similarity of the systems may itself cause training and transition difficulties.

If possible, try to operate the converter side-by-side with the existing equipment for a while. This is particularly important where the application is critical. If parallel operation is projected, be sure to develop a plan for cutover, including instructions on how to fall back to the old system in case of failure. If special cables are required for installation of the converter, don't lose the old ones or there may be no turning back. Parallel operation is impractical where major changes must be made to install the converters, so the higher risk in these cases should be balanced by extensive pilot testing on a dummy application.

Operator training may be affected by the installation of a converter, especially if it is of the terminal emulator type. A manual showing the operating procedures in a **was/is now** form will help those already familiar with another piece of equipment make the change to a new one. This type of training is especially important where some functional differences exist between the original system and the converter system. It is also helpful if the operators can become familiar with the new system in a test/training environment rather than on the job. Some things to watch for in operator adjustment to a protocol converter environment are:

• Keyboard layout, particularly numeric pad and function keys.

• Screen appearance; color, brightness, size, viewing angle, and even overall attractiveness.

• Error symptoms, particularly symptoms of line problems or congestion in the network or serving computer.

Fault analysis in a protocol converter environment can be an unhappy challenge. All procedures for network fault diagnosis and isolation should be reviewed before the installation of a protocol converter and trials should be conducted afterward. The problem of "hiding" errors has already been mentioned, but additional problems may arise in loss of remote test control (the emulator may not provide it) and differences in the visible symptoms of problems. The LED's used on 3270 terminals to indicate host activity will not be present in the same form on a terminal emulator and may not be present at all. It may be necessary or desirable to install EIA interface monitors at critical locations, or to make them available to the test staff. These "black boxes" display the settings of the control signals on the RS-232C interface and give a technician or network supervisor an overall picture of the state of the connection.

■ LOOKING AHEAD IN PROTOCOL CONVERTERS

Protocol converters are unquestioningly becoming more popular and more useful. Terminal emulators now support ASCII terminals emulating IBM 3270s in either bisync or SDLC/SNA, and several vendors plan to make these units act as **packet assembly/disassembly (PAD)** facilities for X.25 as well. Converters are being integrated into other products rather than sold standalone, beginning with front-end processors and moving to multiplexers, network concentrators, and even data switches.

Terminal emulation has been the main area of interest in protocol converters, but delay compensation is closing the gap in attention by users if not in products available. IBM bisync satellite delay compensators are being sought to support the spread of satellite-based **Digital Termination Systems (DTS)** for which over 30 companies are seeking an FCC license to offer. Increasing cable density and microwave congestion in heavily-populated areas such as the northeast corridor have prompted ventures to start a "satellite farm" of earth stations near cities which would serve the communication needs of the business and residential populations. The effect of this on the performance of some protocols has already been discussed, yet it seems inescapable that data communication must either accept satellite paths or pay premium charges for terrestrial links. Delay compensation could save thousands in line charges.

Protocol gateways probably offer the ultimate in the development of protocol conversion. These units will eventually evolve into general-purpose engines to link users and networks regardless of their communication protocol, serving in the real world the ideal of user isolation which the OSI Reference Model will eventually bring to buyers of new equipment with the most modern protocols.

Protocol conversion is not the answer to all of the problems of network evolution. It is often necessary to adjust applications, however slightly, for its optimum use. Improperly selected, it may only move a user from one blind corner to another. But with all of the potential problems, protocol conversion continues to excite the use as few communication products can, and even conservatives agree that it can ease the transition from the residual equipment base of yesterday to the integrated network of tomorrow.

• END

Modems & Multiplexers

An Introduction to Basic Concepts & A Guide To Applications

■ INTRODUCTION

Modems and multiplexers are essential components of data communication networks. In most instances they form the digital path between the communicating parties, whether they be terminals or host computers, using the voice or "analog" facilities of the phone system as a base. Modems actually perform the conversion between the digital signals used by computers and the audio tones which can be carried by the telephone lines. Multiplexers allow multiple data conversations to share a line, reducing the cost of providing a communication connection to each user pair. Both services are so fundamental to successful and economical data exchange that particular care must be exercised in evaluating the products and features and applying them to a user environment. This tutorial provides a basic understanding of modem and multiplexer technology, explans the essential features and operating parameters, and describes the application of both products to the real user communication world.

■ THE PURPOSE OF MODEMS

Computers and terminals exchange data as a series of pulses which represent binary digits, or **bits**. These bits are grouped into larger units called **characters or bytes**, and each letter, number, or symbol used has a unique binary code. The particular set of codes being used must be agreed upon by both the sender and the receiver and is called the **code set**. The most common code sets used today are Extended Binary-Coded Decimal Interchange Code, or EBCDIC, and the American Standards Code for Information Interchange, ASCII. EBCDIC was devised by IBM and is used almost exclusively by the larger IBM mainframes and devices which must communicate with them. ASCII and international versions of it are widely used by nearly all other communication devices.

Binary data in digital form is useful for directly connected devices such as local terminals, but digital data does not pass through the telephone system efficiently. **Figure 1** shows why. The digital signal consists of a series of pulses with very short rise (leading edge of signal) and fall (trailing edge) times. Technically, these quick transitions are a result of high-frequency harmonic signals—often over 10,000 cycles per second. Since the telephone system is designed to carry human voice and not high-fidelity music it cannot reproduce these signals and they are lost in the path to the receiver. The result of this loss is a blurring of the digital pulse, making it difficult to see exactly when it begins and ends. This in turn makes timing the bits difficult or impossible and thus prevents proper location of characters. Data communication cannot occur using digital signals over telephone lines. To use the phone system, it is necessary to "speak its language"—analog.

Modulation • Since the phone system operates on audio tones or analog signals, the first task in using them is the conversion of the digital signal to audio, a process called **modulation**. The receiving station must then reverse the process, demodulating the analog signal and recapturing the digital pulses. The "**MOD**ulation/d**EM**odulation" combination creates the acronym for the device used to perform this process—a modem. Both functions are combined in the modem, allowing either party to send and the other to receive.

The technique used for modulation varies as the requirements of the application change. Simple, low-speed, devices may use modems which employ a form of frequency modulation called **frequency-shift keying (FSK)**. This represents changes in the binary bit pattern by changes in the **frequency** of an audio tone. The line is assumed to be in a steady binary-one or **mark** state when it is idle, and this state is represented by one frequency of

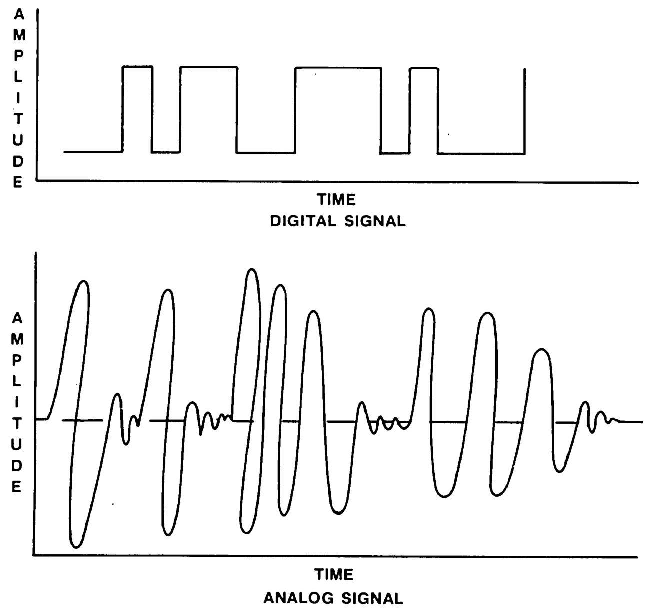


Figure 1 • digital vs analog signals.

tone. When the data bit value "zero" is sent, the modem changes to another tone frequency. This causes a characteristic, almost musical, effect during the sending of data which many users of timesharing systems using this type of modem have come to identify.

FSK modulation works well for relatively low speeds, but as the speed of the digital signal increases the time allocated to shift frequencies reduces so that both the production and detection of the change become more difficult. At 1,000 cycles per second, a single complete cycle of an audio tone requires a thousandth of a second. Since changing frequencies in the middle of a cycle is difficult (to both sender and receiver!), a single bit would have to take at least that long to send, resulting in a data rate of about 1000 bits per second. Since that speed is unacceptable for many applications, another modulation technique must be used for higher speeds.

One alternative is called **phase-shift keying (PSK)**. This uses changes in the **phase** of a signal, or its timing relationship to a fixed reference, to indicate a change in the bit pattern. PSK is used in many medium-speed modems, and is combined with **amplitude modulation (AM)** in high-speed applications to form **quadrature amplitude modulation (QAM)**, the prevailing standard at 9600 bits per second and higher.

Modulation techniques may use all of or part of the frequency range or bandwidth of a telephone channel. If the modulation uses less than half of the bandwidth to carry the data signal, the unused bandwidth may be used to carry the transmission of the other station, permitting both stations to send and receive simultaneously; this is called **full-duplex** operation. **Figure 2** shows some examples of bandwidth use by popular modems.

Baud vs Bits Per Second • Even high-performance modulation schemes cannot accommodate a fast enough transition in signal

Modems & Multiplexers

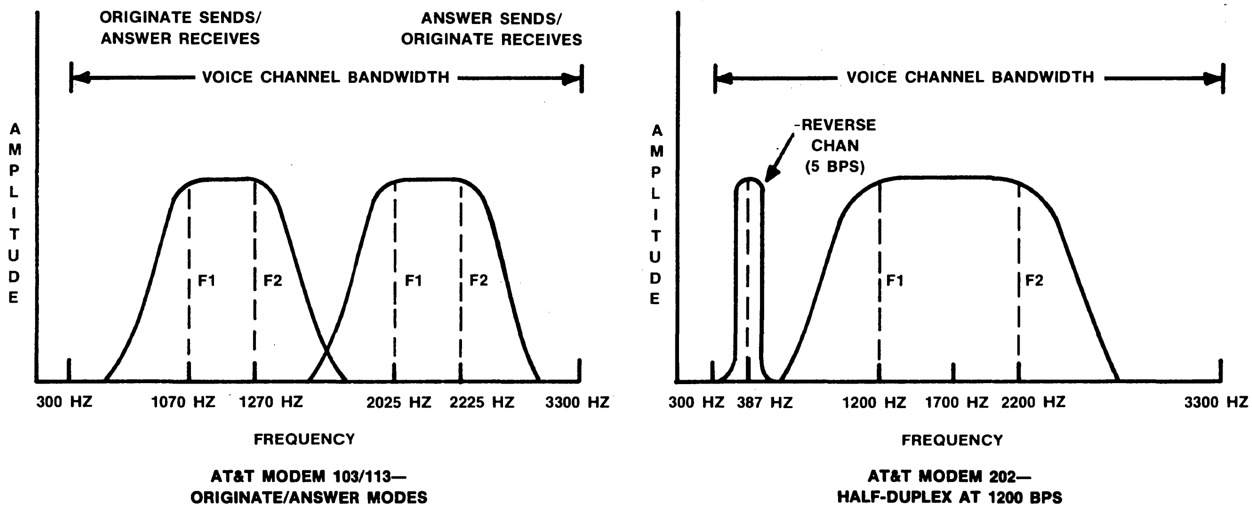


Figure 2 • modem bandwidth usage.

values to track individual bit changes at high speeds. The maximum rate of signal change that can be dependably used is a function of the spectrum of frequencies which the telephone system will pass—the **bandwidth**. This limit is approximately 2400 signal changes per second. Since signal changes per second is an inconvenient unit of measurement, this figure is normally represented using the old telephone term "**baud**." The baud rate of a line is **equal** to the bit-per-second rate **if** the modulation represents every bit change as a signal change. Higher speed modems, however, combine bits into groups of two, three, or even four and assign each pattern a unique signal value. For two-bit groups, a total of four possible bit patterns exist (00, 01, 10, 11), each having its own signal indication. In phase shift modulation, for example, a shift of 90 degrees could represent "01," 180 degrees could be "10," 270 degrees as "11," and no shift as "00." Other values are normally used for engineering reasons, but the principle is the same. Using the bit-pattern encoding technique, a change rate of 2400 baud can support 9600 bits per second (bps) through grouping into 4-bit groups. While the term **baud** is thus **rarely equivalent** to **bits per second**, incorrect usage of it by the computer industry has corrupted the original definition.

Full- & Half-Duplex • Many communicating users exchange data in much the same way as people converse—each taking a turn being the sender while the other receives. This mode of communication is called **half-duplex** while the mode which permits both users to simultaneously talk and listen is called **full-duplex**. At low operating speeds, the portion of the phone line bandwidth which is required to send the data in either direction is small enough that both directions can simultaneously use the same line. This **split-band** operation over a two-wire line, common at 300 and 1200 bps, is now usable up to 4800 bps with some newer modems.

To send data full-duplex at higher speeds, a **second line** is required. This so-called **four-wire service** provides the user with the equivalent of two telephone circuits. It is normally requested as a part of private leased line service, but some modems will permit the use of four-wire operation over two dial-up connections.

If higher speed modems are operated on two-wire connections, the communication **must** be half-duplex. This means that the modems must alternate between sending and receiving. Each time the direction of transmission changes, called **turning the line around**, there is an interval required for the path to electrically stabilize and the modems to **train** (adjust) to the new direction. Modem **turnaround delays** can be a significant factor in cases where the direction of the line changes often, such as with polled lines.

Synchronous & Asynchronous • Whatever the method used for converting between digital and analog, all modems need a way in which the sampling of bit values and the separation of characters can be controlled. If the sending station begins to change the line state with digital bit values, the receiver must begin sampling the output at the proper time or data may be lost or misinterpreted. One way to assure proper data recognition is to begin each character with a dependable transition from the **mark** or binary one state to the **space** or binary zero state. This is called a **start bit**. The receiver, seeing the change, begins the timing of its bit samplings at the end of the start bit. At the end of the character, a "quiet" period of marking line may be assured by defining a **stop bit** interval. This framing approach allows the sender and receiver to operate their stations without a strict synchronization of their bit sampling, or "clocking" as it is called. The start/stop bit transmission method is called **asynchronous** because the clocking is not synchronized between stations.

Asynchronous transmission, or "async" is relatively inexpensive to implement in both terminals and computers. For cases where the terminal and the computer are directly connected, it often provides the best overall means of attachment. But for communication lines, async poses some special problems. The bursty nature of its traffic makes it difficult for the modems to adjust themselves to the conditions of the telephone circuit, "training" for the least possibility of error. To overcome these difficulties, another method of data transmission called **synchronous** was devised. In synchronous transmission, the data bits of one character follow the bits of the previous character without any intervening space and without start or stop bits. So that the receiving station knows when the data **block** has begun and can properly "count off" groups of bits into characters, a special pattern of bits called a **sync** character is sent preceding the data. Since the data is now sent in a continuous stream, the modem can better train for its correct reception. In order to use the synchronous modem techniques, the communicating devices must be capable of sending their data continuously as a block, called **buffering** the data. The buffering capability and its associated cost are the price of the speed advantage of synchronous devices; most asynchronous links are limited to 1200 bps while synchronous communication over the telephone system at up to 16,000 bps can be achieved and is now supported by modem vendors such as Paradyne.

■ MODEM APPLICATIONS

As the previous discussion on synchronous and asynchronous transmission of data shows, there are variations in modem design based on application of modems to business communication. The variations are intended to optimize the cost/performance ratio of

Modems & Multiplexers

the modem, and reflect the different ways that modems might be used. For example, modem use can be classified in the following groups:

- **point-to-point** (two-party) use or **multipoint** (three or more parties)
- **dial-up (DDD)** or **private** (dedicated) leased lines
- **voice-grade** or **wide-band** lines
- **two-wire** or **four-wire** lines
- **full-duplex** or **half-duplex** communication
- **conditioned** or **unconditioned** lines

The categories shown are interrelated so that, for example, the choice of conditioned or unconditioned lines is available **only** for private leased lines. **Figure 3** shows the relationship in these alternatives.

Modems, Medium-Distance Modems, Limited-Distance Modems & Line Drivers • By convention, the term "modem" has come to apply to devices used for modulation/demodulation of a digital signal to permit its passage over conventional dial-up or leased telephone circuits. These devices must operate without a **copper path** (metallic circuit) between users because telephone company equipment including repeaters, switching offices, and microwave or satellite links interrupts the copper path. At least one vendor provides "medium distance" modems designed for telephone circuits at distances up to 50 miles. Medium distance modems do not equalize undesirable line conditions to the extent of regular modems, and are, therefore, cheaper. Less expensive devices called **limited distance** or **short haul** modems can be used where there is an assurance of a copper path between the modem devices. These devices operate in pairs like conventional (long-haul) modems but rely on the more dependable electrical characteristics of the copper path to simplify the modulation scheme. The transmission range is limited by transmission speed and the wire size of the conductors in the copper path. Distance is proportional to the diameter of the conductors and inversely proportional to transmission speed. Wire size (conductor diameter) is specified by American Wire Gauge (AWG) numbers. Conductor diameter diminishes as the numbers become larger. Typical wire size for private lines up to about 15 miles ranges from AWG 19 through to AWG 26. Private lines are also available from the telephone company. Users should realize, however, that conventional telephone circuits are equipped with loading coils for voice transmission and restrict data rates to 9600 bps; unloaded Telco circuits or user-provided metallic circuits support much greater data rates. Some limited distance modems are designed for unloaded circuits only, and can operate at rates up to 19,200 bps at a very reasonable cost. AT&T specifications for transmission over private lines are detailed in **AT&T Technical Reference Publication 43401**. Most limited distance modems meet this specification, which requires the units to filter transmitted signals so they will not cross over and interfere with adjacent wire pairs in the same cable.

For connections of several thousand yards or less, another device called a **line driver** is employed. Properly speaking, this is not a modem at all but a means of sending the digital pulses further than the normal 50-foot limitation of the RS-232C interface. The extra distance is achieved by amplification and other digital techniques. The line driver may also be called a **null modem**, **local distribution unit**, or **modem eliminator**, but should not be confused with devices used only to supply proper external data clocks and proper control signal values to permit a device which normally operates through a modem to connect locally. Line drivers are typically used with synchronous terminals connected to a host computer in the same room. Unlike modems, they are **not** connected in pairs (back-to-back).

Simultaneous Voice/Data Modems • Because of a trend to combine voice and data capabilities in a new generation of "executive workstations," modem vendors have developed modems that accommodate both voice and data communication over the same wire pair. Most of these units use a technique called **data over voice**, whereby data signals are communicated at a higher frequency than voice signals so that the two do not

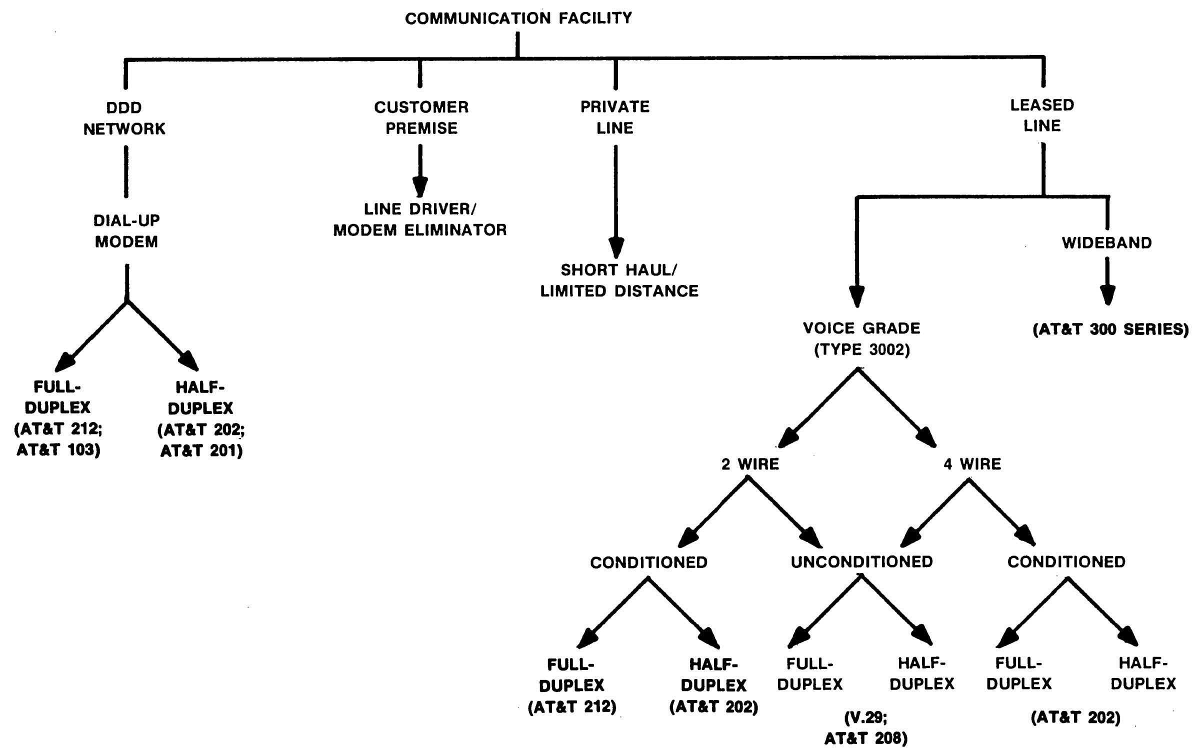


Figure 3 • communication facility vs modem type.

Modems & Multiplexers

interfere. Simultaneous voice/data modems tend to be short-haul devices only, and are useful in PBX environments where users can take advantage of existing telephone wiring and save the expense and hassle of adding dedicated data cables.

Dial-Up vs Leased Lines • Dial-up connections are those made through the normal switched telephone system (DDD or PTN). The paths taken by such connections are not predictable and vary significantly in terms of noise and other important characteristics. Furthermore, a series of echo suppressors are present to prevent the annoying reflection of the speaker's voice. In addition, the local loop from the central office to the connection is always two-wire. This means that both sending and receiving of data at a communicating station must share the same wires. Unless the modems are low-speed models, the bandwidth required for either sending or receiving will fully utilize the single pair connection, so the modems must operate alternately sending and receiving or half-duplex. Turnaround delay on dial-up lines is significant —typically 150 milliseconds.

The connection between the modem and the dial-up circuit may be made directly if the modem is certified by the FCC for direct connection (FCC Docket 19528 Part 68), or via a **Data Access Arrangement (DAA)** if it is not. Either direct-connect or DAA attachment may support automatic dial and answer or may require an instrument for manual dial and answering. Some modems connect to the line by providing a rubber cup set for the insertion of the telephone handset itself. This technique is called **acoustic coupling** and is useful when a modem is to be used with a line normally used for voice, such as in a PBX environment, or where a modem and terminal are to be carried from phone to phone. **Acoustic couplers** are common at 300 bps and available up to 1200 bps, but the variations in the quality of the handset can create unfavorable error performance. The units are almost always less expensive than the combination of a modem and a DAA, but direct-connect modems are now available which are actually less expensive than an acoustic coupler of the same speed.

Leased line modems do not require dial/answer logic since the leased path is fixed. A leased line may have the same bandwidth of a dial-up line, called a **voice grade or Type 3002** line, or it may have a greater bandwidth **(wideband)** such as AT&T Communications Series 8000. Since the path of a leased line is fixed, the circuits involved may be preselected for a particular level of transmission characteristics, called **conditioning**. Leased lines may also be **multidropped or multipoint**. This means that the line has intermediate stations or "drops" between the two endpoints. Multidrop circuits are actually formed by linking **separate** leased lines at the central exchange.

Line Conditioning & Transmission Errors • The telephone system is not a perfect electrical path even for voice communication, and as the speed of data exchange increases it becomes increasingly difficult for modems to manage the precise signaling required for accurate information transfer. There are a number of problems that affect the performance of a circuit. First, the line may fail altogether, normally because of a fault in the local loop or central office. Second, the line may become subject to unusual electrical interference or other sources of random noise, decreasing the ratio between the strength of the signal and the amplitude of the noise pulses, called the **signal-to-noise ratio**. A third source of problems is variation in the strength of the signal itself, independent of noise, called **amplitude distortion**, which may cause periods of "drop-out" when the signal becomes so weak that it cannot be properly decoded at the receiver. The fourth and final source of circuit problems relates to changes in the analog signal itself which is caused by uneven propagation of low and high frequencies. Called **phase jitter**, this condition severely affects modulation techniques that use phase shifts to represent bit patterns—the technique used by most high-speed modems.

One way to minimize problems with leased circuit transmission characteristics is to pay a premium for a line specially **conditioned**. The telephone company will then either select a path which meets the more stringent requirements of the conditioning or specially service a path to meet the standards. Conditioning is available in two types, **C** and **D**, and at several levels within each type; both are available for either point-to-point or multipoint lines. Even with line conditioning, however, AT&T will not guarantee optimum bit-error rates.

Type C conditioning improves the frequency response of the channel and the propagation delay of various frequencies, called **attenuation distortion** and **envelope delay distortion**. Attenuation distortion results from uneven levels of transmission of the frequencies from 300 hertz to 3300 hertz. Type C conditioning assures what in high-fidelity terms is known as a "flatter" frequency response. This means that the difference in volume (amplitude) between two tones of different frequencies is minimized. Envelope delay distortion is caused by the fact that the time it takes for a tone to pass between sender and receiver may not be exactly the same for all frequencies. This delay can cause problems when multiple tones are generated for a series of bits and the tones arrive with a different time relationship from that of their generation. Type C conditioning equalizes propagation delay. Both envelope delay distortion and attenuation distortion can be conditioned to various levels by selecting C1, C2, C3, C4, or C5 conditioning. The higher numbers have the tightest controls on the distortion and thus the best transmission characteristics.

Type D conditioning minimizes line noise called **C-notched noise** and **harmonic distortion**. The former condition causes a poor signal-to-noise ratio on the line, generating random bit errors. Harmonic distortion results from the generation of multiples of an initial tone, similar to two notes an octave apart in music. This "harmonic" may cause the modems to lose tracking completely.

Modems operating at relatively low speeds may not require any conditioning at all, but nearly all 9600-bps modems require some form of line conditioning. If the modem is equipped with internal compensation providing what is known as **adaptive equalization**, Type C conditioning is not required, but Type D is often required. Some new 9600-bps modems will operate satisfactorily even on unconditioned lines. **Figure 4** shows the effect of Types C and D conditioning.

Equalization neutralizes the undesirable electrical characteristics of a communication line that distort the transmission, increasing the error rate and degrading operating performance. Equalization and line conditioning are extremely important to optimize modem performance at rates above 2400 bps. Both act to minimize signal distortion, such as envelope delay and amplitude attenuation, which worsens with increased speed. Equalizers are essential because line conditioning does not totally neutralize distortion; it assures that distortion is reduced to specific limits defined in the specifications for each type of line conditioning. In some cases, equalization eliminates the need for extra-cost line conditioning. It is vital to high-speed communication over the DDD network.

■ MODEM SELECTION PARAMETERS

Modem features and operating parameters span a broad spectrum of applications, and it is necessary to thoroughly understand both the feature and the impact in order to match a modem to a specific application. Some parameters are dictated by the application equipment while others can be optimized through the proper selection of a modem or modem options.

Many modems are designed for either dial-up or leased-line operation, so one of the first issues in modem selection is the line type. Very high-speed data exchange requires leased line facilities, and leased lines are often the best choice where the amount of data to be exchanged is large enough to require connection for hours at a time. The decision to use leased lines is primarily one of economics—would it cost more to lease the line or to pay the dial charges for the time you expect to be communicating?

Another parameter which is normally not selectable is the question of synchronization. Synchronous devices (such as the IBM 3270 or 3780 terminals, all SDLC/SNA devices, and X.25) require synchronous modems. While it is possible to buy modems or devices (converters) that permit asynchronous devices to use synchronous transmission techniques, this is not normally cost justified and may have unwanted application impact unless the asynchronous device communicates in block mode rather than the normal character-by-character mode.

Modems & Multiplexers

SIGNAL IMPAIRMENT	TYPE 3002 UNCONDITIONED	TYPE C CONDITIONING	TYPE D CONDITIONING
HARMONIC DISTORTION	C	NE	C+
AMPLITUDE DISTORTION	C	C+	NE
ENVELOPE DELAY DISTORTION	C	C+	NE
NOISE	C	NE	C+
IMPULSE NOISE	C	NE	NE
PHASE JITTER	C	NE	NE
FREQUENCY SHIFT	C	NE	NE
ECHO	C	NE	NE
PHASE HITS GAIN HITS DROPOUTS	UNCONTROLLED		

C: CONTROLLED AT LEVEL SHOWN
C+: ENHANCED CONTROL AT LEVEL SHOWN
NE: NO EFFECT

Figure 4 • effects of conditioning.

Transmission mode is a modem parameter that may be selectable. As indicated previously, some communication takes place in two-way simultaneous mode (full-duplex) and some in two-way alternate (half-duplex). A very small number of devices communicate in one direction only, called **simplex** transmission. A full-duplex device **MUST** use a full-duplex modem, but some half-duplex devices can profit by being operated on full-duplex lines.

In the previous discussion of full- and half-duplex modems, the time required to **train** the modem pair after a change in transmission direction, or line turnaround, was noted. This time can be quite considerable—150 to 250 milliseconds is not uncommon at high speeds. This turnaround time is equivalent to the time to transmit over 200 characters at 9600 bps! If a half-duplex batch transmission protocol such as IBM 3780 bisync is used at 4800 bps with a modem with a 200-millisecond turnaround, the result would look something like **Figure 5**:

1. the sender transmits a block of 512 characters, requiring 860 milliseconds
2. the line turns around, requiring 200 milliseconds
3. the receiver sends an "ACK," or acknowledgement of reception, requiring about 6 milliseconds
4. the line turns around, taking 200 milliseconds

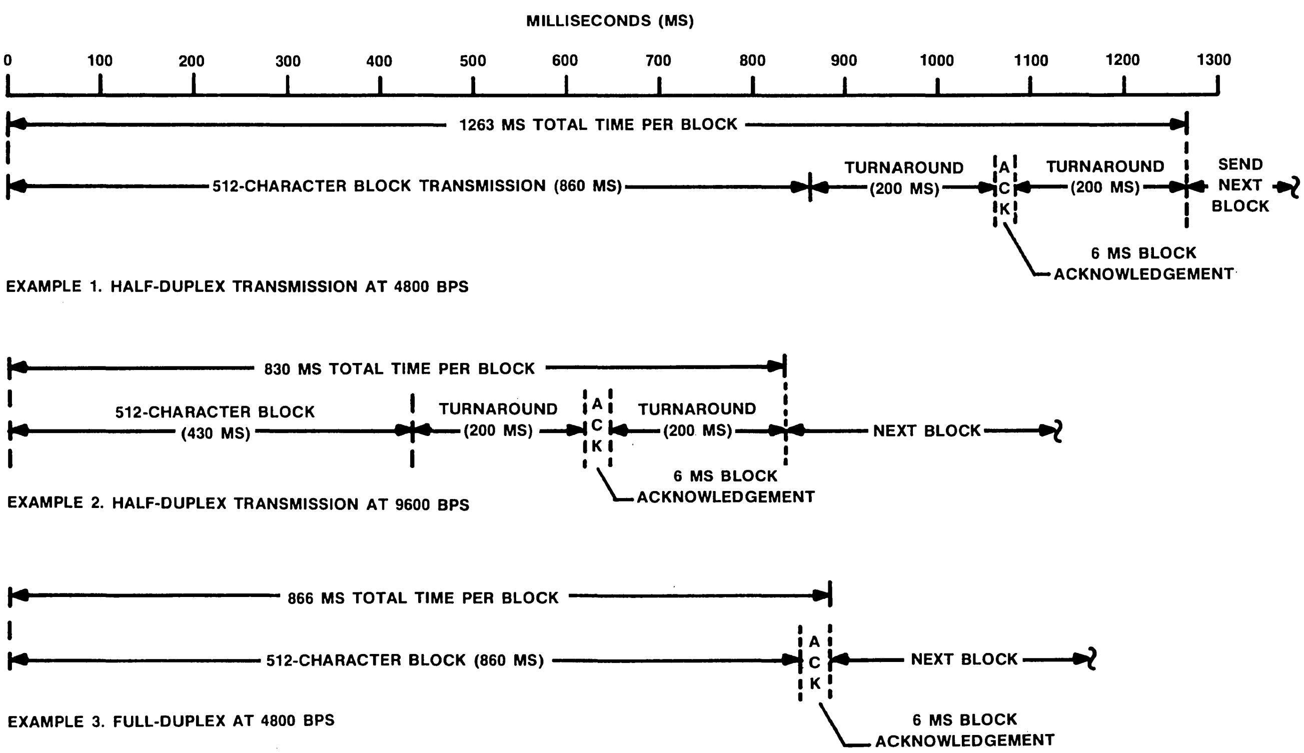


Figure 5 • effects of half-/full-duplex vs speed on transmission time.

Modems & Multiplexers

The sending of one block of data thus required 1263 milliseconds. If we were to use a 9600-bps line and modem pair, the time would be cut to 833 milliseconds since the data and ACK would require only half the time, but the turnaround time would remain the same. But if we were to use a full-duplex modem and line pair we would eliminate the turnaround; since both stations can send at any time, no change in direction or training is required. This would allow us to send the block in 863 milliseconds, only 30 milliseconds more time than doubling the line speed. The difference is even more dramatic with shorter data messages, since the turnaround delay remains fixed while the transmission time for a data block becomes smaller. Thus, users of leased lines and half-duplex protocols should compare the costs of fast modems and half-duplex operation against slower modems and full-duplex operation. On multidrop lines, no single slave station can maintain full-duplex operation without taking the line from other slaves, but a four-wire circuit can still be used to advantage by having the master send continuously on one pair and the slaves send half-duplex on the other. This mode is called **running with constant carrier outbound** and eliminates half the turnaround training delays by allowing the master station to send without training or turnaround delay.

Fast Poll Modems & Clear-to-Send Delay • If half-duplex operation is required, it may still be possible to reduce the turnaround delay on a line through the use of **fast poll** modems. The name for these modems is derived from the fact that one of their prime uses is in a polling environment where a master station continuously polls a multidrop line with many slave stations attached. The time required to poll the entire line may be excessive if the modem turnaround delay is long, so the "fast poll" modem is designed to reduce training time to a fraction of the normal value (usually to less than 50 milliseconds). The term **clear-to-send** delay or **CTS** delay is usually associated with training time because the delay results from the fact that the modem, having received a **request-to-send** indication from the terminal or computer, waits until the training is complete before returning a "clear-to-send" or authorization to proceed with the message. This delay is often controlled by a timer in the modem rather than by actually detecting the successful training completion and may be adjustable to one of several values to accommodate changes in line conditions. Users with half-duplex operations, and protocols that require changing transmission direction often (bisync, for example), would normally run at the shortest delay possible and select modems with minimum delay characteristics.

Compatibility • Modems are used in pairs, and both modems in the pair must signal and modulate in precisely the same way or no communication is possible. When purchasing modems for new circuits it is usually best to buy a pair from the same manufacturer, assuring that the units and options are compatible. When replacing a modem or adding one or more modems to a multidrop circuit, or when adding new dial-up equipment, it may be desirable to use modems of a different make and model. This requires a standard to which all units in question are referenced as compatible, and two such standards exist: AT&T and the Consultative Committee for International Telegraphy and Telephony (CCITT).

AT&T modems are available for all speeds from 300 bps or less to 9600 bps. Each modem is designed for a particular combination of speed, line characteristics, and synchronization:

• **AT&T 103/113 modem** is a full-duplex dial-up modem operating at up to 300 bps in asynchronous mode. The 103 may either originate or answer the call. 103-compatible modems are available for direct connect, connection via DAA, or acoustic coupler connection. The 113, a variant on the 103, is identical in modulation and characteristics but is available either as an originate-only (113A/D) or an answer-only (113C) modem.

• **AT&T 202 modem** is a half-duplex, asynchronous modem operating up to 1200 bps on dial-up lines and to 1800 bps on leased, C2 or better conditioned lines. Because the 202 is half-duplex, it requires that the user devices be able to signal their intention to transmit via the RS-232C Request-To-Send (RTS) interface lead. Many terminal devices cannot generate this signal, so care in applying the 202 is indicated. The 202 is available for dial-up and in a single package as the 202C, or dial-up and grouped in up to 8 channels in a single package as the 202S. Leased-line versions in single-channel packaging are Model 202D/R, and packaged in up to 8 channels per group as Model 202T.

• **AT&T 212 modem** is the solution for users who need more than 300 bps but must operate full-duplex or use terminals that cannot generate a Request-To-Send signal. It operates with either synchronous or asynchronous equipment at 1200 bps, and in 103-compatible mode at up to 300 bps. This dual-speed, dual-mode operation makes the 212 very useful for dialing into timesharing computers or for other applications where many users with different modem speeds may contend for the same set of dial-in ports.

• **AT&T 201 modem** operates at 2400 bps. This is a half-duplex, synchronous modem that can be used on either dial-up or leased lines. The dial-up models are the 201A (which operates only to 2000 bps) and the 201C. The 201B is a leased line version. The 201A is considered obsolete, and most new applications reject the 201 series in favor of newer and higher-speed modems that operate over the same types of lines and at comparable costs.

• **AT&T 208 modem** operates half-duplex at 4800 bps, synchronous only, over dial-up or leased lines. No line conditioning is required. The 208A is the leased-line version; the 208B is a dial-up line version. Point-to-point or multipoint lines are supported with the 208A, and the training time of the modem is selectable at 50 or 150 milliseconds.

• **AT&T 209 modem** operates synchronously full-duplex over leased, D1 conditioned lines at 9600 bps or as a multiplexer/bandsplitter that can allocate 2400-bps segments to four or fewer channels.

CCITT Compatibility • specifies modem compatibility with a specific recommended standard of the Consultative Committee on International Telegraphy and Telephony, Geneva, Switzerland. The CCITT organization is a world leader in establishing recommended standards for telegraph and telephone communication and data transmission. CCITT recommendations define modem operating parameters from less than 300 bps to 9600 bps, but the recommendations below 2400 bps are rarely used in the United States. The following CCITT recommendations define modems ranging in speed from 2400 bps to 9600 bps.

CCITT V.22 bis • defines 2400-/1200-bps modem standard for full-duplex, switched network operation or over a 2-wire leased line using a frequency division technique.

CCITT V.26 • defines 2400-bps modem standard for full-duplex operation over a 4-wire point-to-point or multipoint leased line; identical to AT&T 201 modem except that only 4-wire leased line operation is specified.

CCITT V.26 bis • defines 2400-/1200-bps half-duplex modem standard for switched network operation.

CCITT V.26 ter • defines 4800-/2400-bps full-duplex standard for switched network using an echo-cancellation technique.

CCITT V.27 • defines 4800-bps modem standard with manual equalizer for half- or full-duplex operation over a 4-wire leased line; similar to AT&T 208 modem specification.

CCITT V.27 bis • defines 4800-bps modem standard with automatic adaptive equalizer for half- or full-duplex operation over a 4-wire leased line or half-duplex operation over a 2-wire leased line.

CCITT V.27 ter • defines 4800-/2400-bps modem standard with automatic adaptive equalizer for switched network half-duplex operation.

CCITT V.29 • defines 9600-bps modem standard with automatic adaptive equalizer for half- or full-duplex operation over a 4-wire leased line; defines point-to-point operation; although half-duplex operation is possible with V.29, the long training delay specified (250 milliseconds) makes half-duplex operation undesirable.

CCITT V.32 • defines 9600-bps full-duplex standard for switched network using full echo cancellation.

Modems & Multiplexers

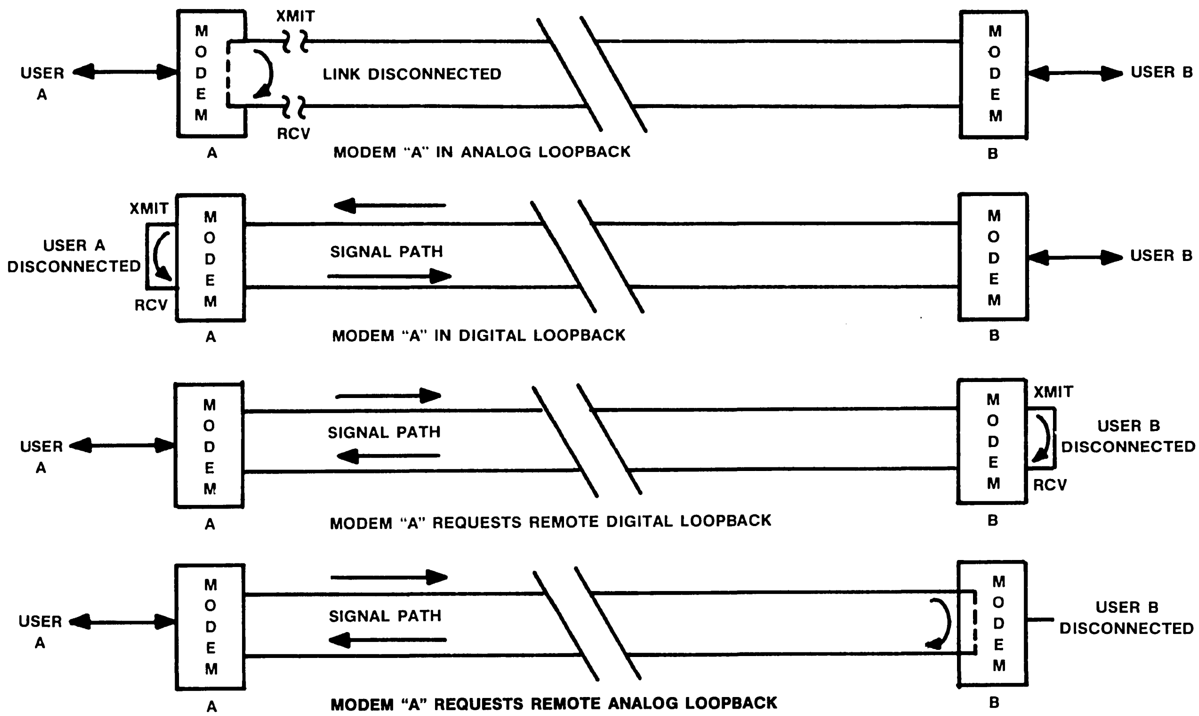


Figure 6 • analog and digital loopback.

■ MODEMS & DATA LINK TESTING

There are many separate elements in the digital path between sender and receiver, any of which can fail and render the path inoperable or degrade performance and cause errors. Some of the elements are owned by the user, some are perhaps purchased or leased from a communication equipment supplier, and some are leased from the telephone company. When a subtle failure occurs, users are often caught in finger-pointing and disclaimers of some or all of the salespersons, installers, and business agents contacted to solve the problem. The solution for many users is the application of diagnostic procedures designed for fault isolation **before** involving outside agencies.

There are several categories of diagnostic tests and procedures a user can apply for fault isolation.

Interface Tests • These tests consist of placing a unit called an EIA interface monitor or **breakout box** between the modem and the computer or terminal. This allows the user to inspect the state of each of the control and data lines in the interface. Often this test will uncover problems with the device itself or the cable by showing that key signals are not being sent to the modem or received at the device. The interface monitor must be matched to the physical interface; RS-232C, RS-449, V.24, etc.

Loopback Tests • Most modems have provision for **looping back** or returning the transmitted signal so that the modem and path can be checked. Loopback may be analog or digital, and digital loopback may be local or remote. Analog loopback connects the modem output to its inputs (with appropriate compensation for channel connections, etc) so that the modem itself can be tested. Digital loopback connects the modem's digital output to its digital input, either at the local or remote end of the line. A test message sent to a modem pair whose remote modem is in digital loopback will be echoed to the sender, testing both modems and the path itself. Some modems will allow the loopback to be set at the remote station under control of the local interface. CCITT V.54 defines a standard procedure for the use of interface control leads to place a modem in loopback and to determine that the modem is in test mode. Circuit 140 (pin 21 on the cable) requests remote loopback, circuit 141 (pin 18) requests local loopback, and circuit 142 (pin 25) indicates that the circuit is in test mode. **Figure 6** shows a diagram of analog and digital loopback, and an example of the use of remote loopback for path testing.

External Digital Testing • The communication line and modem can be externally tested by introducing a data pattern generated by a special device and looping it back so that the pattern is received and checked by the device itself. Devices used for this are called **Bit Error Rate Testers (BERT)**. These devices are available in very small packages and can be used on the spot to verify the quality of a digital connection. Some modems incorporate test pattern generators and comparators for diagnostic testing. Often both random patterns and sequences of specific characters can be generated. Note that most digital error testing equipment is applicable only to asynchronous connections.

Analog Testing • Both lines and modems can be tested at the analog level using signal level and quality tests of the analog signal. These tests provide a measure of the attenuation, signal-to-noise ratio, and other analog characteristics.

Eye Pattern Tests • **Figure 7** presents what is popularly called an **eye pattern** display. This display shows dynamically the **modulation points** associated with each bit pattern. For example, if eight-point modulation is used, representing groups of three bits as a single phase relationship, there will be eight points to the eye. If the eye pattern of a connection is monitored during data transmission, the characteristics of the points will indicate conditions relating to line quality. The presence of random noise on the line causes the points to scatter in all directions into a larger and indistinct circle. Amplitude distortion causes the points to "string out" in an oval whose long dimension points to the center of the "eye." Phase jitter is indicated by the opposite condition, a smearing of the pattern toward the points on either side.

■ MODEM FEATURES

Modems of all speeds and types are available with one or more

Modems & Multiplexers

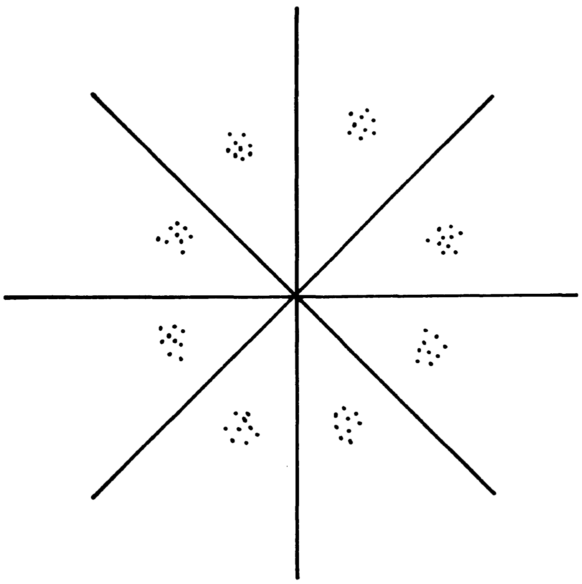


Figure 7 • modem eye pattern.

useful features that may benefit specific applications. Many high-speed modems, for example, offer an integral time-division multiplexer capability. Here are some modem features that should be evaluated as a part of the modem purchase:

Reverse & Secondary Channels • Many medium-speed modems offer a reverse channel, a low-speed path from receiver to sender that may be used to carry flow control commands, acknowledgements, etc. A similar feature available on full-duplex high-speed modems is called **secondary channel** and is sometimes used with modems having built-in diagnostic capabilities to carry information and commands from the master network control point to the remote modems. Reverse/secondary channel usage may require manipulation of a special set of interface control signals, so be sure that the terminals or host computer can support the feature before buying.

Network Control • Many vendors of modems operating at 4800 bps or more over leased lines offer an option of network control which uses a secondary channel of the modem pair to send and receive commands and information between the remote modem and a local network controller. Excessive downtime is a major concern to communication users and the problem has grown as networks expand and gain complexity. Network control through modem monitoring and control offers users a means of reducing downtime by detecting deterioration of lines and by supporting both remote testing of circuits and reconfiguration around failures. Vendors such as Codex, General DataComm, Intertel, Kinex, Paradyne, and Racal-Milgo are producing modems that either contain or can be fitted to support network diagnostic and control modules that link the modems with the vendor's own control system. There is a limited number of control systems that support the equipment of one or more other vendors. AT&T-IS's Dataphone II Service and associated modems provide such network control, and IBM 3860 modems can be used with special application software for network control applications.

Automatic Dialing • This feature, sometimes called **auto-dial**, allows the user to dial out from the computer under program control. The auto-dial feature may be integral with the modem or via a separate unit called an **automatic calling unit (ACU)**. The AT&T-IS ACU is the 801, and the interface between the 801 and computer is defined by RS-366. The CCITT ACU standard is defined by CCITT V.25.

Alternate Voice/Data • If a modem is not capable of automatic dialing, a handset must be used to place a call. Even when auto-dial is available it is sometimes desirable to have operator voice communication end-to-end to verify procedures prior to transmission of data. The alternate voice/data feature provides a handset, dialer, and switch selection of voice or data mode. The use of this feature eliminates extra phone charges for an instrument.

Auto-Answer • Modems expected to receive calls that do not require voice setup procedures per call may use an automatic answer feature to eliminate the need for operator intervention at unattended sites. The modem, on receiving an incoming call, raises an interface signal known as **Ring** and waits for the computer or device to assert the **Data Terminal Ready (DTR)** interface signal, causing the call to be answered. This is required at unattended sites.

Dial Backup • This feature permits one or two dial-up connections to substitute for a leased line connection in the case of remedial recovery from line failure. The central site dials the backup calls, and the remote modem automatically answers the calls and switches to the DDD lines when available. Some units will return to the dedicated line when that line is operational. Dial backup is one of the features controlled by the network/modem control systems.

Error-Correcting Modems • Some vendors are packaging modems with error detection and correction for error-free transmission between modem pairs, a function previously associated with statistical multiplexers, standalone "black boxes," and other sophisticated equipment. Generally designed to operate with asynchronous terminals, error-correcting modems arrange the data stream into a synchronous format replete with data-link protocol and block-check algorithm. When an errored block of data is detected, the receiving modem sends an automatic retransmission request (ARQ) to the originating modem, which then resends the block of data.

■ MULTIPLEXERS & BANDSPLITTERS

Studies of communication between host computers and terminals have shown that in general, the communication circuit is very lightly utilized, as shown in **Figure 8**. An example of this is an interactive terminal used for data entry. The communication path must exist even during the relatively long interval between keyed characters and the much longer intervals required for reference to source documents or other "think time" interruptions. When terminals were rare and the probability of having several in one location was small, there was little that could be done with the idle line time. Modern business, employing a large number of terminals often located in concentrations at branch offices or

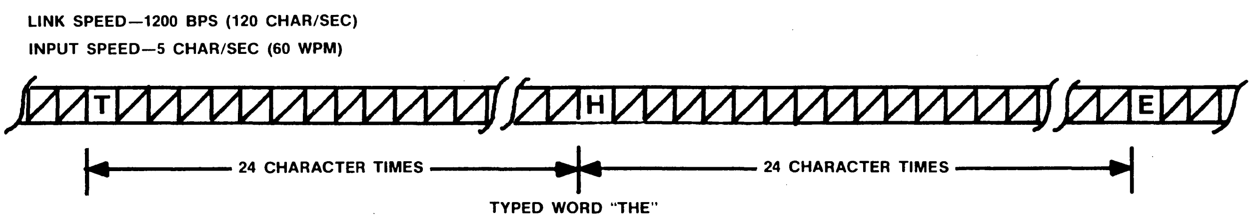


Figure 8 • example of relative line utilization of unbuffered terminal.

Modems & Multiplexers

depots, could in theory share a single line among several terminals. This arrangement not only reduces direct line charges, but it reduces the number of modems required and thus cuts equipment costs as well. One way to share lines is to group terminals into clusters around a single communication controller, the technique used by IBM's ubiquitous 3270 display system. Users with inexpensive asynchronous terminals may be unenthusiastic about trading in existing equipment for the expensive 3270 devices. Besides, such terminals are not designed to operate with cluster controllers. Furthermore, cluster controllers typically use a more advanced protocol such as bisync or SDLC. These protocols are more expensive to support at the host computer level, and an upgrade there adds further expense to a changeover.

The solution to the problem is a device that creates the **appearance** of a group of dedicated lines from a single shared path by giving each of the connected devices a share of the line capacity in a way which is transparent to all the communicating parties. These devices are called **multiplexers** (muxes) or **bandsplitters** and are, next to modems, the most widely used communication hardware today.

■ MULTIPLEXING REDUCES COMMUNICATION COSTS

The benefit of multiplexers stems from simple economics. If a group of 8 terminals are served by dedicated lines from a single host computer, 8 lines and 16 modems are required. A single multiplexer can serve the same terminal population over a single line and a single pair of modems. Of course, this is a simple configuration. Multiplexers can also be used to create complex networks with **tree** structures, alternate paths, multidrop lines, and other complex communication elements; all to best achieve the benefit of reduced operating costs. Large **hub** multiplexers can be used to collect information from nearby users or other multiplexers and concentrate it onto high-speed trunk lines to other nodes, providing multiple data paths or routing through other "hub" multiplexers to bypass points of failure. Terminals with different speeds, protocols, and synchronization may share the same lines, further cutting costs by eliminating parallel lines required only to serve devices with these transmission differences. Asynchronous terminals can use high-speed paths and operate at 9600 bps rather than 1200 bps which asynchronous, full-duplex, modems normally dictate.

Elimination of multiple lines and modems is only one source of savings. High-speed lines are normally not priced in proportion to their capacity; a single 9600-bps link and its associated modems are **not twice the cost** of an equivalent link at 4800 bps. Thus, a technology which allows two slower lines to be combined, even using a fixed division of the line resources, may save enough to pay for itself in a surprisingly short time.

Multiplexers are available in a wide range of channel capacities and resource allocation strategies to satisfy small-, medium-, and large-scale network requirements. Models that support as few as two terminals (normally combined with a modem) can be used by a small communication user, while those capable of creating a multinodal network of thousands of terminals may answer the needs of the largest users. All multiplexers may have added features such as integral modems, line utilization statistics, and additional user ports. The specific features and expansion potential of a multiplexer depends not only on the manufacturer, but on the specific multiplexer technology employed.

■ MULTIPLEXER TECHNOLOGY

Multiplexers combine several user lines into a single line by one of two basic techniques, frequency division and time division. Time division multiplexing is further divided into fixed-allocation systems and variable or statistical allocation systems. Fixed resource allocation has the advantage of low cost—no expensive components are required for decision making or data storage. Statistical allocation of resources is more expensive, but it permits a more efficient use of the expensive transmission facilities.

Frequency Division Multiplexing ● The earliest and least sophisticated method of multiplexing is known as **frequency division multiplexing (FDM)** because each communicating pair of users is given a separate range of frequencies within the link bandwidth. The modulation scheme of each user pair may utilize only the frequency range assigned. The result, shown in **Figure 9**, is similar to the scheme used to provide full-duplex communication on a single voice channel with AT&T 103 modems, but in theory FDM could use two-wire or four-wire circuits. Each of the combined circuits has a range of frequencies separated from other channels by a **guard band** or unassigned range of frequencies allocated to minimize cross-user interference.

The maximum number of users who can be supported on an FDM system is a function of the **total bandwidth** of the composite link and the data rate of individual users. This is because the amount of bandwidth required for a single user is proportional to the speed at which that user wishes to exchange data. Because voice-grade lines are the most popular transmission medium and

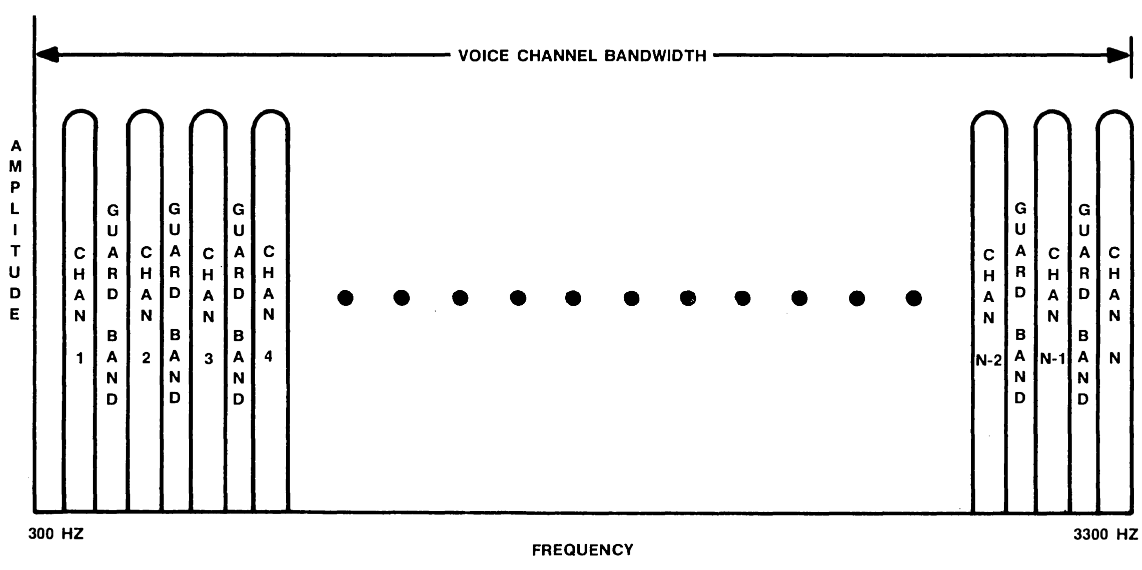


Figure 9 • FDM multiplexing technique.

Modems & Multiplexers

are limited to about 3000 hertz (3K Hz) of bandwidth, FDM techniques are essentially useless above 1200 bps and valuable generally only at 300 bps and below. This limitation on speed tends to overshadow the advantages of FDMs: the elimination of the modem (the FDM itself is a modem), the ability to mix voice and data traffic, and the ability to "drop" off one channel at a given location while passing the rest of the channels through. Frequency division multiplexing is rarely used today.

Time Division Multiplexing • Another way of separating users of a shared link is to give each user access to the link for a small **time slice** according to a rotation scheme. Thus, if a **time division multiplexer (TDM)** is used to combine four channels, each of the channels is given a time slice, one slice out of four.

The sampling rate may be rapid, interleaving the bits from one channel with those of another, or slower, so that only characters are interleaved. Most TDMs are **character-interleaved**, so that a single **frame** on the composite link consists of a character from each channel in sequence. If a channel does not have a character to send, its slot must remain empty so that no confusion on the separation of data and routing to the destination user occurs at the opposite end of the composite link.

Figure 10 shows the operation of a character-interleaved time division multiplexer. Each of the four users is sending data at a rate of one quarter the speed of the composite link. This restriction on the input speed is required since the multiplexer must be able to remove data from the originator end and place it onto the composite link as fast as the data arrives. The input (channel) side of the multiplexer scans each user in turn, taking either a byte of data or a signal that no data exists (usually a character made up of all binary "ones," the marking or idle state of the line). The data is framed and sent to the output (composite link) side of the multiplexer where a similar process extracts each of the input channels and distributes the data to the corresponding output channel. Bit interleaving results in reduced delay on the link and is useful for synchronous transmission, but not asynchronous transmission because it requires sending start and stop bits for asynchronous characters—a waste of channel capacity. Character interleaving increases multiplexer delay because it requires data received from the input (channel) side to be held (buffered) until a complete character is formed, but eliminates sending start and stop bits, increasing the capacity of the channel by about 20 percent.

TDM technology saves money by taking advantage of the fact that line costs do not rise as rapidly as speed increases, thus two slow links are more than twice the cost of a single fast one. It is still a beneficial technology where delay must be minimized and where the input channels are very active so that few time slots are wasted. Most communication circuits have considerable idle time, and the fact that TDM or FDM technology could not address a way to utilize it caused designers to explore alternative methods of allocating time as the declining cost of microprocessors and semiconductor memory made these technologies feasible for multiplexer applications.

Statistical Multiplexers • The result of the design effort was a time division multiplexer which allocates an input channel space on the composite link **only** if that channel has data to send. This scheme can be compared to the classical queueing problems of statistics where a single bank teller served multiple lines. If a line is empty, the teller need not wait without customers; one will be taken from an active line. The technique is called **statistical time division multiplexing (STDM)** but the devices are normally called "stat muxes." Statistical multiplexers offer a higher utilization of the composite link. They also use their microprocessor intelligence to perform error recovery on the link frames, giving a degree of error protection to asynchronous devices that do not normally have such capabilities.

The differences between STDM and TDM technologies are profound in their effects. Because the stat mux has both a microprocessor and memory for buffering, or holding, data it can accept information from a user even if it has no space on the line to allocate (analogous to queueing at the bank). Proper application of statistical multiplexers dictates that the **total** amount of information collected within, for example, a period of a few seconds must be within the data capacity of the composite link. This allows stat muxes to take advantage of the bursty nature of most data communication. If user "A" is sending and there is no time to allocate to that information on the composite link, it will be collected in a buffer with the knowledge that user "A" will probably not send for long and it will be possible to get around to the data a little later when other channels are idle. In **TDM/FDM** systems, the total speed of the input channels could not exceed that of the composite link. With statistical multiplexers, the **average** speed of the input channels cannot exceed the speed of the composite link. The length of the "average" period depends on the buffer capacity of the multiplexer. **Figure 11** shows the operation of a statistical multiplexer.

Buffers & Flow Control • The queue at the bank teller and the buffers of a statistical multiplexer have the same function-storage of work when the **output process cannot keep pace with the input**. In fixed allocation schemes like FDM and TDM the problem was eliminated by requiring that the output resource be sufficient to handle the total input, as expensive as being sure that enough tellers were on duty to handle the peak period without waiting. In statistical multiplexers, semiconductor memory is used to store (buffer) data which has been received but not yet processed. The utilization of this memory is important to the efficient operation of the multiplexer, and is often called **buffer management**.

The first rule in buffer management is that buffers cannot be permitted to fill up. If all the memory available for data storage is used, new information has no place to go and will be lost, causing inconvenience and possible real problems for the communicating parties. Buffer overflow, as it is called, can be reduced or prevented by two means—proper allocation and flow control. Buffer allocation means the division of memory available among channels active to minimize problems. Some multiplexers

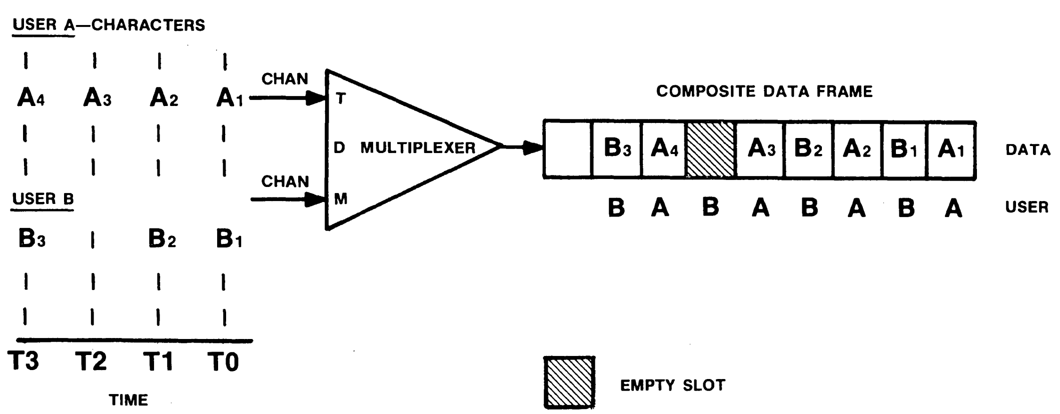


Figure 10 • character interleaved TDM operation.

Modems & Multiplexers

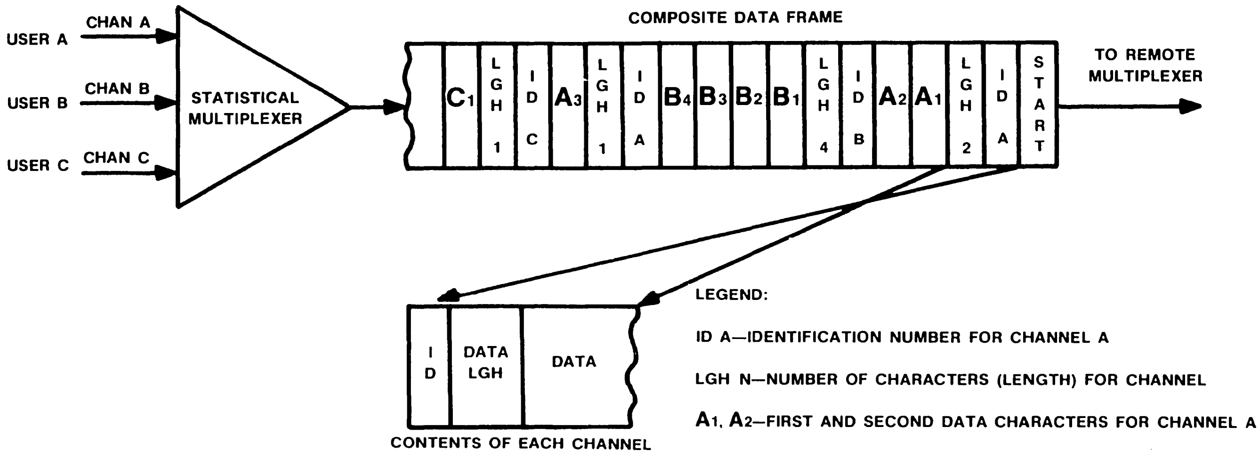


Figure 11 • statistical multiplexer operation.

have a **single pool** of buffer space for which all channels compete equally. This can allow a single channel which sends data very fast, a computer-to-printer link for example, to acquire nearly all the buffer space and cause other channels to fail. Other schemes assign each channel a fixed amount of space, preventing the overrun from affecting other users but possibly wasting buffer space on a slow channel while a fast one is losing data due to insufficiency. If buffer allocation is the only way in which a multiplexer treats the overflow problem, the preferred solution is to support user-defined buffer allocation so that channels with no need for extensive buffering can contribute their normal share of the total memory available to those who do have a need.

Even user-defined buffer allocation is not a perfect solution to the buffer problem. If there is a temporary outage on the communication path which causes a delay in sending data, it is possible that data will back up beyond any reasonable buffering level and be lost. The solution to this is a means of regulating the information flow into the multiplexer; this is called **flow control**. Many user devices have the ability to recognize a signal to stop sending data and another to resume transmission, some can generate such signals, and some can do both. Multiplexers designed to simulate a dedicated path from user to user will faithfully pass such signals, but would not normally expect to generate them. The flow control option of statistical multiplexers allows the unit itself to generate a flow control request when there is danger of buffer overflow on a specific channel, or on all channels of a **pool system**. Flow control may be accomplished by sending a special character (normally called an X-OFF) to the user device, or by changing the state of the RS-232C interface lead called Clear-To-Send (CTS). Normal data flow is requested by sending an "X-ON" character or restoring the CTS signal. Both the exercise of flow control and the resumption of data flow are normally controlled by the level of buffer utilization. Flow control is initiated if the buffers reach a high percentage of utilization, typically 80 percent, and removed when utilization falls below a lower threshold, typically 60 percent or less. If the user device ignores the flow control attempt, the buffers will eventually overflow and data will be lost.

Multiplexers may also respond to flow control. If a user printing device is connected through a multiplexer to a computer, a line error may cause a large block of data to be collected and sent as a unit when the error is corrected. This may cause the destination device to lose data unless **it** flow controls the multiplexer. Both types of flow control are essential in statistical multiplexer applications which have a high ratio of input speeds to composite link speeds, called the **overbooking** or **compaction ratio**, especially where printing devices or other devices that do not involve a dialog exchange are connected.

If flow control is available on the multiplexer, it is desirable to have **individual** channel buffers rather than a buffer **pool**. This will prevent all users from being flow controlled at the same time due to high utilization of a single user. If flow control is **not** available, the pool method will result in the least chance of data loss on any channel by allocating buffer space where it is needed.

Link Protocols In Statistical Multiplexers • Unlike the TDMs, statistical multiplexers are intelligent devices and communicate across the composite link using a communication protocol. The main advantage of this has been noted—error recovery procedures on the composite link will provide error detection and correction to asynchronous protocols which lack this capability. They also provide a second level of error recovery to a protocol such as IBM BSC (bisync) which already has its own, a doubtful advantage paid for in additional multiplexer delay.

Most multiplexer vendors use a proprietary form of link protocol which employs the same zero-bit insertion and deletion used by the so-called bit-oriented protocols SDLC/SNA and HDLC/X.25 CCITT. This has led many of the vendors to advertise that their protocols are "similar to X.25 Level II," or "HDLC-like," or "a form of SDLC." In fact, very few multiplexers have a link protocol which is at all like any of the standard bit protocols, and even fewer have one which is actually compatible with X.25 or SNA. The claim has led users to plan to connect a multiplexer to an SDLC or X.25 host computer, saving a local multiplexer and asynchronous computer ports. **This rarely works.** Users who desire to connect a multiplexer directly to a host or front-end processor should insist on seeing such a configuration in operation. This is especially true of X.25 claims, since even Level II compatibility would **not** provide access to public networks or most host X.25 packages. **Few of the current multiplexer vendors presently have a link protocol which is X.25-compatible.**

Statistical multiplexers that **do conform** to CCITT X.25 Level III are called **PADs (Packet Assembler/Disassemblers)** because they **packetize** asynchronous data into a synchronous format suitable for transmission over an X.25 packet network. Some PADs will also support input for IBM BSC, SDLC, and other synchronous terminals, but vendor-specific protocols are not supported by the CCITT. Parameters for configuring asynchronous devices and interfacing them to PAD and other DTE equipment are specified in **CCITT Recommendations X.3, X.28, and X.29** (see Technology Report 400-700, **X.25 & Packet Network Concepts and Applications**). Another recommendation, **CCITT X.121** address coding, specifies an address sequence for device interconnection to foreign X.25 packet-switched networks.

Users must realize that PAD multiplexers from different vendors vary greatly in their degree of support for such features as permanent and switched virtual circuits, CCITT PAD recommendations, and common stat mux options. Some models do not support call answering, and require X.25 software installed at the user's host computer to perform software demultiplexing/depacketizing. On the other hand, some vendors provide lavish

Modems & Multiplexers

functionality for their PADs, going beyond the CCITT recommendations to support a variety of equipment pseudo-standards and user applications. Adoption of revised or future PAD standards, however, could make this equipment incompatible with the mainstream of packet technology. Finally, PADs destined for connection to public data networks such as Telenet, Tymnet, or Uninet must be **tested and certified** for proper operation by the appropriate network vendor.

Diagnostics • The ability to run diagnostic tests on multiplexers, the composite data path, and the associated remote equipment supports user procedures to isolate failures, reducing downtime. Because multiplexers are typically microprocessor-controlled, they offer greater opportunity for sophisticated performance monitoring and diagnostic testing. The monitoring capability is often overlooked as a source of diagnostic information, but may give advance warning of circuit or equipment degradation, and permit scheduling preventive maintenance at an off-hour period.

Status Reporting • Multiplexer performance monitoring varies according to the size of the unit and the vendor. The four-channel multiplexers and modem/bandsplitters may have little or no capability to collect information on the line or report abnormal conditions, while the large hub or nodal units normally offer extensive channel statistics gathering, reporting of transient or permanent errors, and control of reporting thresholds. Users should look for these features in multiplexer performance monitoring:

- **Event reporting** of any unusual conditions such as line failures, buffer overflow, or configuration errors.
- **Statistics** on line errors and buffer levels and utilization, so that the performance of the line of the system as a whole can be monitored.
- **Supervisory console support**, so that reports can be produced on a CRT or printer rather than via an LED display. This is especially important for multiplexers larger than 4 to 8 lines.

Diagnostic Testing • Multiplexer diagnostics provide the ability to run tests on the circuits in a multiplexer environment from a central point. On very inexpensive 4-channel units these test capabilities may be unavailable or limited to a "go/no-go" test of the path to and functionality of the remote unit. More sophisticated units will allow a form of **Bit Error Rate Testing (BERT)** via a pseudo-random bit pattern or "barber pole" character test pattern. This can be done by placing the remote multiplexer channel in loopback and initiating the test via the supervisory console or maintenance panel. The returned data is tested for errors, and the resultant bit error rate indicates a level of link performance. This type of test can be used to spot degrading circuits, either the composite link or the individual user "tail circuits" to or from the multiplexer. Many multiplexers will also permit the user to measure the path delay across the composite link. This is very useful in conjunction with the buffer levels of the multiplexer to evaluate the performance degradation from loading. If the path delay test is run when the buffer levels are relatively low, the delay should be minimal because competition for the composite link is low. At high buffer levels indicating inbound data rates are temporarily exceeding the composite data rate, the delay of the multiplexer will be higher. Its exact value is probably less important than the trend in values. A multiplexer user who normally experiences path delays of, for example, 100 milliseconds and finds that at the same time of day delays now average 300 milliseconds should look to the line and equipment for a reason. Is there a high link error rate, effectively reducing the data capacity of the composite line? Is a terminal "streaming" data? Some multiplexers will allow the central operator to read the control signals at the remote and local stations to assure that the state of the devices is normal. This can save a costly on-site evaluation of a terminal whose power is off or whose cable is unplugged.

It is important to realize that multiplexer diagnostics are limited to the portion of the network within control of the multiplexer. Loopback testing, initiated by the multiplexer, can loop the local or remote **multiplexer port**, but rarely can the multiplexer control remote tail circuits with their own pair of modems. This is true even if the modems are integral to the multiplexer. Some vendors offer a degree of integrated network management with their own modems and multiplexers, but this uses a separate network control system. Multiplexers can also adversely effect the ability of modems with network control features to operate because they interrupt the path. Users who either have or wish to have network control modems should evaluate their use with multiplexers before purchasing.

Most multiplexer diagnostics apply to the user channels rather than the composite link directly. Users can test the high-speed path independent of the multiplexers by using BERT testers, analog line testers, or modem testers. They can also loop back several remote channels and compare the results. If a high error rate exists on several channels, the composite link should be suspected while a high rate on only one channel would indicate a problem with that channel's path. Where line error figures on the composite link are available, they provide the best and earliest indication of a problem with the combined path. If these indications are positive, more complex bit-error rate testing or component testing can be conducted.

A Simple Trouble-Shooting Procedure • Point-to-point multiplexer systems can be tested by the user for any but the most subtle of problems. Assume that a user has reported a problem with the system and the operator or technician is to check it out. The following procedures should be followed:

1. Check the indicator lamps on the local multiplexer to insure that all external indicators are in their proper state. Be particularly sure to note any indicators that signal the composite link is down (usually called "link active" or something similar) or that the unit is in test mode (in which case it will not handle user data). If any of the indicators are not in their expected state, refer to the manufacturer's manuals for indications of conditions which could cause the problem.

2. If the multiplexer "logs" or stores events or error reports, check the supervisory console or event log source to see if any unusual conditions have been reported. Some units will report line errors, initialization due to power failure, etc. If an entire multiplexer initializes because of a power failure, it may leave control signals or flow control status of the connected users in any improper state, requiring initialization of the user devices as well.

3. If possible, check the buffer levels and flow control status of the multiplexer. If all channels are operating with buffers full and are flow-controlled, it indicates either a link error which has caused data to collect or that a single user has been "streaming" data into the multiplexer either because of an error or an excessive data rate. If possible, check the remote multiplexer buffer levels. A link failure will make this check impossible because it isolates the remote unit. If the remote buffer levels are also very high, it indicates a high error rate on the composite link. If they are low, or at a normal level, it probably means that a single device has introduced too much data.

4. If a composite link condition is suspected, check the modem indicators. Some modems will give an indication of excessive errors, and nearly all will report a complete loss of communication by the state of the "carrier" indicator. In either case, running more complete modem diagnostics according to the manufacturer's recommendations is indicated.

5. If the user channel is suspected, attempt to read the control signals of the channel involved at both the local and remote ends. A printer which "hangs up" and flow controls the multiplexer may cause some models to buffer data until the buffer pool is exhausted. If the control signals are abnormal, take the devices involved out of service temporarily and see if the rest of the system is restored to normal operation. If not, try to get an inspection of the remote device. Is the paper jammed? If no abnormalities can be detected, it may be necessary to initialize one or both of the user devices. Since this will probably result in loss of data, be sure to inform the users to rerun any jobs or check the status of data entered.

6. If a single user is involved in a problem but user-level initialization will not cure it, it may be necessary to initialize the multiplexer. If possible, this should be done on only the channel involved, but many multiplexer vendors initialize only the entire unit. If you cannot isolate the procedure to the user who has failed, try to let the other users gracefully sign off. This may be possible if you disconnect the failed user. If it is not possible, reinitialize the

Modems & Multiplexers

entire unit and attempt to restore use, one user at a time. If the system still fails and no external indications of a problem are detected, the multiplexer itself should be suspected and the manufacturer's test procedures followed.

■ MULTIPLEXER CONFIGURATIONS

Multiplexers can be connected in three basic ways, as shown in **Figure 12**. The simplest and most common configuration is the **point-to-point system**. Where a series of co-located terminals must connect to one or more computers located at another point, the point-to-point multiplexer can eliminate the need for multiple circuits and modems and thus reduce costs. Point-to-point systems can use either TDM or statistical multiplexers, depending on the protocols and utilization. The number of user channels involved in such systems may range from two to 48 or more. Point-to-point systems save in line or modem charges, but introduce a single point of failure which many users find objectionable. Additionally, they primarily benefit users who have a single point of terminal concentration rather than those who have terminals more distributed.

Users who feel that a single point of failure is unacceptable and have a slightly more distributed network of terminals may find the "nodal" configuration of multiplexers the answer. At the simplest level, these systems provide dual links between a single pair of stations. The communication load is shared between the links as long as both are operational, but shifts to the single operating link if one fails. The single link, of course, cannot maintain the same data rate and some additional delay can be expected. This arrangement is also referred to as **load balancing**.

More complex nodal networks can be built from multiplexers which have the ability to handle composite links to other multiplexers as well as direct user channels. A popular low-end version of this system permits a **three-node network** with a single master multiplexer and a pair of slaves. The master unit would typically be located with the computer facility and the slave units at remote terminal locations. This same function could be served by a pair of point-to-point multiplexers, but there are three benefits to the nodal configuration. First, the master multiplexer and two slaves will probably cost less than four separate multiplexers. Second, the master unit can control both slaves from a central point, providing improved network management. Finally, communication between a station on one slave and one on the other can be supported, an important consideration if there are some computer resources located at all three locations.

The ultimate in nodal configurations are those supported by large units often called **network processors** or **network concentrators**. These units allow a user to build a multinode network with such features as dual-link connections with load balancing, alternate routes which can bypass a failed unit, advanced statistical and diagnostic support, user-specified connecting addresses, and protocol conversion. Such units provide some of the economic advantages of packet networks to users of non-packet equipment.

A final multiplexer configuration is the so-called **multidrop multiplexer**. As **Figure 13** shows, this device consists of a master

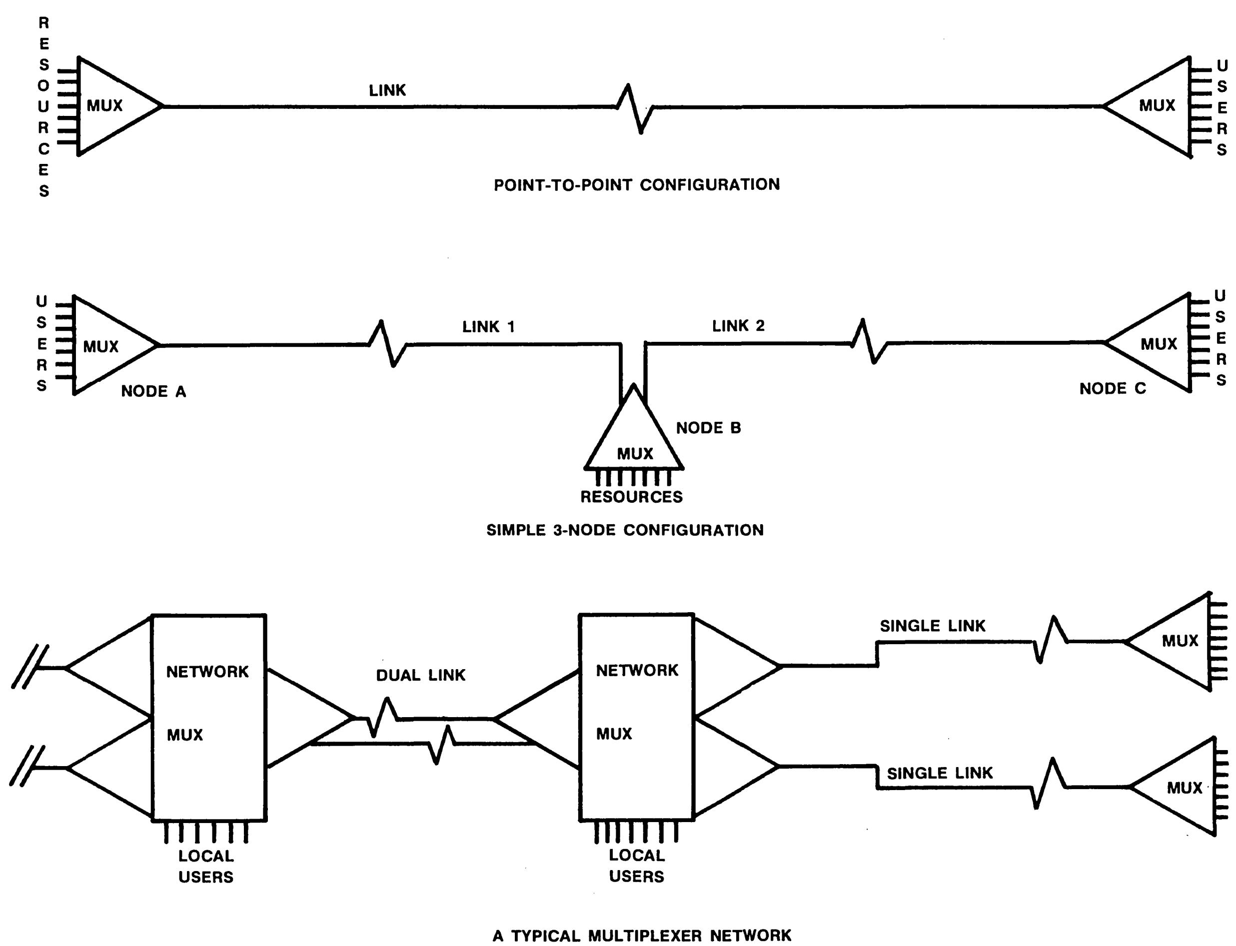


Figure 12 • typical multiplexer configurations.

Modems & Multiplexers

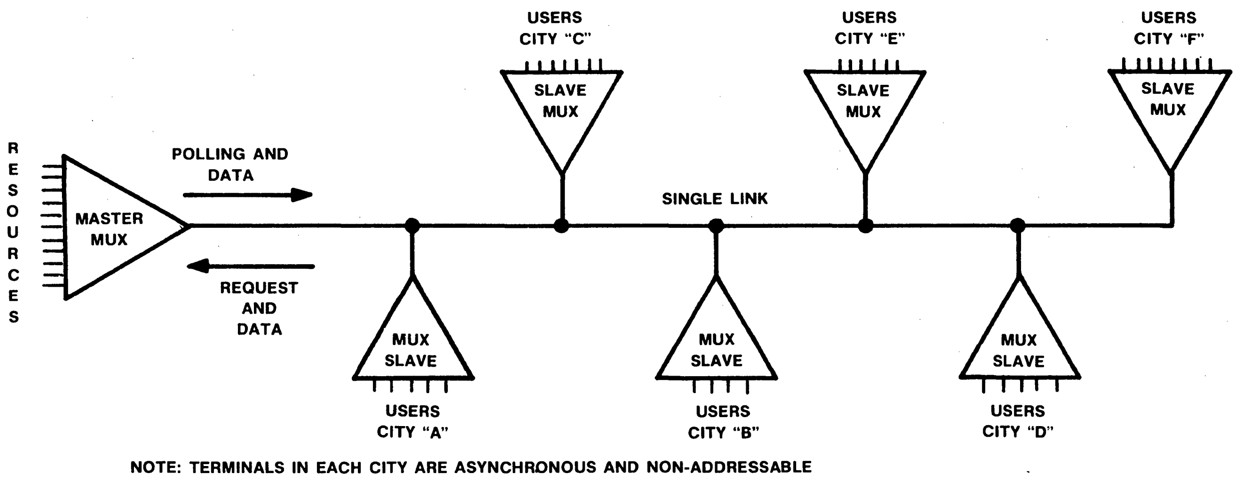


Figure 13 • typical multipoint multiplexer configuration.

unit normally located with the computer installation and a series of "slaves" distributed along a multidrop line. Each slave can support multiple terminals, providing the advantages of polled, multidrop lines and point-to-point multiplexers in concert. In addition, inexpensive and normally non-pollable asynchronous terminals can thus be accommodated in a polled environment. The disadvantages of this type of multiplexer is the relatively long delay which results from combining the normal delays of a multiplexer with the delays associated with polling. Users sometimes find that application programs which use key-stroke editing (backspace, delete, cursor positioning, etc) cannot be handled efficiently because of these delays.

■ STATISTICAL MULTIPLEXER FEATURES

TDM technology supports only point-to-point connections and provides for little feature selection beyond the number and speeds of the input links and the speed of the composite path. This is because most TDMs do not employ extensive microprocessor intelligence and the incremental cost for such features would be high. Statistical multiplexers, having extensively applied microprocessors to the user channels, resource scheduling, and composite link management tasks, are often able to provide special features at no cost other than the programming required. Some of these features are standard equipment at no cost while others are options or associated only with extra-cost models.

Parameterization • On TDMs and inexpensive statistical multiplexers, the data rate, character structure, and other user channel operating parameters are normally set by **plugs**, **DIP switches**, or **front-panel key pads**. While these methods are low in cost and satisfactory for a few channels with constant configuration parameters, they become **unwieldy** and eventually **impossible** as the number of channels and frequency of channel parameter changes increase. Many systems reduce this problem by having the channel information programmed at the master unit and **downline loaded** to the slave at initialization of the multiplexer composite link. Those with a supervisory console option can use this console to change parameters under software control, a much **easier** task than manipulating keypads or DIP switches.

Supervisory console parameter entry, or keypad entry of information, has **no** hardware-readable plugs or switches associated with the settings and hence must **store** the channel configuration in a form of memory. Any power failure or problem would cause parameters stored in normal semiconductor memory (called "RAM") to be lost. One alternative to this is to provide a section of RAM with a **battery backup** power supply. If the power to the unit is removed, the battery pack (if fresh) will power the RAM holding the configuration information for many days. Other systems use a form of Programmable Read-Only Memory (PROM) or Electronically Erasable Programmable Read-Only Memory (EEPROM) to store the parameters.

Automatic Speed Control • This feature, called **Auto-speed**, **Auto-baud**, or **Automatic Baud Rate Recognition (ABR)**, allows the multiplexer to determine the speed, parity, start and stop bits of an asynchronous device and regulate itself accordingly. It typically is used for dial-up connections where the precise terminal speed is unknown. Many host computers provide this feature, and when multiplexers are introduced in such an environment they **must** provide automatic speed settings since the host computer may no longer answer the call. There are two conventions for automatic speed control, the popular **carriage-return** form and the **Memorex form**. All rely on the terminal to transmit a known character so that the speed setting can be made. Most ABR schemes will operate only to 1200 bps because dial-up asynchronous connections above that rate are quite difficult to support.

Priority Control • Applications that combine critical transmission with those of a less critical nature may benefit from a scheme to prioritize certain channels so that they are serviced before others. This is primarily useful for asynchronous data, since placing a low priority on a synchronous channel might cause a timeout and require operator intervention to restore communication. Batch traffic such as printing is often selected for a lower priority than interactive terminal traffic, for example. Vendors such as Timeplex are beginning to respond to the user interest in this area by providing priority scheduling schemes under user control, and others such as Infotron provide automatic prioritization of synchronous traffic in some multiplexer models.

Data Compression • A technique that compresses data transmission into fewer bits or characters without losing information. Common methods include coding redundant characters; coding long ASCII numeric characters into binary; and coding frequently recurring symbols into shorter bit sequences (called **Huffman Coding**). Care must be taken to ensure that data compression techniques match user applications: most data compression algorithms are optimized for either textual or numeric data, and the wrong match has been known to actually slow transmission speeds instead of the opposite, desired effect. Some multiplexer vendors provide data compression on a switch-selectable, channel-by-channel basis while others provide it as a standard feature on all channels only. Users should test data compression techniques on their own data traffic before committing to the equipment.

Alternate Routing • Sometimes called node-bypass, alternate routing is an option on some multilink/multinode multiplexers. Upon a link failure, alternate routing allows the associated channels to be re-routed through an alternate link, bypassing the failed link/node until it again becomes operational.

Modems & Multiplexers

Load Balancing • Load balancing is used in dual-link, point-to-point configurations. Two links share the load of one communication path, usually in dynamic fashion so as to avoid overloading one of the links. The advantages of load balancing are a greater throughput rate than would be available through only one link, and the ability to route all data traffic over one link in case the other fails, but at only **half** the throughput rate. The disadvantage is the cost of maintaining twin data circuits and associated modems.

Split-Channel Speed • Also called **asymmetrical data speed**, this option supports videotex terminals or modems which transmit and receive data at **different** channel speeds. Common receive rates are 75 and 150 bps, while the transmit rate is usually 1200 bps. Applicable to Viewtron, Viewdata, and other public or private-access videotex networks.

Spoofing • Designed to optimize IBM BSC or other bisynchronous protocols for transmission over satellite links, spoofing mimics the control information of these synchronous half-duplex protocols to alleviate unacceptable satellite delay characteristics. BSC requires an acknowledgement from the host upon proper reception of each block of data before another data block can be sent; over satellite links, the delay between sending a block of data and receiving acknowledgement from the host can average between 1 and 2 seconds. Using a spoofing technique, the remote multiplexer acknowledges transmitted data blocks immediately before they are received by the host, while the host multiplexer performs similar control emulations. A block of data containing errors, however, prompts a negative acknowlegement and imposes a critical situation for the spoofing mechanism which must retransmit the errored block in its proper sequence.

Bandsplitter • Some statistical multiplexers permit the composite link to be divided using a TDM scheme. The portions can then be allocated to the use of the main multiplexer as a composite link and to other users who do not want or need the resource allocation and error recovery of a statistical multiplexer. An example of this is a user with asynchronous terminals sharing a location with a single SNA cluster controller. Since the SNA device is operating full-duplex and has a high line utilization, nothing is gained by combining it statistically with the asynchronous traffic and delay may actually be increased. On the other hand, a single high-speed line can be divided into two TDM slots with an integral **bandsplitter** and the SNA device and composite link each given one slot. The two do not compete for resources since each has a fixed allocation of half the speed of the line, and the SNA link is not subject to statistical multiplexer delays due to frame assembly, queueing, or error recovery. Bandsplitters are normally bit-interleaved TDMs that divide channel bandwidth into a fixed number of parts, which can then be distributed between the composite multiplexer link and any users who have high-utilization traffic to pass. The concept is only useful when some of the users at a given location have such high activity that statistical multiplexing of their data would cause excessive delay to other users, since the time slots allocated to one of the bandsplit channels are not available to others even if they are idle.

Bandsplitters are often used on lower priced modems to provide synchronous channel support. Since synchronous data arrives in blocks with no start or stop bits, it is more difficult to process and the hardware components needed for it are more expensive. Statistical multiplexers normally cannot "concentrate" as much synchronous data onto a composite link as asynchronous data, but bandsplitting as a means of sending normal terminal traffic on synchronous devices is inefficient and should be carefully evaluated.

Switching & Port Contention • The relationship between input users and output users in a point-to-point multiplexer is normally fixed in that a given channel on the slave unit will connect to a given channel on the master. In some cases it may be desirable to have fewer multiplexer ports on the computer side of the unit and let a larger terminal population at the other **contend** for these ports. The function performed is similar to that of a telephone rotary in that a terminal, becoming active, attempts to access one of the computer ports in the same way as a dial-up call might **hunt** for a free line. **Port contention** is very useful where a multiplexer is replacing a **large number** of dial-up connections. Users of port contention must define the way in which the multiplexer can detect an attempt to connect and a way in which path disconnection is to be requested, normally the setting of a control line on the RS-232C interface such as Data Terminal Ready (DTR).

A further capability may be offered in nodal multiplexer systems. Since there are multiple nodes and multiple users, it is necessary to define the connectivity of the multiplexer—who is connected to whom. This is normally done via supervisory command, but some

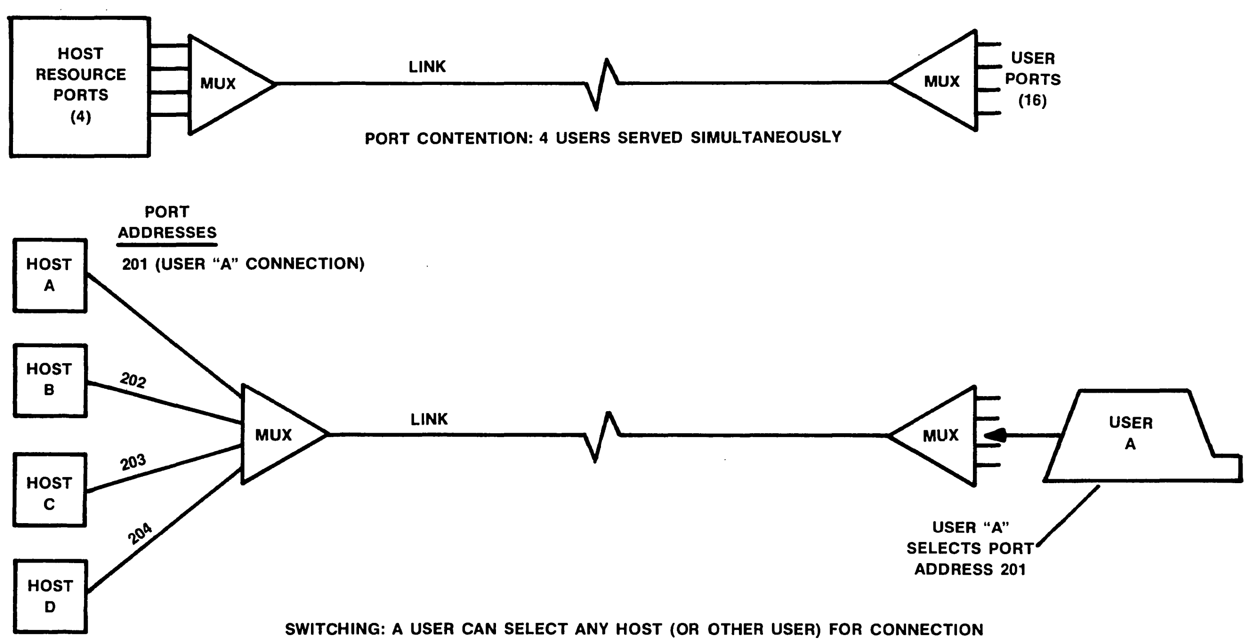


Figure 14 • examples of port contention and port switching.

Modems & Multiplexers

vendors will allow users to **select** parties to whom they connect by entering an address. The multiplexer then performs a function similar to a telephone switch and establishes the path to the requested station. This switching procedure is typically referred to as **channel routing**, and can be performed between local channels as well as between remote channels. **Figure 14** shows both switching and port contention.

Both switching and port contention require a means of signaling connection requests and disconnect requests and for handling a **busy** condition. Switching also requires that the multiplexer support a dialog with the terminal operator to read the address. Since connection is under user control, switching systems often have security features such as **closed user groups**, also called **restricted resource groups**. These permit users and resources to be assigned user group codes which the system will store and recognize. Each resource can be identified with one or more groups of authorized users and the multiplexer switching logic will bar any connection request from a user who is not a member of one of the authorized groups. Establishing closed user groups is a supervisory function that can be performed at a supervisory console or another terminal, subject to password access to the facility.

• END

Business Communication Networks— Putting It All Together

Strategies For The Selection & Integration Of Communication Products

■ INTRODUCTION

Although the data communication user of today has access to product selection and evaluation information on thousands of network products, building a network is more than selecting its elements. This article describes the building of a user network, from setting policy to network management and control, with emphasis on how things go together. Host software, transmission facilities, modems, concentrators, multiplexers, and network control systems are described in the context of a business communication environment.

■ THE PICTURE & THE PROBLEMS

If you're a data communication user, an authority no less than AT&T has surveyed your situation and defined the major problems which you face. In its application for the fully separate unregulated subsidiary later known as Information Systems, AT&T listed what it found to be the four major problems facing communication users today: the application orientation of existing networks, the lack of expandability, the lack of control, and the high capital investment required to support new applications.

Application orientation doesn't sound like much of a problem. After all, aren't networks supposed to be related to an application? Not in today's mode of thinking. Most of the users of data communication evaluate each separately and cost justify it either independent of other uses or incrementally with something which has already been done. This often results in the new network either duplicating parts of an existing one or growing out of it without regard for other uses.

Why do users tend to **grow** application networks? According to AT&T, it is because of traditional project costing and an understandable tendency to take the path of least resistance. It's easy to say that modern communication theory favors the development of a communication network as an application-independent corporate resource, but it is rather difficult to force a user with hundreds of terminals to restructure or convert an existing network because it is tied to a single use such as order entry.

The second user problem was the lack of expandability typical of user networks. This again stems from the tendency to plan for a least-cost current network in an environment where future applications may involve business situations which the current planners are not responsible for nor aware of. When your company plans its national inventory control system are the people studying electronic mail for corporate staff present? If you are a typical user, probably not. Thus, the potential for the expansion of the network to handle electronic mail is likely to be missing, and future planners will call your network **inflexible**.

Lack of network management and control facilities in a network often stems from the reliance on host or front-end processor to provide such capabilities. As networks become more complex and cost pressure mounts, new technologies such as multiplexing may be employed to reduce costs. Since multiplexers create network elements that the host computer cannot control, the user loses the central control of his network from the computer site. **Figure 1** shows the effect of adding a multiplexer/concentrator in a network. The new network is more reliable, but some of the paths are **invisible** to the host computer unless they are switched in. These "phantom lines" may be vital in the strategies of fall-back and recovery, yet the host software has no way of monitoring their state. In addition, newer modems have valuable diagnostic and control features that did not exist when the application was designed and, therefore, are not supported by the computer network control software. While all of these enhancements may be used in existing networks, control of the new facilities may be totally separate from that of the earlier and more conventional routes.

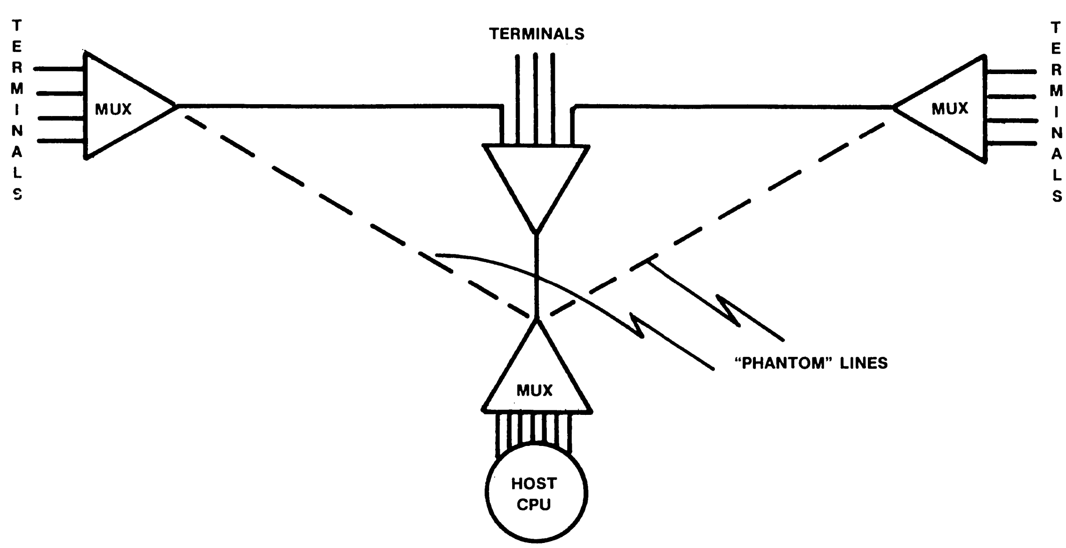


Figure 1 • "phantom lines" hidden from host control.

Business Communication Networks—Putting It All Together

The final problem in AT&T's bag is that of the capital entry threshold for new communication applications. Many users find that good business projects cannot be carried out because the cost of providing the communication facilities is excessive or because the business value of the project cannot be accurately determined. Although this can sometimes be solved by utilizing the facilities of another application, such **piggybacking** is often difficult to sell to the users of the existing facilities, who are concerned about loss of response time, increased failures, and in general of any changes to their comfortable environment.

What are the solutions to these problems? AT&T-IS, of course, says that its Net 1000 service provides them, and to a large extent this is true. Net 1000 is not the only solution, though. The advantage of Net 1000 is that it is a generalized communication resource planned by experts and designed to provide all of the economic and managerial benefits of a network without demanding an excessive capital investment for new users. Any company can provide some or all of these advantages in a private network by careful planning of the network elements. For some, public networks such as Net 1000 may provide the most cost-effective way to communicate. The decision to use public or private facilities is often made very early in the planning cycle, but network problems cannot be solved by looking at the parts of a network, they must be solved as the total network.

■ PLANNING & POLICY IN COMMUNICATION NETWORKS

Data communication is not something that a company uses for its own intrinsic value, it is a tool in the information processing and management policy of the business. One of the first steps in planning communication facilities is thus to define that policy.

You hear and read much about **MIS** or Management Information Systems, office automation, transactional systems, and the like. **Which are you?**

One major element in establishing an **information identity** is classifying the value of information to your business. Some companies use data only as a means of operating; buying, selling, paying, receiving. Data communication may be needed to gather in **transactions** which are the electronic equivalent of paper invoices, checks, etc. Other companies must analyze information in order to do business. Insurance companies, for example, must perform accurate statistical analysis of loss experience in order to set rates. Banks may require an accurate picture of the status of a corporate client in order to approve loans. This type of user may not be able to predict in advance exactly what will be valuable and what will not, and which pieces of data must be correlated in order to support business decisions.

Transactional users place a high value on the timely processing of data since data is a part of their chain of cash receipt and product delivery. Online data entry is popular with users of this type because it improves the clerical productivity, and concepts of office automation or distributed processing are gaining favor here because the structure of information movement is fixed by the applications themselves. A cash receipt may be entered in the accounts receivable system and the general ledger systems but probably not in the inventory system. This fixed flow allows the separation of processing power and its movement closer to the point of transaction—distributed processing. Transactional users, having less reason to gather their data to a central point for processing or correlation, can often employ networks that have no single point of collection—the peer network.

Informational users, on the other hand, view the correlations of pieces of datum as important as their "production" use. They often have very high volumes of data requiring a method of batch entry such as optical scanning or magnetic character recognition, and the cross-matching of information required for generating "ad hoc" reports requires a central storage and processing operation or a distributed intelligence architecture not yet commonly available. These users demand hierarchical networks with remote processing clusters feeding area controllers which in turn feed data to a cluster of host computers.

If the general characteristics of each user type don't help you classify yourself, there are a few key **symptoms**:

- Businesses which obtain a large portion of income from investment of receipts are normally informational in structure.
- Manufacturers are normally transactional users.
- Retail operators with very low margins, such as food stores, are often informational users while those with higher margins are transactional.
- Smaller companies tend to be more transactional in nature while larger ones are more often informational.
- High-technology companies are likely to be more informational.

Once you have identified your classification as a user of information, it should be a key element in your communication planning. Transactional users should generally migrate toward **peer-distributed networks** using host computers and access methods which support real-time transaction processing, the so-called **teleprocessing monitors** such as IBM CICS, for example. Informational users have a need to integrate their data communication and data processing strategies more closely and should investigate **network architectures** such as IBM SNA, Digital Equipment's DECnet, or Honeywell DSA. These will allow a much more flexible use of communication resources.

■ NETWORK ARCHITECTURES & ACCESS METHODS

Most users have little choice in the computer to use for their networks—the one they already have must serve. But within the computer system there may be a choice of communication support packages, and the proper selection here is important because it serves as the basis for the rest of the network. The host software can be divided into the following categories:

- **Access Methods.** These provide a convenient way for the user to interact with the communication hardware. They are normally concerned only with data communication, and provide little or no application or database services.
- **Teleprocessing or Transaction Monitors.** These provide a framework within which a user can develop programs to process single requests for action, or transactions, from a communication line. They include a higher level of communication services that further insulates the application programmer from the technical details of communication lines, and may provide additional services such as database management and screen building.
- **Network Architectures.** These complex packages join user program and communication facilities so that local devices and remote terminals are almost indistinguishable. The application program uses the communication facilities as resources, and the resource is managed by the network architecture itself, independent of its users.

Access methods generally require a program to use the communication line as a device, reading and writing to it to communicate with the users it serves. While the program is loaded, all data from the line is routed there. More sophisticated systems will allow the control of a device by a program, allowing other devices on the same line to be handled by other programs. In monitors or network architectures, the communication software often allows the user to enter a series of requests or commands that are dynamically routed to the particular program possessing them. A single terminal can thus update an accounting file, read the status of an order, and change the address of an employee without requiring a central computer operator to switch the device from one computer, computer port, or program to another.

Selection of host software support requires that you as a user evaluate not only your needs but also your capabilities. Most vendors offer access methods in varying levels of complexity, and the programming environment differs dramatically between a simple access method and an advanced network architecture. In general, the network architectures allow less sophisticated application programmers to work in a communication environment, but will cost more in direct software fees and in configuration of the more complex hardware normally required. On the other hand, the "simple" access method may cost very little, run in a minimum configuration, and be nearly impossible for your staff to learn to use.

If you are a small user, probably transactional in your use of data communication, you will probably find a teleprocessing monitor

Business Communication Networks—Putting It All Together

or transaction package the best choice. These packages, such as IBM's CICS or Informatics' TAPS, will provide your programmers with a consistent communication environment within which applications can be developed. The selection of a specific package should be made on the basis of the features of the monitor, particularly the ease with which it adapts to various protocols. A monitor that handles IBM binary synchronous communication will not serve a user with plans to employ another protocol such as SNA.

The following features should be evaluated in TP monitors:

- **Is it limited to a single vendor, single operating system, or a single access method?** Such limitations are common in the packages provided by hardware vendors, but restrict the growth path of the user. If you are certain you won't change vendors or operating environments, restrictive packages may provide all you need at lower cost.
- **Will it handle the protocols in use or planned?** Many TP monitors use a host access method as a basis and have no restrictions in protocol not inherent in the computers, while others do not support all protocols.
- **Is database support included?** The more complete the operating environment, the more easily it is used by relatively inexperienced programmers. Real-time communication applications are more than just line handling, and if your programmers cannot handle the data communication part they may not be able to handle the real-time data management either. Be particularly careful of file and record lock facilities to prevent collision in accessing the same files from multiple users.
- **Is there a screen builder program?** The design and development of good screen formats for data entry is a major task and has a significant impact on the productivity of the users of the system. Some packages allow programmers to develop a screen from a set of simple application-oriented definitions, and this saves considerable time and debugging.
- **Is there a query language?** The ability to retrieve ad hoc reports from the database is a benefit, particularly to informational users. Transaction users can use the query language or "bread and butter" reports like transaction registers and daily activity logs, saving programming efforts for more important tasks.
- **Are there provisions for security against unauthorized access?** Password protection keeps a stranger from accessing the system, but not necessarily an authorized user from seeing sensitive data in another area. File and program level security or closed user groups are valuable in keeping the wrong people out of your files. If the TP software doesn't provide it, you'll probably end up writing it yourself.

Users with more sophisticated needs will probably find that a total network architecture will serve them best. These extend the services of TP monitors to what can be called the "logical device" concept, where a program can operate with a local user, a remote user, or a data file and not need to know which is which. There are limitations to this, of course, but most of the mainframe and minicomputer vendors are working to provide such facilities for their systems. The international standards work on protocols such as X.25 has advanced to the point where truly user-independent interactions, as defined by the OSI Reference Model, are now possible. Vendors are busily at work developing OSI products, although none are yet available and probably will not be available until 1985.

There are some key points to consider in the selection of a network architecture:

- **Equipment supported.** Since most architectures are still vendor specific it is important to be sure that the equipment which the vendor supplies is capable of handling the entire communication and user application, and that the cost is reasonable. Remember, a network architecture also establishes your operational environment for business programs.
- **Device support.** Will the architecture support the devices you plan to use or already have? Will there be any restrictions in the features provided by the architecture with off-brand or unusual protocols?
- **Network structure.** What type of connection structure does the architecture expect? If terminals cannot directly enter the host computer, the number of communication and cluster controllers may impact costs significantly, especially if the number of terminals at any one location is relatively small. If the architecture does not support host-to-host connections easily, it may be unsuitable for many multihost users.
- **Network management.** Since the users are almost completely isolated from the communication facility, the network architecture must manage this itself. Are the reporting and diagnostic facilities of the system adequate? Can spare lines or alternate routes be utilized? Is it easy to place redundant equipment in the network and switch to it? External network or technical control is generally possible even with network architectures, but non-integrated control is often difficult to coordinate. How do you tell the network architecture's facilities manager that you've switched a line? What do you have to tell the users, if anything?
- **Design direction.** Nearly everything in data communication is evolving toward something, including your own operation and the architecture you select. If they are not going the same place, an eventual migration to another system is required. **With network architectures, you do not change concepts easily.** Explore the ultimate goals of the architecture before committing to it, because the application programs you develop under it will probably be much less transportable than non-communication programs that perform similar functions.

What about access methods? Why leave them to last? Most access methods are primitive-level interfaces to the communication hardware which require an experienced data communication programmer to use. If your environment is such that this type of individual is likely to be available, access method programming may be a reasonable choice. Unless your company is large, it probably is not. But if you decide to use access methods directly, don't let the fact that they are relatively primitive lead you to expect that they are also very flexible.

Access methods can be broadly classified as direct or queued. Direct structures require that the user program issue "reads" and "writes" directly to the communication line, a practice which increases programming complexity and tends to dedicate a given line to a single application. In queued structures, the access method manages the line on your behalf and delivers messages from or to it when requested by the programmer. A "read" or "write" command in such a system actually reads from or writes to a queue of messages. Some systems permit the queueing of data to a disk device if the application program that must receive it is not loaded or temporarily unable to accept data.

Some computer systems bundle access methods and host operating systems together, or offer several operating systems which in turn are supported by several access methods. It's normally best to view these combined products as combined products rather than trying to separate them. Catalog the languages available in each system, the number and type of special-purpose facilities like database management, and the communication implications of each. Don't forget to consider any restrictions on multiprogramming or multitasking which their operating system imposes. In particular, check the number of tasks that can be run simultaneously and the relationship between the number of tasks and the number of lines or terminals.

A final point on the evaluation checklist for any communication software is its ability to adapt to various transmission media. In many of the complex network architectures, access to public data networks or satellite links can be transparent to the user, while others will either not perform with specialized transmission systems or will perform with an unacceptable level of degradation.

■ ANALOG NETWORKS—THE COMMON CARRIERS

Most data communication is still maintained over dedicated or switched telephone circuits, and very few users can plan a network that totally excludes these paths. The term **analog** is applied to these lines because they carry frequencies in the audio range, such as the human voice. Analog lines can be classified according to the range of frequencies they can carry; sub-voice grade cannot carry the human voice properly, voice grade carries the frequencies required for normal conversation. The total span of frequencies carried by a circuit is called its **bandwidth**.

Business Communication Networks—Putting It All Together

Since computers and terminal devices used in data communication output digital data rather than audio or analog data, the output must be converted to analog tones through use of a **modem**. The type of conversion, frequencies used, and various technical features of the modem determine the maximum speed which the modem link can support. Modem speed is restricted by the bandwidth of the channel (the greater the range of frequencies available, the higher the speed) and the noise level on the line.

There are many kinds of modems, and the points of differentiation have to be examined by prospective users or there is a high risk of improper selection. To do this, we must introduce some terminology; half- and full-duplex, dial and leased lines, synchronous and asynchronous, and point-to-point or multipoint.

Analog circuits may be operated as two-way alternate paths, called **half-duplex** or two-way simultaneous paths, called **full-duplex** as shown in **Figure 2**. Half-duplex operation allows data transmission in either direction, but only in one direction at a time. With full-duplex, data can flow in both directions at once. Half-duplex connections are normally sufficient for terminal-to-computer interactions because human operators tend to either read the terminal data or key their own but not both at once. Such read-change-send interactions are sometimes called **logically half-duplex** because simultaneous communication is contrary to the interaction of the two stations regardless of the capabilities of the path.

Full-duplex communication is often required for computer-to-computer exchanges because the high volume of information moved makes it undesirable to restrict a station to either sending only or receiving only, even for a short time. While half-duplex data paths can be supported by a standard dial-up connection (sometimes called **two-wire** because of the number of voice-carrying wires in the circuit), full-duplex communication can be handled over such paths only at low-to-medium speeds, currently, to 4800 bits per second (bps). This is because the bandwidth restrictions on data capacity of the line apply to the total of the data sent at one time in either direction. Full-duplex operation above 4800 bps requires two circuits, one in each direction.

Another characteristic of analog communication is the way in which characters are sent on the circuit. For low-speed applications, each character can be marked with a **start bit** and a **stop bit** and sent whenever it is available. This method is called **asynchronous** because the sender and the receiver do not have to exchange precise information on when data is being sent, called **clocking**. But async, as it is normally called, has a high overhead because of the need for start and stop bits, and at very high speeds it is difficult to keep the modems at each end of a path properly tracking with no data or data in random spurts. A solution to this is to exchange clocking information between the modems so that each station knows exactly when the other is putting data on the line. This is called **synchronous** communication, and nearly all high-speed transfers over analog circuits are synchronous.

Analog circuits can be leased from the phone company or accessed through direct distance dialing (DDD). The relative economics of the two methods depends on the amount of service you need and the distances involved. Here are some rules of thumb in evaluating leased lines versus dialed lines:

• Infrequent connections (less than several hours per day total call time) are almost always more economically handled by dial-up service or a high-volume user program such as WATS (Wide-Area Telephone Service).

• High data rates (over 4800 bps) will nearly always require leased lines.

A type of circuit that always requires leased lines is called **multipoint or multidrop**. This is a phone line that follows a path with "stops" or "drops" along the way, to which terminals or other devices are connected. The line, being shared between users rather than dedicated to a single user, is normally less expensive than point-to-point service for all the stations, but the management of multidrop service is more complex, and there are obviously limitations on the response because of competition for the line. **Figure 3** shows the advantages of multipoint or multidrop lines.

Selecting a combination of these features for your particular environment isn't as complex as it might appear. Here are a few basic rules:

• Check with the equipment vendor first. There may be a recommended (or even required) set of modem/line options for your system. At least such a check will normally reduce the alternatives list.

• Slow terminal devices, such as printers, normally operate on asynchronous circuits at speeds of 1200 bps or less. A printer without a keyboard may operate satisfactorily at half-duplex, but most keyboard devices expect the computer to **echo** the character or it will not print locally. This requires a full-duplex modem.

• High-speed devices such as computers, or terminal systems such as IBM's 3270, require synchronous modems. If the operating speed is over 4800 bps, the modem manufacturer may recommend leased, **conditioned**, lines. The conditioning process assures that the line's performance will be within a tighter tolerance than normal lines, and separate charges are assessed for conditioning depending on the type and the performance desired. If you're selecting a modem, be sure to check the conditioning requirements carefully; it may be more expensive in the short term to buy a modem with high-quality equalization logic that eliminates the need for conditioning, but it may pay for itself in the long term.

• Most multipoint lines and many point-to-point lines are **polled**, which means that a master station must send a request for data to each slave station attached at frequent intervals (several times per second is not uncommon). If half-duplex lines are used, the master modem must **train** to each new slave, and the slave answer requires that the direction of transmission be changed. This process is called **turning the line around**. There is a built-in delay in modems called **clear-to-send delay** which holds up the sending station for a time period to allow the modem to train. This delay is often quite long, approaching a quarter of a second. To

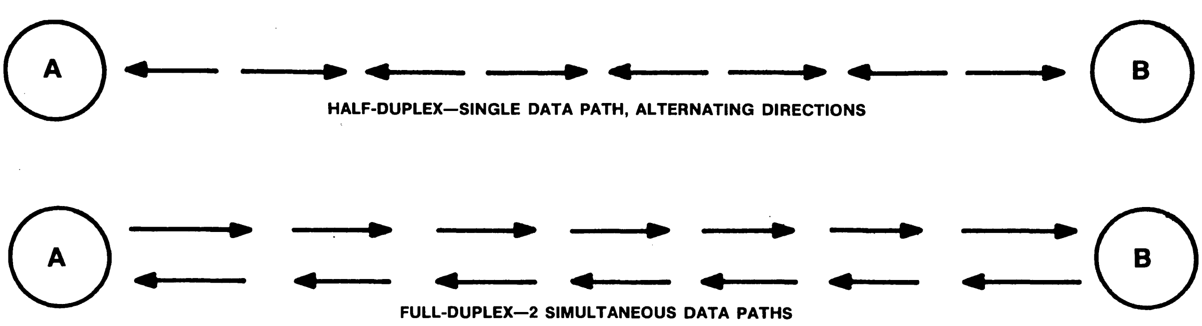


Figure 2 • full- and half-duplex transmission.

Business Communication Networks—Putting It All Together

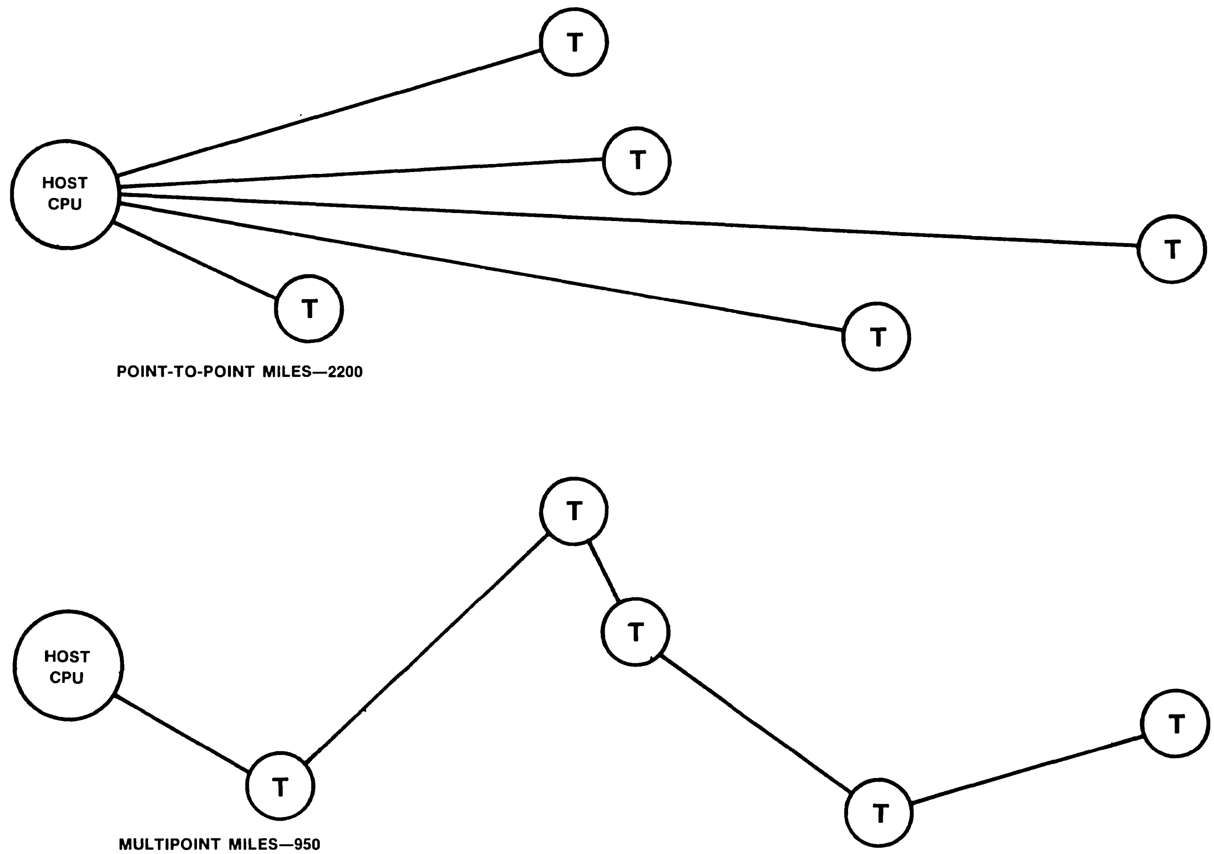


Figure 3 • multipoint line application.

prevent the delay from slowing the polling of the line, special modems called **fast-poll modems** are available which train quickly. Consider these if you have a line with many stations.

• Modem or turnaround delay also affects data capacity of a line. A quarter of a second delay is an entire block time to a batch terminal operating at 9600 bps. When long sequences of data are to be sent, for applications such as remote printing or RJE, it is almost always wise to consider full-duplex service. This will eliminate the turnaround delay and dramatically cut the time required to complete a big print job. This rule applies only to high-performance lines that use error-correcting protocols such as IBM bisync.

• Most users will need to buy modems in pairs, one for each end of the circuit. If one station already has a modem, as would be the case if you were connecting to a timesharing service or service bureau, the modem you purchase **must be compatible** with the existing one. There are two main standards of compatibility, AT&T-IS and CCITT. Many vendors advertise as **103/113 compatible** or **AT&T-IS 209 compatible**. There is a high probability that such units are compatible with AT&T-IS modems, but a lesser chance that they are compatible with each other. The CCITT, an international standards organization, defines modem standards for speeds from 200 bps to 9600 bps, and modems that comply with one of these standards are normally compatible with others that do. It's always a good idea to try any combination of unlike modems first, and as a general rule **if you have to buy two, buy two of the same**.

■ PRIVATE NETWORK DEVICES—MULTIPLEXERS & CONCENTRATORS

Users with heavy communication loads, large numbers of lines, or unusually high reliability requirements may find that even leased line networks are not adequate either from a cost or performance perspective. Point-to-point and multipoint networks may seem to duplicate circuits when many terminals are scattered in a relatively small area but none are co-located. This duplication cannot be avoided within the constraints of point-to-point or multipoint architectures, so a new network structure must be proposed—the nodal network.

Most users tend to think of networks as having the classical **star** structure shown in **Figure 4**. This may be the best network design for a few widely separated devices, but better line economy can almost always be achieved by introducing some points of concentration in the network. Network architectures such as SNA, DECnet, or packet networks achieve this by network elements called **nodes**. These nodes are an active part of the network and cooperate with the host computer or computers to move information efficiently and recover from line failures through alternate routing.

Users without a network architecture can gain some of the advantages of nodal networks by using devices such as multiplexers, network processors, or concentrators. The distinction between these devices has become blurred as vendors offer new features on each, so some definitions of the terms must be provided:

• **Multiplexers** are applied in pairs to collect a series of lines into a single line at one end and fan them out to the same number at the other. Multiplexers do not change the protocol or the number of paths, so their use is largely transparent to the user's application and equipment. Although multiplexers may provide additional services such as switching to backup facilities or alternate routes, the primary purpose of multiplexers is to reduce line costs by combining many relatively lightly utilized paths into a single, more heavily utilized path, thereby eliminating the cost

Business Communication Networks—Putting It All Together

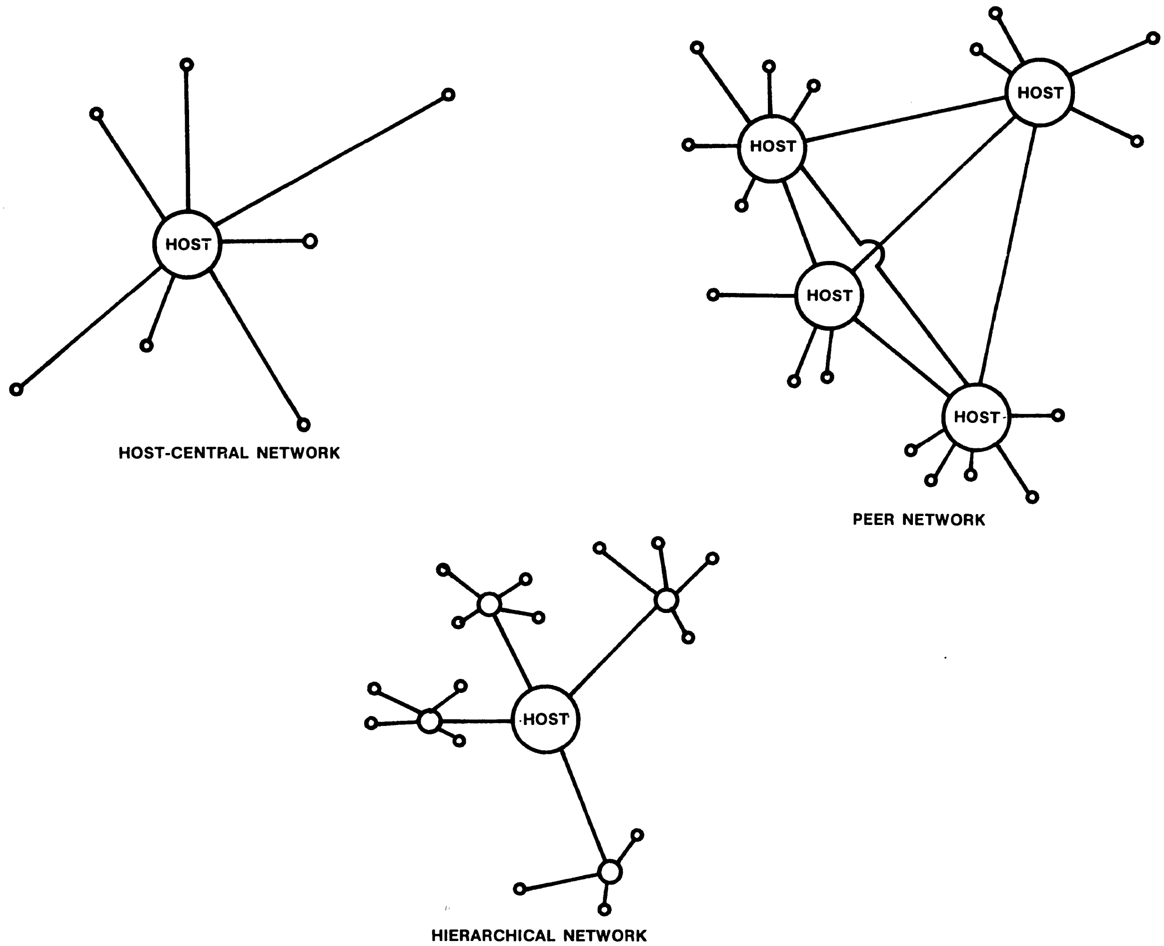


Figure 4 • central, hierarchical, and peer networks.

of additional lines.

• **Concentrators** collect multiple lines into a single line, but do not pair with a partner to split them out again. If the protocol of the concentrated devices permits multiple users on a line, the effect may be only to reduce the number of ports into a computer at the end of the concentrated link. If not, the concentrator must use a multidevice protocol on the concentrated line, and the user at the other end must support that protocol rather than the one used by the individual devices being concentrated.

• **Network processors** are multiplexers or concentrators that add features such as routing data to multiple "trunk" lines and/or protocol conversion. Lines may be run from one network processor to another to provide alternate paths in case of line failure or to balance communication loads. Network processors are more or less transparent to the user, depending on the features selected.

Multiplexers are probably the most popular communication device apart from modems. **Figure 5** shows the effect of placing multiplexers in a user network. The remote stations are grouped into cluster points where a multiplexer collects them and moves them over a single line to the central site. Here other multiplexers restore the original number of lines so that the computer configuration does not change. The effect is a reduction in the number of lines, and therefore, in the cost of the network. As the cost of lines increases and that of multiplexers decreases, the circumstances that justify their use become broader. Some users can actually save money by multiplexing as few as two lines.

There are several types of multiplexers that differ according to the technique used to share the common line. The oldest scheme, called **frequency-division multiplexing (FDM)**, separates users by assigning them different frequencies within the analog bandwidth. This scheme is much too restrictive and is seldom used today. Another method, called **time-division multiplexing (TDM)**, gives each user a time slice on the common line. This slice, normally the space required by a single character, is permanently assigned to the user, and goes empty if the user has no data to fill it when it is offered. The fixed nature of the allocation for both TDM and FDM is unappealing when the use of the lines to be multiplexed is very light. Since typical utilization of terminal links is 10 to 20 percent, both TDM and FDM waste line capacity by granting it to terminals that statistically are unlikely to need it. Fixed allocation of line capacity, such as TDM or FDM provides, is most useful when the lines being collected are already heavily utilized. In this case the cost savings is achieved by taking advantage of the fact that a single circuit of high data capacity (9600 bps, for example) is not twice the cost of two circuits at half that speed. Small TDM devices called **bandsplitters** are often used to combine several slower lines into a single high-speed line. Bandsplitters are usually combined with modems for maximum economy, and may be useful where up to four low-speed circuits with high utilization or stringent response requirements may be combined into a single trunk.

A more efficient concept for multiplexing now used by most

Business Communication Networks—Putting It All Together

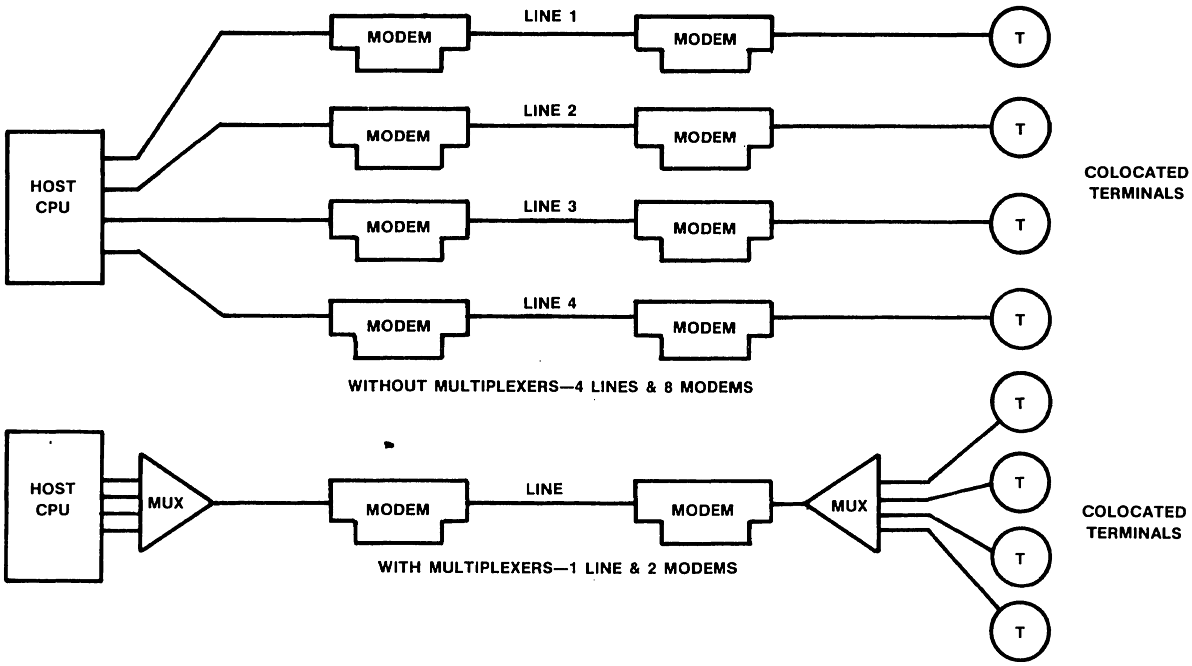


Figure 5 • application of multiplexers.

vendors is called **statistical multiplexing**. This system collects data from the incoming lines and inserts the data **and an identifier of the source line** onto the common trunk. The opposite multiplexer then extracts the data and source ID and gives the data to the equivalent line on the other side. If a line presents no data, it is not allocated space on the high-speed link and that space can be used by other stations. **Figure 6** contrasts the FDM, TDM, and statistical techniques.

The main advantage of a statistical multiplexer, or **stat mux** is easily illustrated. In TDM or FDM techniques, each incoming line must be allocated enough capacity to handle its maximum rate of data delivery. For example, a TDM could combine eight 1200-bps lines onto a single 9600-bps line, but not 10 or 12. This is because the TDM has no way to utilize the empty slots of one user with data from another. In a stat mux, the limit on the number of lines that can be combined is the **actual** data traffic, not the potential traffic. If your 1200-bps lines are actually utilized at 20 percent, then each only requires 240-bps of high-speed line capacity. Great! That means 40 lines over a 9600-bps link! Not exactly. The average data rate over the lines may be only 240 bps, but the data will not be evenly distributed. If the data on all lines arrives at random intervals, there will be times when the number of characters waiting will exceed the capacity of the common line. This causes a "queueing delay." There are a series of formulae to predict the queueing delay for various conditions, and all relate the delay to the **average utilization** of the common high-speed link. If the delays are plotted on a graph, a curve of the shape shown in **Figure 7** results. As you can see, when utilizations reach about 70 percent the delay climbs rapidly. In our previous example of 1200-bps terminals at 20 percent utilization, we would reach this 70 percent high-speed link utilization at about 28 devices, still a considerable improvement over the eight allowed by TDM.

If you would like to calculate the limit of the number of lines a statistical multiplexer could combine, follow these steps:

1. Your high-speed line is probably synchronous, so its speed divided by eight is the data rate in characters per second. For a 9600-bps line, this is 1200 cps. Since there is some overhead in statistical multiplexing, use 1000 cps as a better guide. Now, if we want to stay off the sharp climb of the delay curve, we must figure 70 percent of that, or 700 cps is our target utilization.

2. Your input synchronous lines also use eight bits per second for each character per second, so divide each of their speeds by eight to get their capacity. Now multiply that figure by their utilization. If you don't know the value, assume that 3270-type lines are 40 percent utilized and 2780/3780 batch/RJE lines are 60 percent utilized. Add up all the results.

3. Your asynchronous lines use 10 bps per cps, so divide their speed by 10. You can assume that they are 20 percent utilized unless you have better figures. Add these numbers up, too.

4. Now, add the total of the synchronous and asynchronous lines. If the value exceeds the high-speed line capacity (700 cps in our example of 9600-bps trunk lines), the delay is likely to be noticeable and perhaps prohibitive.

For example, if you have four asynchronous lines at 2400 bps running IBM 3270 terminals and 10 asynchronous lines at 1200 bps, your calculation would look like this:

1. High-speed line capacity is 700 cps.

2. Synchronous load per line is 2400 divided by 8 or 300 cps at 40 percent utilization, or 120 cps. Times 4 lines, this is 480 cps.

3. Asynchronous load per line is 1200 divided by 10 or 120 cps. At 20 percent utilization, that's 24 cps per line or 240 cps for all 10.

4. Adding the sync and async load, 480 plus 240 is 720 cps, which is slightly over our target of 700. When utilization is this high, it's a good idea to check the utilization figures carefully or to drop some of the lines out of the multiplexer.

Multiplexer delay is not a popular issue with vendors or users. Most manufacturers state that their delay is relatively short, one or two character times. This would mean that the delay of a character at 1200 bps would be about eight one-thousandths of a second, or eight milliseconds, a figure users could easily accept. Anyone who has ever measured multiplexer delay finds that the manufacturer's figures are inaccurate. For the older style of fixed-allocation multiplexers such as TDMs or FDMs, the manufacturer's figures are accurate, but for statistical multiplexers under what would be considered normal loads,

Business Communication Networks—Putting It All Together

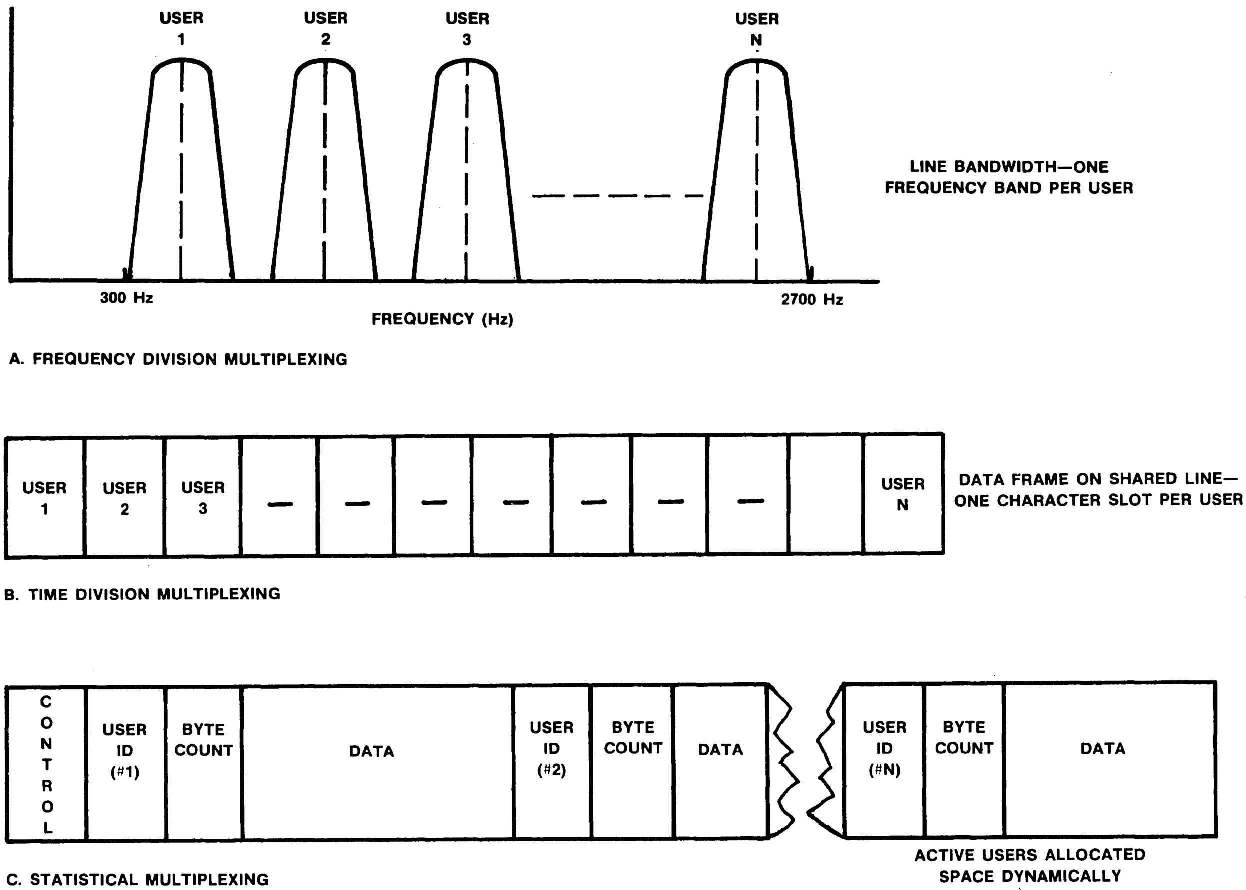


Figure 6 • multiplexing techniques.

delays are normally greater than advertised, and increase with loading as shown in **Figure 7**. Vendors, rather than deal with the delay issue, prefer to relate statistical multiplexer performance to the ratio of the total input speeds to the link speed, a measure called **compaction ratio** or **overbooking ratio**. While these published ratios are useful as a guide, they do not serve as a point of vendor selection because the design differences in the units tend to become unimportant in their contribution to delay at higher levels of utilization, where all units of all vendors approach the statistical curve already described.

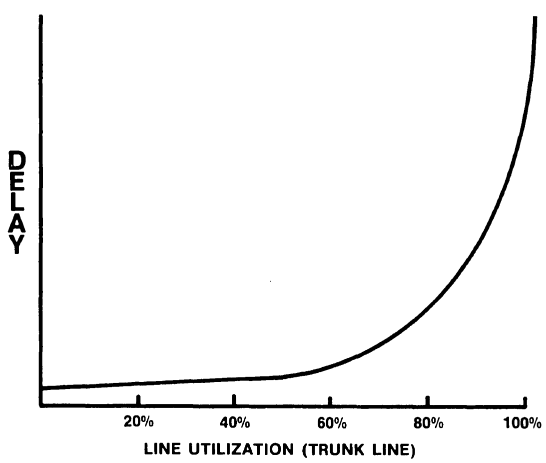


Figure 7 • statistical multiplexer delay.

Concentrators, once relatively popular devices, have fallen into a decline because of the problems in finding a suitable concentration protocol. Because concentrators do not operate in pairs to restore the original line configuration at the host in the way multiplexers do, the host computer must process the concentrated protocol directly. Recent interest in the international packet switching protocol, X.25, has led to the availability of concentrators which use X.25 as their **output protocol** and either asynchronous or binary synchronous (BSC) as their inputs. These units, called **PADs or BPADs** for "packet assembly/disassembly" facilities, provide many of the advantages of multiplexing at a lower cost **if the host computer has X.25 support**.

Buying and using devices like multiplexers or concentrators is often difficult for first-time users because the network concepts they employ are not familiar. Here are some guidelines in evaluating your application for private network devices:

- Evaluate your traffic pattern as the first step in the selection process. Networks are normally built for reasons of line economics, and lines that are either already fully used or that serve isolated areas are poor candidates for combining into networks. If you have lines that are less than 50 percent utilized, that operate under 4800 bps, and that serve users who could be grouped into geographic clusters you should consider private networks for cost savings.

Business Communication Networks—Putting It All Together

• Configure your network to take advantage of natural points of concentration in locations where equipment already exists. **Figure 8** shows a user communication network with the new lines for application of concentration/multiplexing shown as dotted paths. While the **best** point to concentrate a cluster of users is its center, for minimum line charges, it is uneconomical to build an office at a new location merely to serve as the location of the concentrator. If your business is service-oriented and you have lines to customer locations, finding a point of concentration will be a special challenge. Customers may object to having lines to other users terminate at their location and the other users may fear security breaches if the data is proprietary. Be sure to cover these issues, as well as insurance of and access to your equipment, before making a decision.

• Some communication protocols, especially those used by electronic funds transfer equipment, may have very short **timeout** periods. This means that the host system expects the terminal to respond to its inquiries or commands very quickly. Any form of resource sharing used to build networks may introduce **unacceptable delays** and may cause the devices to fail to operate. Polled asynchronous protocols designed along the lines of the Burroughs Poll-Select or NCR standards **may** have these short timeouts. **Avoid using any form of resource sharing if the timeout interval is less than one second, and try to benchmark the products in your environment if timeouts are less than three seconds.**

• Some multiplexer and concentrator vendors limit the aggregate data rates of all the input lines. For example, a vendor may say that the multiplexer has 16 channels of input with an aggregate limit of 76,800 bps. This means that no more than 16 lines can be multiplexed with that device, and that the total line speed of all lines can not exceed 76,800 bps. Since that is an average of only 4800 bps per line, users with requirements for higher speeds might find this unit unacceptable. **Don't make a selection on this basis alone.** Most users report that the ratio of the total input data rate to the rate of the high-speed multiplexed link should be less than 8 or 10 to 1, even on lightly used lines. Performance may limit your application before vendor-enforced restrictions apply.

• Some special devices used in data communication require a nearly complete set of control leads from the device interface, normally the 25-pin RS-232C standard. Most multiplexers and concentrators will handle only two to five control leads, so devices that need more may not function. The secondary channel leads are almost never available, so if you need such facilities you must **seek alternative solutions** to multiplexing/concentration. Most users will not find this a problem.

• Some network products perform protocol conversion so that unlike devices can communicate. While this can be a benefit to multidevice users, **it must be evaluated carefully**. Just stripping off the IBM bisync control characters from the start and end of a message will not make the data within compatible with a DEC terminal. Protocol conversion within network products is most useful in permitting host-to-host connections where data presentation formats can be negotiated by the programmers involved.

• **Don't neglect the TDM-type device** for applications where the lines to be combined are 4800 bps or less and highly utilized, or where wideband (56K bps) service is used. A TDM is less expensive than a statistical multiplexer or concentrator and offers less delay. Use it where input line utilization is so high that the statistical sharing of the resources is improbable.

Once you have settled on an application and general type of product, look for the following characteristics in each vendor offering:

• **Cost.** Most users buy multiplexers and concentrators to save in-line costs. Don't let feature shopping lead you away from the primary goal.

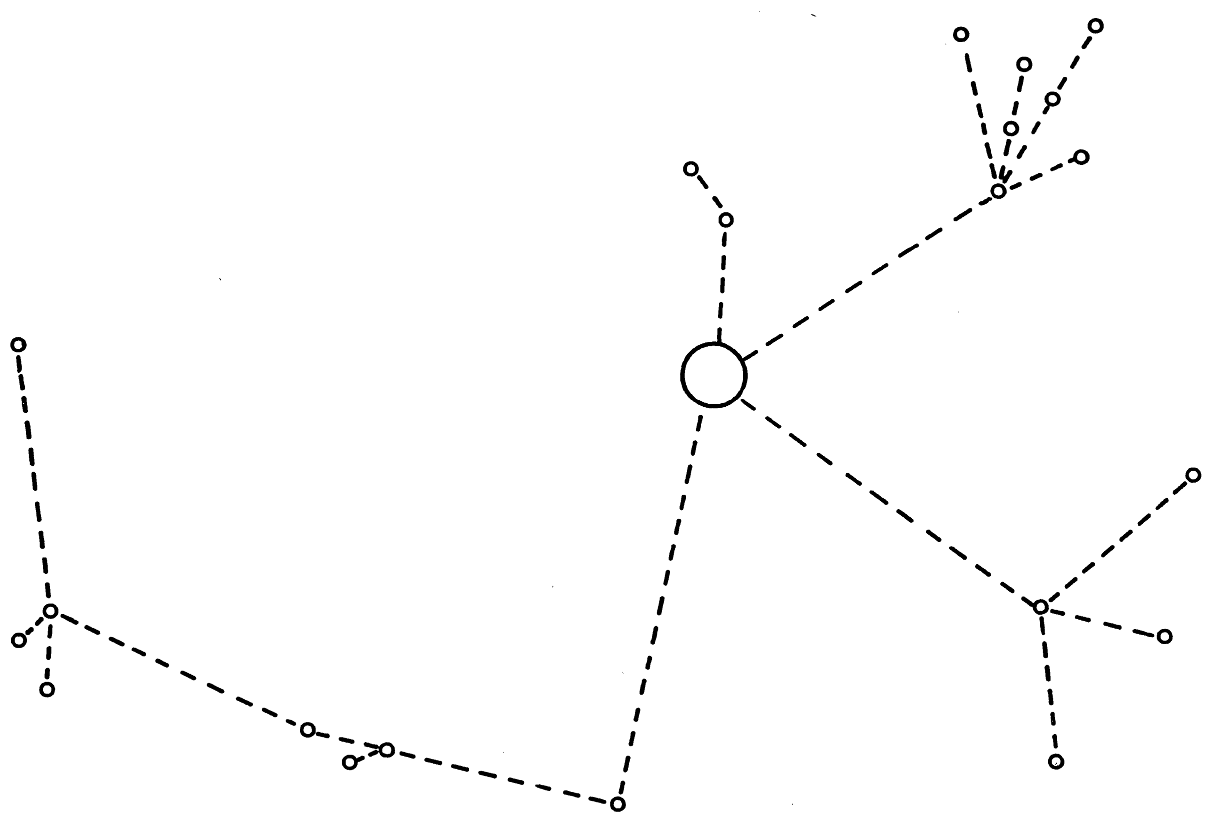


Figure 8 • building a network.

Business Communication Networks—Putting It All Together

• **Protocol support.** Be sure that the devices you select support all the protocols you intend to use with it. If you have any defined expansion strategies, be sure that the devices either support them or will not take part in the expansion. In some cases it is justified to buy equipment that will be obsoleted in a few years if payback has been reached, but try to find something you can continue to use even after network changes occur. Also, be sure that the multiplexer/concentrator can handle the problem of **echo** for asynchronous characters. Most async applications use a host echo of the character to cause the keyed character to display to the user. The delays inherent in multiplexing may require an echo at the multiplexer to maintain operator keying speed. You must **disable your CPU echo in this case** or characters will display twice.

• **Flow control.** Asynchronous terminals and printers may respond to requests to pause in sending data if the rate of data arrival is temporarily too high for the CPU to buffer. This function can be used by the multiplexer to prevent loss of data when many users try to send at the same time, choking the shared link. If you want to send data to a printer, it is almost always a good idea to have flow control in the multiplexer because there is no "conversational" interaction that would naturally cause the system to stop sending after a short message. With keyboard terminals, flow control is useful where the messages sent by the CPU are very long. The value of flow control increases as the ratio of the input data rates to the high-speed link rate increases. **If you are concentrating many terminals that are lightly utilized, you probably need it.** Otherwise a competition for the shared resource which results from, for example, everyone starting at the same time in the morning causes the multiplexer or concentrator to overflow its internal storage and data is lost.

• **High-speed line characteristics.** It is highly desirable that the devices you select operate on a wide variety of transmission paths. Be sure that the protocol used on the concentrated line is relatively immune to satellite delay, and that high-speed operation is supported. Almost all these devices provide for operation to 9600 bps, but some provide standard support for 19.2K bps, and some will provide a special feature to run at speeds up to 72K bps. If the cost of expansion is not prohibitive, the confidence such growth potential can provide is considerable in an environment where usage is tending to increase.

• **Network management.** Private network elements are often transparent to the user and thus to the existing host facilities for management of the network. Be sure that the devices you select have the facilities for testing both themselves and their shared data path, and that the tests can be conducted from a central point and with a high degree of user-friendliness (ease of operation). For small users, built-in test and configuration controls are the least expensive, but larger systems operate best with a terminal device as a supervisory console. If you buy a unit with support for more than eight lines you should have at least the option for supervisory terminal support.

• **Upward mobility.** Most users initially install networks that are "star" point-to-point in nature. When the first concentration devices are placed in this environment they tend to preserve the initial star structure, but as the user grows, the network may evolve into a truly "nodal" structure. **Figure 4** shows a network evolution. Unless you are sure that your network will not grow to a nodal structure with alternate routes and fallback devices, you should be sure that the product line of your chosen vendor will support such growth. Does the unit support multiple remote devices into a single host-support device? Can several multiplexed lines be combined into a single line or split to alternative destinations? Can users request connection to other users in a data switching mode? Multiplexers and concentrators do not mix well between vendors, so try to find a vendor with the features you need to grow.

• **Beware of buzzwords.** Most vendors advertise that they have a line protocol "similar to X.25 Level 2" or an "SDLC-/HDLC-like protocol." It is not even particularly useful to have a line protocol **exactly** like X.25 Level 2 unless it is also like X.25 at Level 3, since your host computer package is unlikely to support Level 2 only. To have something which is only similar to X.25 or SDLC is even less useful. X.25 is a good output protocol for concentrators because they will send it directly to a host computer. Multiplexers which use X.25 as an internal protocol can be used interchangeably as concentrators (by taking one of the pair away and running the high-speed line directly to the host) but may have a greater delay and a higher price than those that use a customized protocol designed for maximum efficiency in a multiplexer environment.

• **Direct bus attachment.** Some vendors provide special forms of multiplexers which at one end of the connection will attach directly to the channel or bus of a computer, replacing a communication controller. The DEC UNIBUS connection is the most common of these. Since the cost of a communication controller for the host is sometimes quite high in itself, combining a multiplexer with it provides additional savings. The risk is the possibility that the host will be replaced with a model that is not compatible with the multiplexer/controller.

• **Multidrop multiplexers.** Some multiplexers can be used on multidrop lines, offering the advantages these types of links provide to users of asynchronous terminals that normally cannot be multidropped. **Don't use multidrop multiplexers with multidrop terminals.** Delay with these devices can be considerable, even with little load. They work best with applications where a block of information is sent to and from the terminal rather than strings of single characters.

■ DIGITAL COMMUNICATION—SATELLITES, DDS & DIGITAL TERMINATION SYSTEMS

Analog systems based on the public telephone network have the advantage of being available between almost any two points in the country and throughout most of the major industrial countries of the world. But analog is limited by the characteristics of this same system to about 19.2K bps. Users who need higher data rates must look to digital transmission. Digital transmission is most often based on pulse-code modulation, or PCM, which is a high-speed sampling technique being phased in as the standard means of sending information on microwave trunks and thus through the "backbone" intercity/interstate phone network. Because PCM works equally well with digital data (it is itself digital-voice data that must be "digitized" to be carried) most of the analog system problems with noise, bandwidth, conditioning, etc are eliminated. Digital transmission, particularly satellite digital transmission, is becoming increasingly popular as the number of voice channels increases and the local system becomes more overloaded. **Figure 9** shows a chart of the relative costs of high-speed transmission facilities, demonstrating the importance to users of weaning away from the traditional analog circuits.

The basis of most digital transmission, whatever the service type, is the so-called **T1-carrier**. This is a 1.544-million-bit-per-second TDM channel that can be subdivided into 24 speech channels carrying PCM-coded voice at 64K bps or data at up to 56K bps.

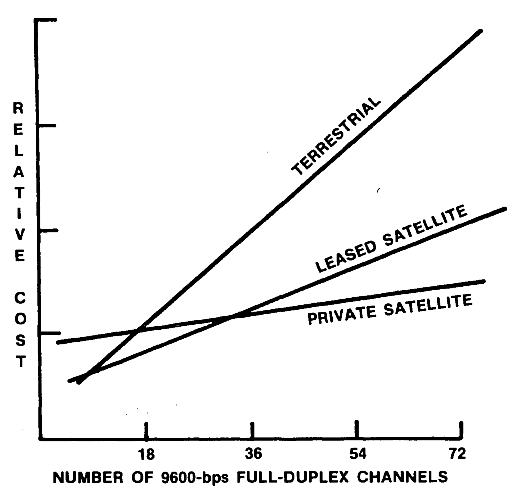


Figure 9 • satellite versus terrestrial costs.

Business Communication Networks—Putting It All Together

Each of the 24 channels can be divided down in any way that the users agree upon. For example, the 56K-bps channels can be time-division multiplexed to serve five users at 9600 bps.

Most users cannot directly utilize T1 data rates, although commercial TDM-type T1 multiplexers are available. AT&T Communication's ACCUNET Digital Service and SKYNET Satellite Services are available in most major cities at rates up to 3M bps. Digital circuits have no modems and an extremely low error rate, making them ideal for high-volume users. Digital service should be investigated by any users who operate in areas where it is available, if leased circuits are being considered, and modems have not yet been purchased. Be sure to check for competitive digital services as well—other companies offer a digital service, and satellite carriers may also be an option in some service areas.

Digital transmission is offered by satellite carriers through the use of a series of geo-synchronous satellites to provide high-speed communication potential thoughout the world, and it offers relatively low-cost user-to-user service between major population centers. Normally, a customer must connect to an access center (earth station) via a leased line. Satellite digital transmission is significantly less expensive than terrestrial for transcontinental usage and the cost trends are much more favorable, but satellites introduce a propagation delay because of the long distances involved (a hop is about 51,000 miles and takes about one-third of a second) which has a major effect on the performance of the channel.

Most high-performance protocols have error detection and correction capabilities, involving a periodic exchange of acknowledgements of correct reception. In this type of protocol, the sending station must eventually wait if these acknowledgements are delayed, since to continue to send would be to defeat the error recovery system and probably cause the other station to reset the link. This situation is sometimes called "closing the window." When a protocol is used with terrestrial links the delay in the communication path is very small and has less effect on the window than, for example, the modem characteristics. In satellite channels, a transmission and acknowledgement requires a round trip on a path with a third of a second delay, causing over a half-second of total delay. Protocols such as IBM's binary synchronous (BSC) which require that each block of data be positively acknowledged must therefore delay over a half-second for each block. This delay time may exceed the actual time used to send the block, so the effective capacity of the channel is dramatically reduced. Protocols such as SDLC/SNA and CCITT X.25 have a greater immunity to delay because they permit more data blocks to be outstanding without an acknowledgement. But even the 7-block window of these protocols is just enough at 9600 bps, and cannot provide full performance at 56K bps.

Users with protocols not efficient for satellite operation may employ a form of satellite delay compensation. These devices are actually protocol converters and transform the delay-sensitive user protocol to an internal protocol suitable for long path delays. This requires that the user protocol be emulated to the extent that the acknowledgements for data are delivered to the sending station by the compensator at the local earth station and **not** by the remote user. At present, several satellite carriers offer compensators as a part of their service. There are also a number of commercial devices that can serve as a form of delay compensator, but **most concentrators and multiplexers that claim to have a protocol immune to satellite delay will not render a user protocol immune also**. This means that the device will not introduce its **own** sensitivity but will not cure your own either. If you plan to buy a compensator, make sure that it really is one.

A new type of service is being licensed by the FCC to provide high-speed digital transmission. This service consists of a series of city-wide **Digital Termination Systems (DTSs)** which may be linked together by satellite paths to form a national Digital Electronic Message Service. Over 30 companies have filed with the FCC for authority to provide DTS, and the first group of licenses have been issued. The DTS concept, shown in **Figure 10**, uses microwave distribution locally to bypass the leased phone lines used for conventional and satellite carriers. The elimination of the **last mile** local loop that reportedly causes most of the problems in data transmission is expected to improve performance and cost, and to shorten wait times in congested metropolitan areas.

Current digital services are essentially forms of leased-line service, and thus, most appropriate to users with relatively high data volumes. Some satellite carriers, requiring user-purchased earth stations, are usable only by the very largest firms. Pricing on

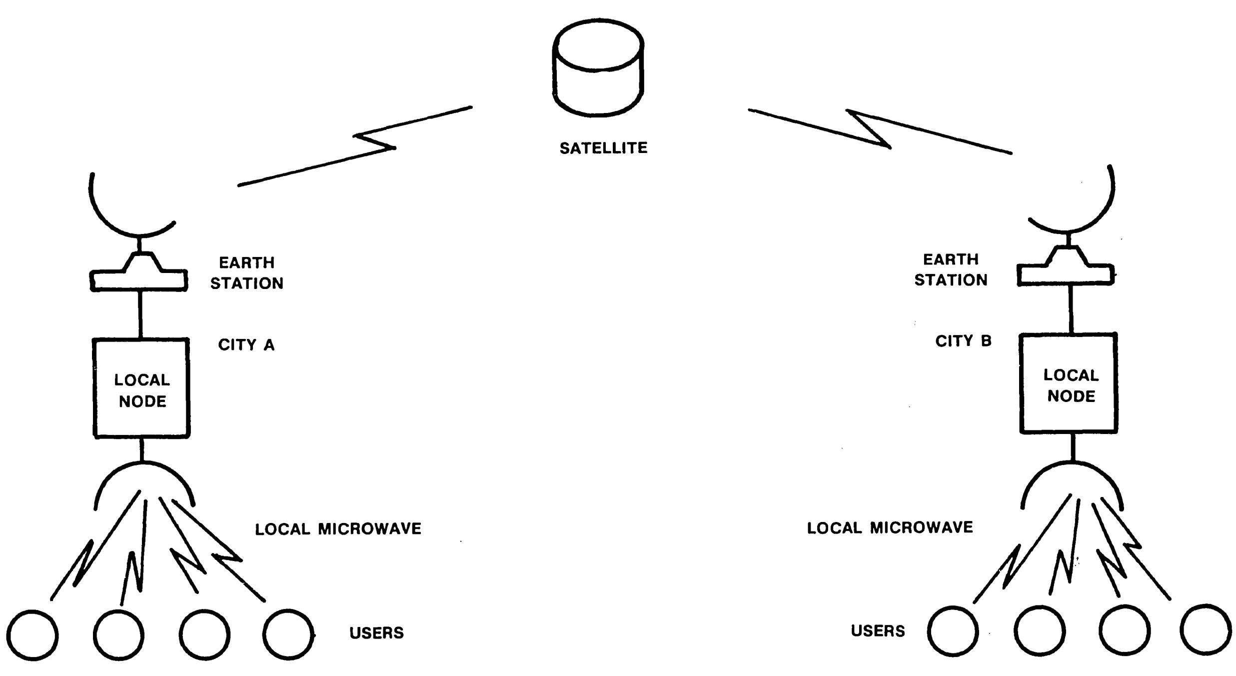


Figure 10 • a typical digital termination system.

Business Communication Networks—Putting It All Together

DTS is not yet available, but it appears that it will be competitive with leased lines.

■ PUBLIC PACKET CONNECTIONS

Public packet networks are to dial-up service what digital services are to leased—and more. Phone service is billed based on time and distance, making dial-up links expensive for users who call frequently or who must call long distances. Public packet networks are generally billed on a combination of time (for dial-in users rather than those who lease a local connection to the network) and data volume. Because the charges are distance independent, some organizations that must collect information nationally may find that public packet networks offer them the only viable method of data communication.

Public packet networks are made up of a heavily interconnected series of computer data switches called **packet nodes** or **packet engines**. These switches collect user data in bundles called **packets** and transfer them from node to node until they eventually reach the node attached to the destination. Terminals may access such networks by dialing into a public access port or by leasing a **local access line** to the network. Host computers, since they normally must be online to the network at all times should a terminal "call in," are connected by leased lines.

Figure 11 shows a chart of the relative line economics of leased connections or dial-up circuits versus public packet. As you can see, nearly any connection more than two states in length can be expected to be less expensive with packet networks. Why use leased lines then? First, very high-volume users who can expect to fully utilize a high-capacity leased trunk will probably find it less expensive. Packet charges are based on allocating that trunk cost over many users, and of course a reasonable profit is added. Other problems that can impact potential users are more subtle: packetizing problems and network delay.

Packet networks switch packets, so if your device doesn't happen to communicate in blocks of data (asynchronous terminals, for example, normally work a character at a time), either you or the network must form the data into packets. This can be more difficult than it sounds, because grouping data into messages introduces a delay while the data is being collected into the packet. If the computer-to-terminal dialog happens to require that one of the collected characters signal an action, and that character is being held up for packetizing, no action will take place and the operator will assume a system failure. For example, if you are keying an account number of up to 10 digits followed by a carriage return, the computer will probably look for the carriage return as the signal to process. If the packet network does not recognize that the return is the last character you will key, and holds the data so that a nice large packet can be collected, the operator will never receive a response. The process of building packets is called "packet assembly/disassembly" or PAD; it can be tuned to a particular application by varying the

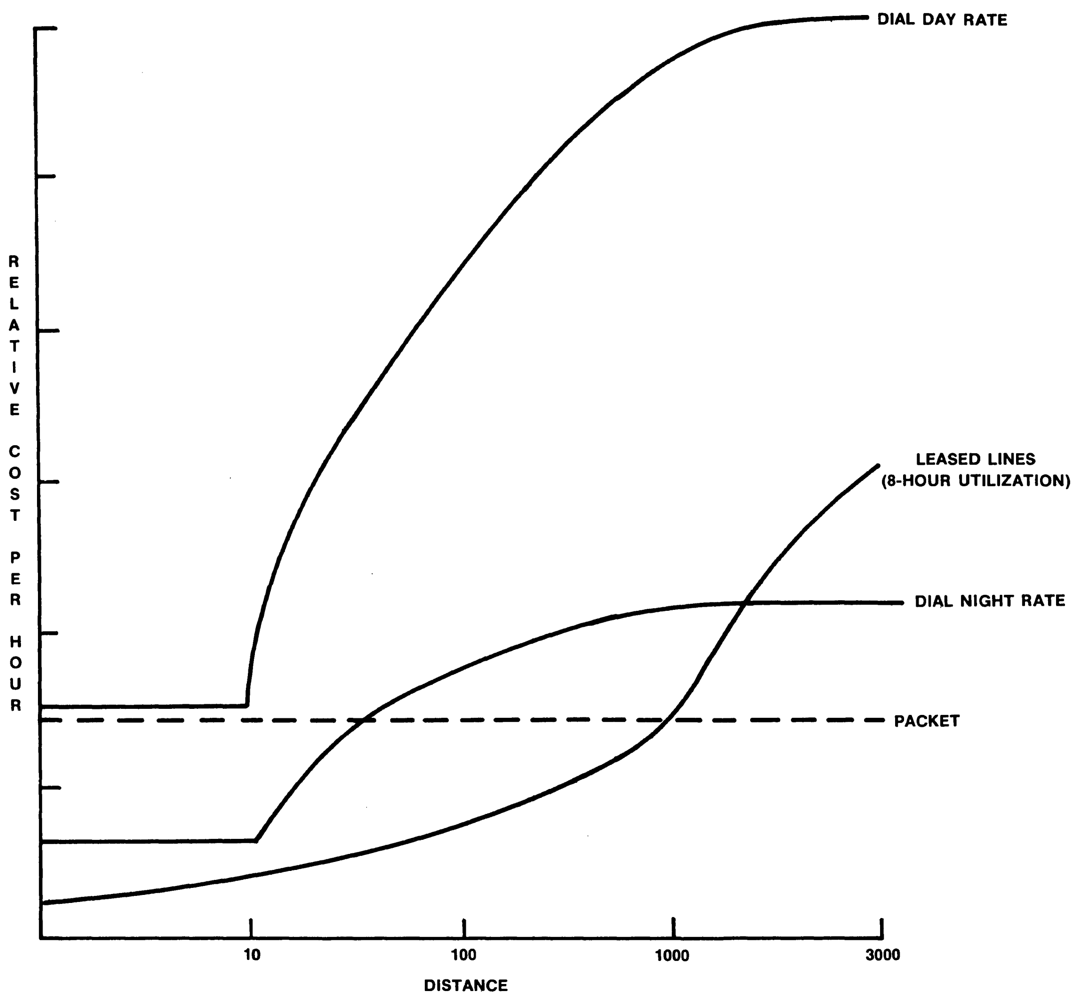


Figure 11 • relative communications costs.

Business Communication Networks—Putting It All Together

value of **PAD parameters**. Pad processing is defined by a series of international standards, X.3, X.28, and X.29, but various network operators such as GTE Telenet may also have their own.

Host computers can attach to a packet network directly in packet mode, using the international packet interface standard known as X.25. Attachment in this way normally reduces the network charges because X.25 can support multiple conversations over a single link. If your computer or front-end processor is **certified** by the network operator, it can be attached directly to the network, otherwise it must be connected through a **network engine** which adds significantly to the costs.

Network delay may be a factor in the use of packet networks. The time it takes to collect the data into packets and to handle the packets in each node can be considerable, creating a problem not unlike the satellite delay problem, but complicated by the fact that the shared trunk facilities of the network may not be able to allocate enough resources to your call to support the rate of data delivery your host or terminal could supply. Users with high data delivery requirements may find that leased line service is necessary, and it may even be less costly.

A newer form of public packet service is available, offered by the unregulated subsidiary of AT&T, Information Systems. This service, called Net 1000 Service, is designed to provide public network features to users who could not normally afford the public packet charges, a feature the company calls **low entry threshold**. Unlike most public networks, Net 1000 Service has a message storage mode that lets users deliver batches of data to the network at one point in space and time, and to extract them at another. If the application programs and terminal operators can adjust to the fact that no actual terminal-to-system dialog is taking place, this could allow a series of dial-up terminals to enter data to be accessed later by a dial-up host.

Public networks are not everyone's answer. The following questions can help you determine their applicability to your network needs:

- Is your network geographically distributed? If you have local connections, public packet is probably not your best choice.
- Is your network currently supported by dial-up lines, or do you expect the use of a new network to involve mostly short connections? High volumes and long connect times are less economical in public packet than low volumes and short calls.
- Does your computer already support X.25? Can it be made to do so? Generally, the most economical means of attaching a host to a network is through a certified X.25 interface. If you cannot do that, the alternatives may be too expensive.
- Is packet service available throughout your network coverage? Telenet, Tymnet, and Uninet each cover nearly 300 major service cities, but you may require connections elsewhere.

Mixing packet and dial connections is possible, but not always easy with all host computers. Check your system's capabilities before you commit to anything.

■ NETWORK MANAGEMENT & CONTROL

Most users depend on statistics from and features of their host computers or front-end processors to control their networks. For simple network structures or for users who can adopt a single-vendor or plug-compatible network architecture this may be the best solution. For users who have begun to introduce such devices as multiplexers or concentrators into their network, it cannot be fully effective because these devices are unknown to the computer, and thus, uncontrollable. Where transparent devices offer network control features, these features may offer an alternative to host control, but it is difficult to integrate several different control packages satisfactorily. In general, if you cannot get the desired level of network control or statistics from your host access method or network nodes alone, you should consider a separate and comprehensive network control system.

Using a network control begins with defining precise objectives. As a user, you have various network control and management features available, some of which may not be applicable or cost justified. Some of these options are:

- **Switch or plug.** This capability, called **patch panel**, provides connections between lines or modems and computers or terminals. This allows you to switch lines or ports on the system when a failure occurs, or to change the division of resources between systems.
- **Line statistics.** This capability provides a way of gathering traffic information and error statistics on a link. The statistics are useful in communication planning, in allocating costs, or in detecting a rising number of line errors which could indicate a deteriorating link.
- **Line monitoring.** This consists of a data line monitor that displays data and control activity on a line, and optionally "T" connections to allow each line to be monitored without disconnecting it. Line monitoring is useful only if the users of the facility are relatively sophisticated.
- **Network monitoring.** This allows the relatively unskilled user to determine the status of lines and other network elements from a central point. The information is normally relayed in more user-friendly displays than line monitoring would provide, and the status of the entire network can be determined quickly. This capability is normally available only through the use of a computer-controlled device, either a network node itself or a separate system with links into various network devices. Most systems such as AT&T-IS Dataphone II Service use special diagnostic/control modems that have a separate narrowband signaling channel (**secondary channel**) for the exchange of network status information.
- **Fall-back and recovery.** Advanced network control allows recovery from failures to be initiated through global, predefined steps under user control or automatically upon failure detection. The recovery may be limited to failures in the links, or may extend to more complex rerouting associated with node or host failures.

Most small users will not require complex network management. If you have less than 10 lines into your computer and do not have a requirement for high reliability and extensive spares, you probably don't need any network management capability beyond the limited statistical information on the lines which probably is available from your host.

The initial step in network control may be either to a patch panel system to permit line switching or to a means of line testing and monitoring, such as a data line monitor or a line statistics monitor. These can serve as a way to gather information on the condition of the line and to adjust the communication configuration when a problem is detected. One of the primary benefits of this step is associated with use of leased lines. If precise information on the error rate or type of failure experienced is available, and can be demonstrated to a repairman on request, phone company cooperation in correcting a problem with a leased circuit can be improved dramatically, increasing your uptime.

More sophisticated network control is likely to be expensive, so the best alternatives should be explored. First, does your access method provide any form of control? Network architectures such as IBM SNA will offer considerable flexibility in routing and fall-back. Second, some private network devices such as concentrators may provide network control. While this type of control may be difficult to integrate into the existing network if host control is also supported, it may be cost-effective.

A final option is the acquisition of a full **network control system** such as is offered by AT&T-IS, Codex, General DataComm, Timeplex, Kinex, Paradyne, or Racal-Milgo. These systems operate in conjunction with vendor-supplied modems to provide a means of gathering network information and conducting testing from a central location. The control information is carried on a low-speed secondary channel which may either share data path transparently or use other pins of the RS-232C interface.

If you are evaluating network control and network management systems, be sure to consider the modem implications if you have a network in place. Some systems will not operate with other vendor devices, and users rarely install modems with network control capability unless they install the entire system. This means that you probably do not have the modems already in place. Network control in a modem is almost certain to increase its cost over comparable units without it.

The following checklist is useful when evaluating network control systems:

Business Communication Networks—Putting It All Together

• **How many individual network elements will the system support?** Don't buy something you will outgrow.

• **What type and speed of modems are supported?** Your primary data communication circuits all have speed and other modem characteristics associated with them. Be sure that the system you choose will support **all** your devices. For example, many of the systems will operate only with 2400-bps to 9600-bps synchronous modems. Chances are that networks with low-speed circuits could not justify the costs of such a system, but there are obviously exceptions.

• **What type of information is available?** You should be able to get alarm reports on problems, line statistics (character and message counts), and more technical displays such as the "star" pattern used to detect line or modem problems with phase jitter.

• **What is the output device?** Reports on a printer or a printer port on the system is a must for records of activity.

• **What types of lines are supported?** Will the system work with multipoint circuits?

• **Is remote testing supported, and in what form?** You should be able to set the remote modems into loopback for testing.

• **What is the total cost?** Be sure to get a quote on the equipment, the modems, any special cables, and the monitoring device. Include the costs of any special bridging needed to route the special control paths around existing network devices, switches, etc.

■ TRENDS IN COMUNICATION SYSTEMS—PLANNING FOR THE FUTURE

The explosive rate of growth in data communication technology demands that users plan their present networks in the light of projected trends, both in their own usage and in the field as a whole. The first step is to project your own demands on the communication network and the data processing facilities it supports, and the second to evaluate the trends in communication that might impact your present and future.

Standards in data communication are probably **the most significant force users should consider.** Organizations like the International Organization for Standardization (ISO), the Consultative Committee for International Telegraphy and Telephony (CCITT), the Institute for Electrical and Electronic Engineers (IEEE), and other have been working to provide definition and direction for the rapid growth in the industry. Standards are being worked on for network architectures, local area networks, integrated digital networks, message switching, videotext, and other areas. If you are moving into an area where standards exist, try to be sure your equipment complies. If standards do not yet exist, try to find out what progress is being made.

A second area to guard is that of technology. Microcomputers and integrated circuits are finding their way into nearly every aspect of data communication. Certain products, such as modems, have actually tended to get less expensive through time due in part to these advances. Before you buy in anticipation of a future need, check the cost trend for the product type and hold off if costs are declining. In buying for a current need, try to select a product which has a state-of-the-art design to lessen the risk of technical obsolescence.

Finally, evaluate the trends in the environment; interest rates, line costs, satellite charges, etc. Some of these trends are shown in **Figure 12**. In general, the costs of terrestrial links are almost certain to increase significantly because of the competition for microwave and cable space. Until fiber-optic cables become common, satellite carriers will be more and more economical in proportion to their earthbound counterparts.

For the far future, most experts believe that network architectures such as SNA and DECnet and international standards such as CCITT X.25 or the OSI Reference Model for Open System Interconnection will be more important in product development and thus to users. Labor costs, continuing to rise, will increase the need to boost productivity with point-of-transaction computing power, so distributed processing will increase. So, obviously, will data communication. The value of information will increase, making it important to collect and disseminate it efficiently and to make studied decisions in areas where the quality of information might be affected, such as its timeliness.

■ SUMMING IT UP—DEFENSIVE PLANNING

Probably the central requirement for future planning in the current environment is flexibility. If you select a path, make it a decision which provides you with as much latitude as possible and which tracks what seems to be the center of the fan of possibilities that grows in width as it extends into the future. Being exactly right, exactly optimum, is unlikely in a field as rapidly changing as data communication, but being excessively wrong is relatively easy.

New users of data communication should plan their data communication facilities around their total data processing strategy, and set a long-term goal in terms of a network architecture or access method that can be worked toward with reasonable costs in the present. Try to adopt the most standardized solutions you find, not the most elegant.

Current users should evaluate their network changes in terms of a long-term strategy for migration and not just on the basis of lowest cost. Where there seems to be a clear direction, follow it. Otherwise, avoid committing yourself even if it means living with existing equipment a little longer.

Finally, all users must conduct realistic postmortem evaluations of failures, not to access blame but to prevent their recurrence. Each time you make a mistake, examine the decision process that led to it, the implementation process that brought it about, and the

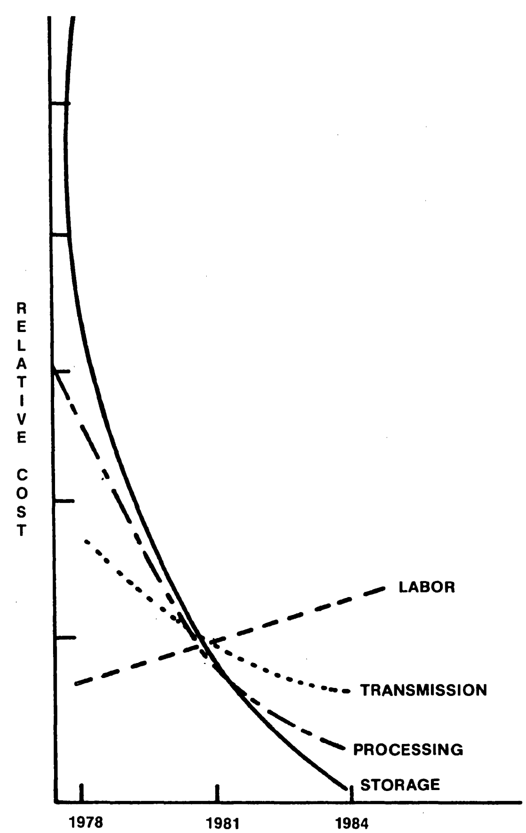


Figure 12 • cost trends in information processing.

Business Communication Networks—Putting It All Together

operational environment in which it failed.

Data communication is a complex field, and integration of the network or of network to computer system is a task to be approached with care. Only by being as knowledgable as possible on your own needs, the technology, the trends, and the products can you hope to make decisions that will serve your company today and in the future.

• END

Carrier Services—When To Break With Tradition

Guidelines for Matching User Requirements to Common Carrier Alternatives

■ INTRODUCTION

The majority of today's data communication, other than local connections, is supported over voice-grade circuits supplied by AT&T Communications. The advantages of using the telephone system for data exchange are many, not the least of which is the almost universal availability of the facilities and the uniform high reliability of such circuits. But, voice-grade telephone circuits are not always the answer. Pure digital paths have been available from both non-AT&T and AT&T sources, and wide-band analog circuits have also found a place in high-volume applications. Even more complex services are available from the so-called "value-added carriers"; packet communication, protocol conversion, store-and-forward message services, and forms of data processing and storage within the network. In addition, more and more companies have entered the communication market as suppliers of voice circuits, often at lower costs than the basic AT&T service.

What is the best transmission buy for the business communication user? Obviously, the answer depends on the application, geography, business trends, and other issues. This report examines the carrier services available today, their probable evolution in the near future, and their application to the problems of the user.

■ GROWTH OF COMMUNICATION—APPLICATIONS & SERVICES

No one doubts that the data communication industry is growing rapidly, and that this growth is due to an explosion of user need for the rapid exchange of information to support business applications. The revenues of communication carriers, estimated at **over 80 billion dollars in 1982**, is expected to rise by over **10 billion dollars per year through the 1980s (Figure 1)**. Computer usage is increasing at all levels of business, and more and more systems are linked with terminals or other computers via communication lines.

Applications of data communication are varied but can be classified as interactive, mail, or batch. Interactive applications are those which require a direct man-machine dialog and are often found where the collection of transaction information takes place at the time of the transaction, such as would be the case with an automated retail sales position. Mail applications are those where a user sends a request or a response via a communication channel to a **mail server**, which in turn enters into a dialog with the destination user to deliver the mail. Batch applications are those where the data communicated is collected manually and delivered in bulk for processing, or where results of a computer process are collected in bulk and delivered for manual review.

Examples of the application groups are as follows:

• **Interactive data entry.** These applications use communication facilities to collect information from human operators, recording business events called **transactions**. A transaction might be a sale, an addition to stock, or any other action which effects the status of the business. The direct man-machine interface allows the computer to check information accuracy and consistency online and thus reduce errors. If the data entry is a part of the transaction process itself, a business record of the transaction is made with its completion, eliminating expensive and error-prone flows of paper records.

• **Interactive inquiry.** An opposite of interactive data entry, inquiry applications use a terminal to retrieve information from a computer rather than enter it. In inquiry applications, the most important consideration is the usability of the resulting display. MIS terminals, inventory status terminals, and many credit authorization applications are examples of inquiry systems.

• **Electronic mail.** Much interest has been generated in electronic mail as a part of the drive toward the automated office. In its basic form, electronic mail allows a user of a system to enter a relatively short message which is free form and intended for human review. The **mail** is then held by the system until the destination user is able and willing to receive it. More complex variations on electronic mail add an automatic filing procedure, levels of mail priority, and **reminder** messages to users who have not picked up their mail.

• **Batch data movement.** There are still many transactions which cannot be carried on interactively with a computer, either because the circumstances of the transaction do not permit it or because of lack of economic justification. Many businesses write checks manually and enter the accounting information in bulk. Batch communication applications are characterized by periods of extremely high utilization when the batch is being sent.

In spite of the variations in applications, which would seem likely to impact the type of communication services which a user might employ for information movement, communication users have traditionally turned to AT&T for their answer—data transmission paths. A dial-up circuit with a properly designed modem can operate at 4800 bps, a respectable if not spectacular data rate, and is available nearly anywhere in the United States for a small installation charge. This availability was the key ingredient in a partnership between data communication and AT&T Communications which still exists today.

Voice telephone circuits are pervasive and generally reliable, but not designed for data transmission. Computers communicate in pulses called **digital signals** rather than in smooth audio-frequency waves such as make up the human voice. To use a phone circuit for data, a translation from digital to the audio, or **analog**, signal is required. The translating device is called a **modulator/demodulator** or **modem**, and is probably the most common piece of data communication equipment in existence.

Modems are an additional expense in the exchange of data, so their elimination has been a recognized path to increasing the savings of communication applications. An early solution was to devise a system to transmit the digital pulses directly, the so-called **digital service** most popularly known as **DDS** after the acronym for AT&T Communication's **Dataphone Digital Service**. But digital and analog services are both **data pipes** which provide a path for information exchange. For this document, we'll refer to data communication services which provide a digital or analog circuit for connecting users without any regard for the mode of use as **path services**.

Is a reliable and inexpensive path the best answer to data communication? Many users would cheerfully settle for it, but improvements in computer technology have introduced another form of network in which data moves between computerized switching **nodes** which handle the information as **data** and not as some collection of digital or analog signals. Some of these computer switches handle text data for message services such as Telex or TWX. Use of computers for these services permits the message switches to save information as it is sent and hold it for delivery to a currently unconnected or busy user. It also permits more efficient use of trunk circuits between cities by saving a message until the trunk is free. In path-service applications, transmission facilities must be available to support any attempt at communication. Computer switches have storage capacity so that the facilities needed for transmission relate to the average data load and the tolerable delay rather than to the peak load. Experts have studied these computer switches and concluded that the **packet switch** form offers the lowest possible transmission costs (**Figure 2**). Packet switches or other communication services

Carrier Services—When To Break With Tradition

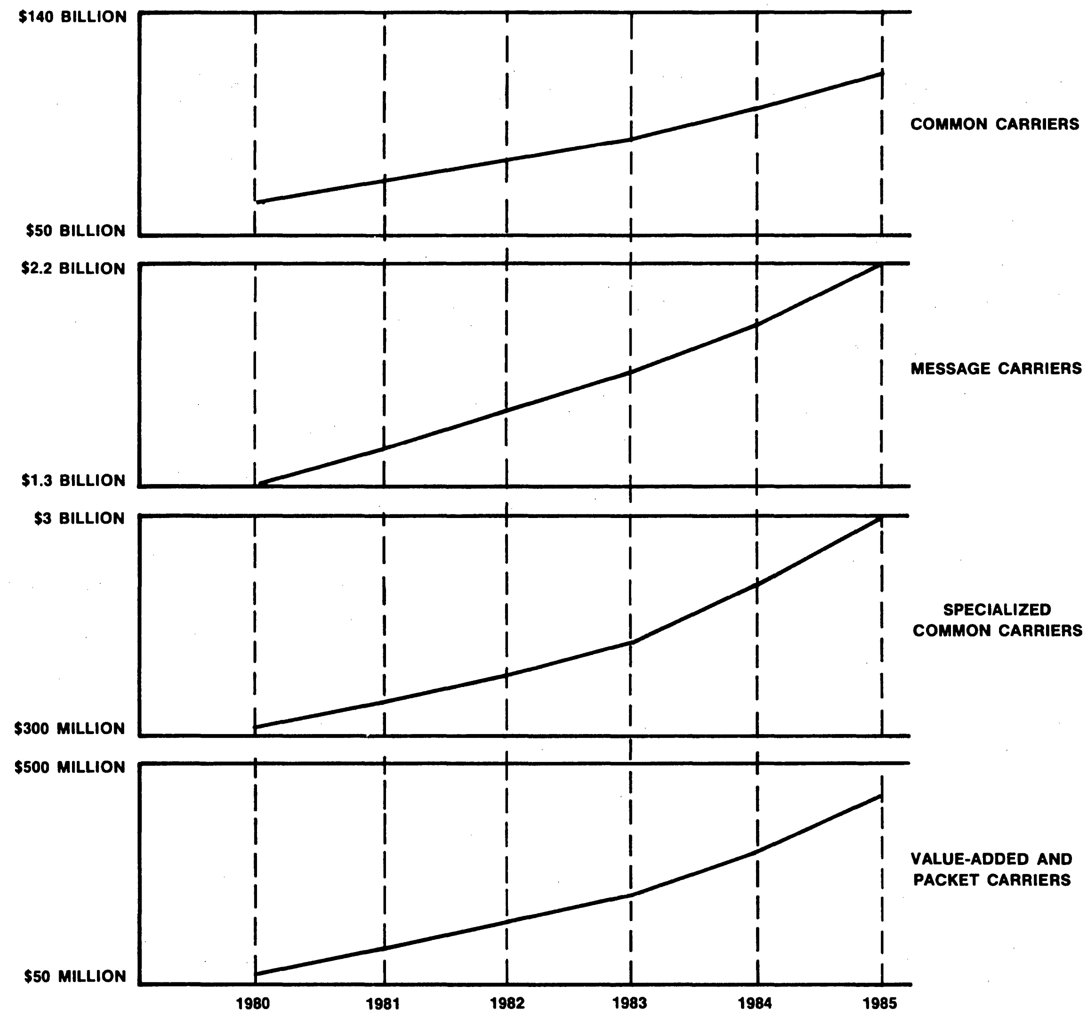


Figure 1 • carrier revenue trends.

which are based on computers and can read the user data may also offer additional services such as protocol conversion, storage of data for later delivery, password or other security protections, expanded billing information, or other services sometimes grouped as **value-added services**. This report refers to all such computer-based networks as **value-added networks**.

Transmission cost is not the only factor in data communication, however. The goals of the automated office have made it apparent that there is considerable information movement taking place in business in non-data forms. Voice communication is still more pervasive than data and likely to remain so for the forseeable future. More and more offices are also finding that services such as facsimile can offer an excellent means of transferring information where the destination user does not need a copy of the data which a computer can read. With these services increasing in popularity, more and more users are looking back at basic path services and the integration of voice, data, and facsimile traffic.

A summary of the user's alternatives for data transmission is as follows:

1. path services
   - through a common carrier such as AT&T.
   - through a specialized common carrier such as MCI, RCA, WUI, etc.
2. message data services such as Telex/TWX
3. value-added services
   - through a public packet network.
   - through a specialized network offering custom protocol or data format support.

Technological alternatives and their associated uncertainties are a serious problem for the communication user of today, but they are probably secondary to the currently bewildering regulatory position in the communication industry. The reorganization of AT&T has eliminated the stability of the backbone of data communication, and all the elements are shaking. For business, it has meant the elimination of a stable planning base for future projects and expansion of current systems. For vendors, it has meant a frantic effort to retain flexibility of product line so that a quick redirection can take advantage of a market opportunity created by yet another regulation, announcement, or rumor.

Computers are almost universally recognized as essential to the operation of a modern business of any scale. Data communication is becoming equally essential, and the effective use of the basic transmission facilities which are available to the user can improve

Carrier Services—When To Break With Tradition

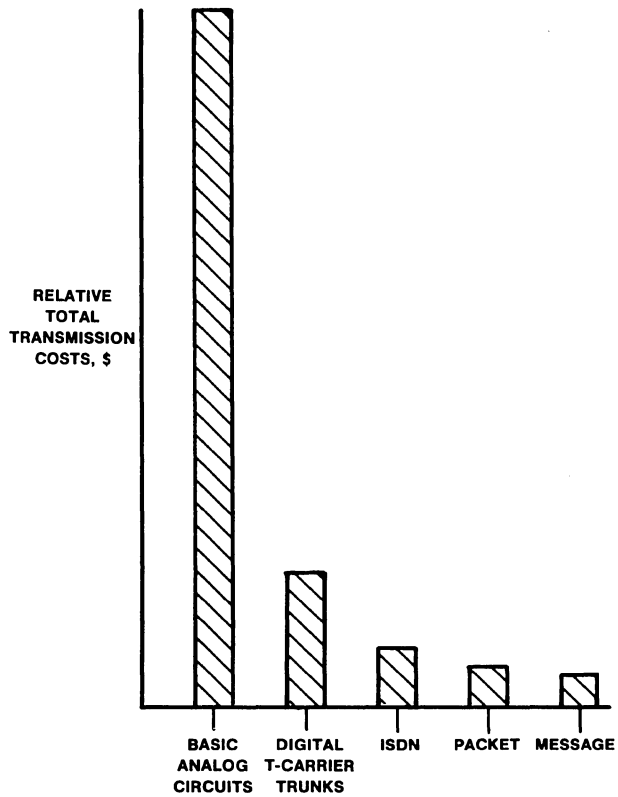


Figure 2 • transmission economics.

service, lower cost, and increase reliability. Improper use can literally put a company out of business.

■ DATA FORMS IN COMMUNICATION

The first issue in the analysis of the best vehicle for data exchange is the form which the data takes. Business information can be categorized as follows:

• **Voice communication**, normally handled through a central company exchange point such as a PBX, and making up the majority of the telecommunication load. Voice exchanges are normally associated with requests for action or information at a high priority and must be supported by facilities sufficient to ensure that access to the resource is rarely blocked. Voice communication tends to rely heavily on mutual speaker recognition and nuances of inflection, and hence must take place over facilities with sufficient fidelity to provide accurate reproduction of the primary speech frequencies. The phone system provides transmission of frequencies between 300 and 3,000 cycles per second (Hertz), and this is normally considered the minimum standard.

• **Message communication**, handled through Telex/TWX systems, electronic mail, etc. Message services are normally used for a priority delivery of a document which might otherwise be sent via conventional mail. The messages are normally brief, and the data rates required for this service are quite low. Telex lines are still sometimes called **sub-voice grade** because they offer less bandwidth and therefore data capacity than a standard telephone circuit. Although it is possible that users of message services might enter into a form of conversation by sending short and alternating messages in a form similar to a dialog on citizen's band radio, most message services assume that the receiver may accept the message without actioning it immediately.

• **Computer data exchanges**, transferring information between computer and terminal or between computers. These exchanges differ from the voice and message exchanges previously mentioned in that the information is in a computer-readable form at one or both ends of the connection. Data exchanges can be similar to voice connections when the terminal and computer interact in short conversational transactions, or similar to a message service when a batch of data is accumulated and sent at one time. All data exchanges have a common element—the information must be accepted and delivered in digital form because the devices served are digital devices.

• **Image-oriented services**, such as facsimile or videotext. These services represent movement of data in a form which represents its **visual attributes** rather than its data content. A computer terminal will display the letter **e** upon receipt of a binary code (11001001 in the ASCII code set) in the proper digital form. The appearance of the **e** will depend on the display characteristics of the device. A facsimile or videotext system may send a letter **e** as well, but it will do so in a way which conveys the way the **e** appears at the sending station and possibly lose the fact that **e** is being represented at all.

Examination of the above data forms suggests that there are two basic kinds of information being exchanged, **human-oriented** and **computer-oriented**. Human-oriented information is sent and received to preserve a sensory characteristic of the data (its sound or appearance), while computer-oriented information is sent and received in ways to preserve its internal code structure, the binary digits making up letters and numbers. The business mix of these data forms is shown in **Figure 3**.

It is difficult to make a single method of data transmission serve with maximum efficiency over such a spectrum of needs. Analog, or voice, circuits have the ability to transmit the human voice well because the information content of a word per unit of time is relatively small. Rapid-paced speech communicates at a rate of about 3 words per second, or about 15 characters per second. Visual information can be absorbed at a much greater rate, so it is not surprising that a system designed for voice cannot carry it at its potential speed. Computers can exchange information at nearly a million characters per second, far beyond the capacity of a voice circuit.

Communication facilities of sufficient capacity to satisfy the potential rates of information absorption for all forms of data would seem to require a user to own a television station and a battery of high-capacity fiber optic cables, in addition to a normal voice telephone system. There are obviously economic discontinuities in the matching of transmission capacity to the intrinsic speed of the data form involved. These influences have tended to crystalize into two trends, selection of multiple transmission methods for each form of data based on cost/performance trade-offs and selecting a single transmission method which will serve all methods satisfactorily. The former method is favored among large companies whose needs in each area are sufficient to justify dedicated and optimized facilities, while the latter is most common among users whose volume of exchanged information in each form does not justify a dedicated facility. A bank clearing 100,000 funds transfer items daily can surely justify a data network, but a small office transmitting 20 line items daily would be better served by having its data communication facilities share a phone circuit.

What particular transmission or carrier services do a particular set of business practices or data forms justify? That question, like many in the evaluation of communication alternatives, is answered first based on the analysis of needs.

■ NEED—THE BOTTOM LINE

Although some users can design new communication applications based on the most economic transmission facilities available, most find that other considerations such as the manual supporting operation or the computer software development dictate the basic structure of the application. Each system has a set of intrinsic communication requirements based on the way the system is to be used, the data volumes expected, the skill level of the users involved, and the geographic distribution of data sources and processing/storage points. The first step in the evaluation of the services required for a particular application is to formalize these **hidden** requirements.

Evaluating an application, whether already automated or a

Carrier Services—When To Break With Tradition

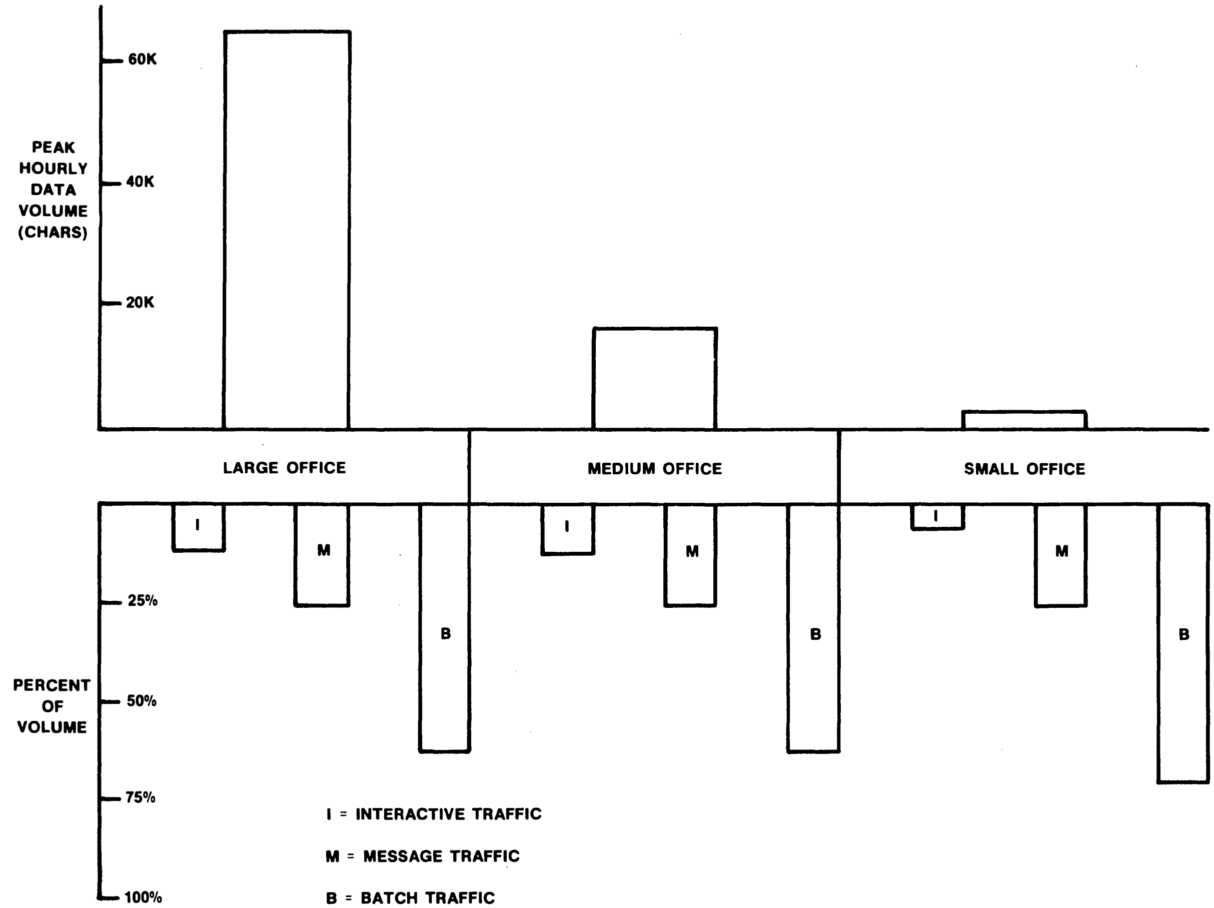


Figure 3 • business traffic distribution.

candidate for automation, begins with the evaluation of the **use** of the system, not with the technical characteristics. The following questions can help define the way in which the application would use data communication.

1. **What form does the original data to be input to the system take?** For example, does the source of data consist of a handwritten form, a printed document, verbal or visual input, computer-formatted data?

2. **What form of results are expected?** This is always a difficult area to resolve, since many data processing operations will tend to give their internal requirements for processing and control as though they were final results. Some questions will help to identify the real requirements. First, is there a requirement that a computer-readable copy of the document, transaction, or input be maintained **for actual use elsewhere in the business**? Second, is there any need to preserve the **appearance** of the input document in the results (a microfilm image)?

At this point, it is possible to examine the input and output requirements and draw some preliminary conclusions. If the data source and results are both to be **human** form and no computer-oriented uses of the information are expected, then it is possible that the transfer of information from sender to receiver could be in graphic or visual form, facsimile or videotext, rather than as terminal/computer **data**. Even if the information **is** transported as data, it may be practical to use a message carrier such as Telex/TWX for the exchange rather than a traditional communication circuit capable of high performance, since human-human exchanges are typically relatively low-speed interactions.

If the data must be in computer form at one or both ends of the connection, then the application is more a traditional communication application, and other questions will apply.

3. **What is the skill level of the input personnel as it relates to the difficulty of entry?** One of the most significant of the hidden costs of a data input operation is the cost of undetected errors. Experience has shown that the input process is the time to detect and correct errors; the source document is already in the hands of a person, and a dialog with the computer at this point can prevent errors from multiplying through later processing. Interaction between computer and operator implies a very quick response time and thus a communication system with minimum delay. The impact of transmission delays on interactive, critical applications can sometimes be minimized by providing intelligence at the terminal itself, but relational editing of input data against data already on file must normally be done via interactions with the database itself, via communication link.

4. **What is the skill level of the users of the output as it relates to the complexity of the output format?** If it is necessary to portray a very complex output to a relatively unsophisticated user, some form of visual aids may be required. This in turn may require either a form of chart/picture output such as can be provided through videotext or an interactive dialog to receive the data. The former may be sent economically through a special graphics-oriented carrier service, while the latter must probably be supported by a dedicated path.

5. **What data volumes are expected?** Many of the specialized

Carrier Services—When To Break With Tradition

and value-added carriers lease facilities from AT&T Communications and are able to reduce their cost to the user through more efficient use, which is to say lower idle time. A user with very high utilization may find that some alternative carriers will not provide significant (or any) savings because of the way in which the service is billed. Public packet networks like Telenet and Tymnet, for example, bill in part based on data volumes and not on connect time and distance. This favors users with a wide geographical separation who make interactive connections where relatively little information is exchanged. Data volumes may also affect the practicality of a service; if the maximum data rate the service can support could not handle the volume, it is obvious that that particular service is unsuitable for the application.

6. **What are the operating schedules and response time requirements of the application?** Most users find that their applications have peak usage periods. It is important to know the traffic of the network and its distribution over time, since the impact of a 1,000 transaction day is greater if 900 fall in the last hour! The issue of operating schedules also leads to questions of response time. Human-engineering studies on data entry applications show that operators are relatively sensitive to delays in system response. In a **production keying application**, the delays tend to disrupt the rhythm of the operator, who spends time watching the screen to see if the system has caught up rather than watching the document and keying. In more complex interactions, the operator can mistake the lag for a loss of communication and attempt to repeat the request, an action which will be inefficient at best and potentially disastrous in cases of file changes and deletions. All data has a range of expected response (**see Figure 4**). Another schedule issue is differences in operating hours between users resulting from differences in work rules, time zones, or equipment usage schedules. These differences may make the delivery of information difficult because both sender and receiver may have only a small period of mutual availability.

7. **What trends can be identified in the application?** This includes not only expected changes in volume but changes in processing mode, equipment, level of personnel, sensitivity of information, cost of resources and cost sensitivity of the business as well. Very few communication applications can fully cost-justify in less than three years, so it's important to avoid major changes within that period. There is no reason to acquire unnecessary transmission capacity in expectation of needing it in several years, but there is every reason to be sure that the type of transmission facility you select will support your needs without expensive changes over the business life of the application.

The preceding questions on the skill levels of the users and the characteristics of the application classify the use into one of the following categories:

1. **Interactive/critical.** This group includes applications which have complex user/system interactions in proportion to the skills of the system users and which are critical in terms of response times.

2. **Interactive/non-critical.** These are applications where the interaction is not complex in terms of user skills and where the response time of the system is not critical.

3. **Batch/critical.** Applications where the information may be sent with little or no user-system dialog but where the volume of information requires a relatively high-performance data path.

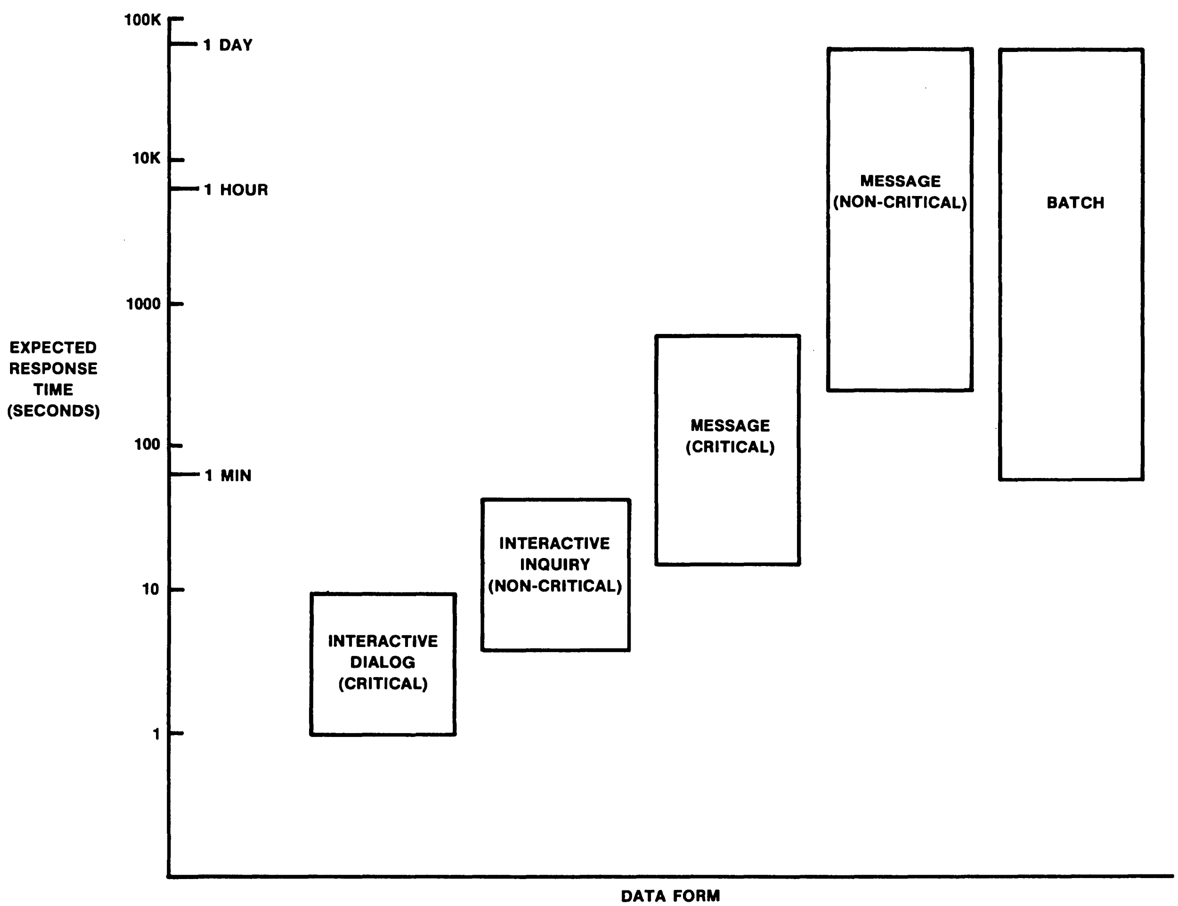


Figure 4 • response anticipation ranges by data form.

Carrier Services—When To Break With Tradition

4. **Batch/non-critical.** These are batch applications where the data volumes are sufficiently low to make any reasonable data transmission rate satisfactory.

5. **Message/critical.** These are batch data applications where the users cannot simultaneously connect to the system because of differences in business hours or limitations of resources, but where a relatively large amount of information must be exchanged.

5. **Message/non-critical.** Message applications where the amount of data exchanged is within reasonable transmission limits.

The final set of application questions deals with the geography of the business.

1. **Where are each of the data sources, process points, and data users located?** A map of locations is useful, particularly if major cities are also conspicuously marked. Many communication services are available only in or near major population centers.

2. **Where are other user facilities not involved in the operation located?** Sometimes buildings of other divisions of a company can be used as a communication drop or switch point, and sometimes data processing points can be relocated to an area better served by communication carriers.

3. **Are there any private communication facilities already installed between any user facilities?** Map all leased lines, tie lines, private copper circuits, satellite or microwave circuits, **any** communication ties at all. Many users overlook such facilities if they belong to other divisions.

The completed network map may answer many of the services questions by inspection. If the network is relatively small and consists of a series of point-to-point connections, communication analysis can be done almost by treating each connection separately. If the network is highly interconnected, having many points of intersections of circuits, a more complex planning task will clearly be required, and the range of carrier services which may apply will be greater.

Let's summarize the analysis of needs and the range of results.

• Communication applications can be classified by the type of interaction between user and system which the communication path must support. Interactive use implies a dialog which is sensitive to delays in responses to user requests and thus requires a transport which has little additional delay. Message use permits more delay between sender and receiver to optimize communication efficiency because a complete message will be sent at one time, and real-time response by the addressee is not normally expected. Batch use implies data collection outside the communication system and transmission in bulk. Batch applications typically require high-performance paths because the volume of information is very high and the impact of even a small decrease in the effective data rate of the line is multiplied by a high volume to produce a long total delay. Interactive traffic is sensitive to single-transmission delays because of the back-and-forth nature of human interactions. Batch traffic can tolerate occasional high delays on a single transmission if the average rate of data movement is high. Message traffic can normally tolerate considerable transmission delay if the users themselves are not delayed in the entry of a message or the receipt of it; if the delay is introduced in the movement of the complete message between elements of the network itself in getting it from the area serving the sender to the area serving the receiver.

• Normal sensitivity to delay in any application is increased if the interaction of the user to the system is complex or if the user is unskilled. The high concentration levels either condition requires are more easily disturbed by unusual reactions of the system, and unskilled users may interpret lack of response as a system failure.

• Tight schedules and high volumes also act to increase the sensitivity of an application to delays. Systems which regularly cause reports to be delivered late or employees to work overtime can be both expensive and resented by users. Neither situation is desirable, and in combination they may force a business to re-evaluate a decision, often at a significant cost.

• Networks which are widely distributed geographically and carry low volumes of information often cannot utilize coventional path services effectively because such services tend to bill on combinations of connect time and distance. Highly interconnected networks with high-traffic volumes will utilize data paths more completely, and the complex interconnections may be more difficult to achieve with value-added network services.

■ CARRIER SERVICES—FIRST LOOK

Communication carriers are companies who provide users with the ability to move information from place to place. The services which these carriers provide varies from the familiar telephone line or other basic path to a complex set of data storage and processing services, coupled with information movement and protocol and format conversion.

The most familiar carrier service to users is the data path, taking the form of a dial-up or leased analog phone circuit, a digital data circuit, or microwave or satellite connection. Path services are usually billed by combinations of connection time and distance and are supplied by AT&T Communications, Bell Operating Companies (BOCs), independent telephone companies, specialized common carriers such as MCI, and satellite carriers such as SBS, RCA, and American Satellite.

A second form of carrier service is the interactive switched network or public packet network. These services use computer nodes to route user data from place to place, and the nodal computers may provide additional services such as protocol conversion. **See Figure 5**. Packet networks are very popular internationally, where business rates for the use of the telephone system for data exchanges are higher. In the U.S., the popularity of the packet network has been growing slowly but steadily as users discover its cost saving potential. Most public packet networks bill primarily on data volume exchanged, so the distance between users is not a factor in communication cost.

Similar in concept to the public packet switch networks are the special format networks which cater to a particular type of traffic, either a specific protocol (**public packet networks are not necessarily X.25 networks**; most users of them do not employ X.25 at all) or to a particular form of data such as graphics or videotext. The charging basis for these networks vary.

Networks also exist for text-/message-oriented communication. This type of service implies that the network is the intermediary between sender and receiver, accepting data at one place and time and delivering it at another. The ability to leave a message for a destination which is either not connected or busy can be invaluable, and the printed text of such a message can be treated as the equivalent of a letter, subject to copying and manual review by multiple users. Messages may also be directed to multiple users from a single transmission.

Leaving messages in the network almost inevitably leads to leaving other forms of data in the network, and to the use of network-based computer resources for processing of information in some form. Microcomputer network services which offer subscribers the ability to send and receive messages from one another also offer long-term data storage as well as files of information which any user can access and copy. IBM and AT&T Information Systems (formerly American Bell) have incorporated more sophisticated processing capabilities within their networks. The logical structure of these networks is shown in **Figure 6**.

In addition to the carrier services relating to the type of transmission capabilities and pricing of data movement, most carriers offer **secondary services** which may be of significant benefit to users:

• **Billing and accounting.** Communication resources are expensive, and businesses normally wish to allocate their cost to the internal users based on the actual charges the user incurs. Some networks provide accounting codes or user identifiers within the company account as a whole, so that the task of breaking down the cost of the service is done by the network.

• **Protocol conversion.** Many companies have several generations of communication terminals and computers from several different vendors. In some cases, users who need access to two different computer-based applications have two different terminals with different keyboards and different phone circuits to support the access. Network-based protocol conversion allows a

Carrier Services—When To Break With Tradition

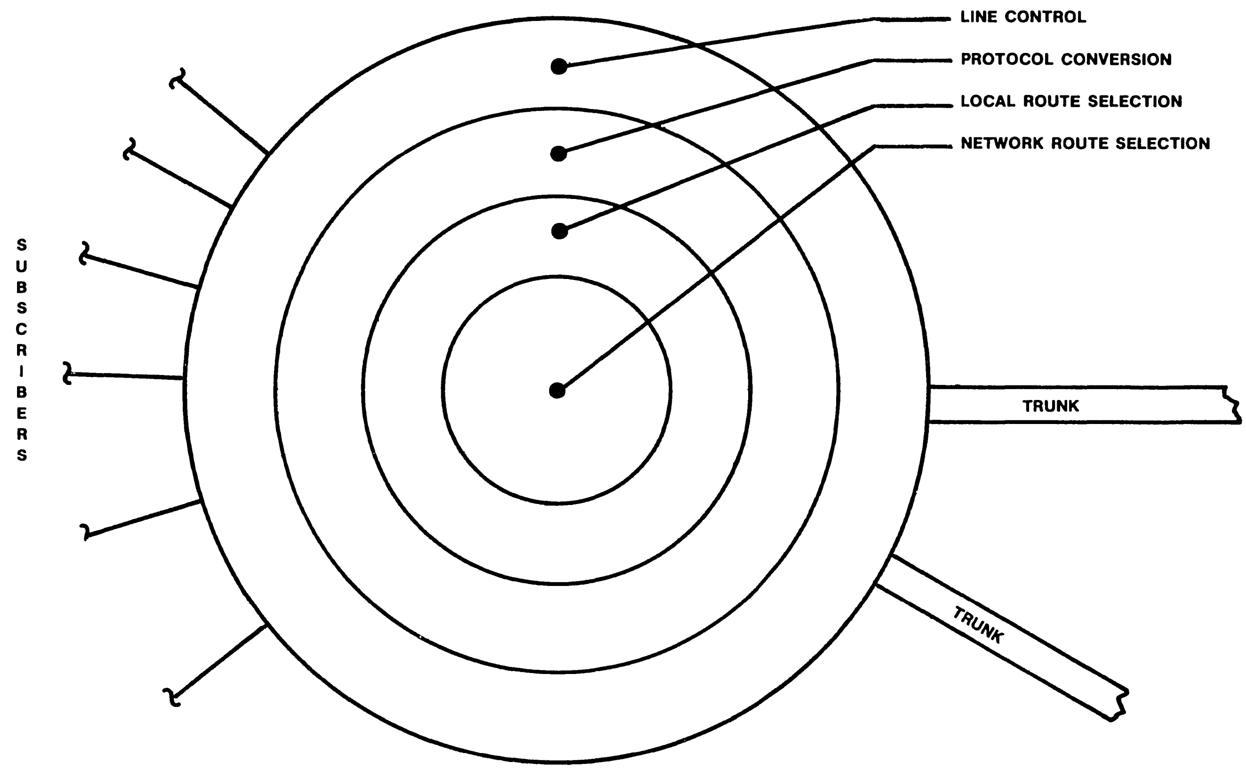


Figure 5 • a packet node structure.

single-user terminal to access multiple-host computers which support different and incompatible protocols. Protocol converters also **immunize** a user from the effects of **long network propagation delays** such as are found in satellite networks.

• **Switching and connection features.** Many users desire to treat data connections in the same way as voice connections—dial-up to start and then hang up at the end. Automatic dialing and answering performs this function through the path services of a common carrier, but dial-up connections at high speeds are not presently available. Most computer-based data networks support data **calls** between users, and have protection schemes to prevent a user from receiving a call from an unauthorized station. This connection security may be important in networks where outsiders could gain access through network switching since most computer applications do not have passwords or other forms of connection authentication built in.

• **Reliability/diagnostic capabilities.** Some business applications are very sensitive to service interruption, particularly at certain times of the day. While dial-up path service from a common carrier can quickly replace a call if the line fails or degrades to the point of unusability, leased line subscribers may suffer a loss of service under such conditions which lasts several hours. Value-added networks normally offer users redundancy within the network itself, providing alternate routes around a failed network trunk. The advantages of this service are somewhat lessened by the fact that the local loop between the subscriber and the central office is probably the greatest single point of failure. Since this loop is also required to connect the user **to** the value-added network (most use leased lines to attach subscribers to the serving node), backup of internal network paths does not protect from the primary source of failures.

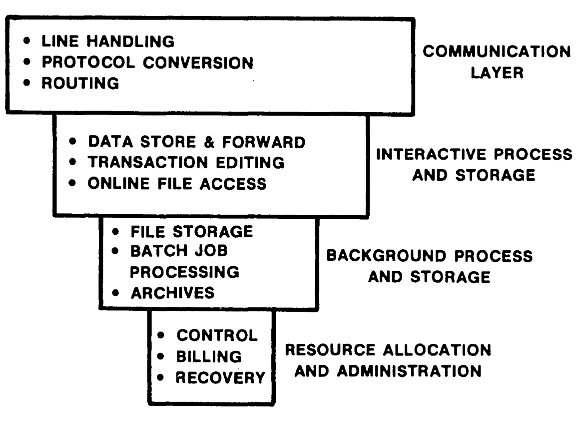


Figure 6 • logical layers of a processing services network.

■ BASIC PATH SERVICES—THE COMMON CARRIERS

The backbone of any communication system is the **vehicle which carries the information** itself, in whatever form it takes. Information can be transmitted in many forms, all of which require some form of governmental guarantee of access and freedom from interference if they are to serve the general population. Telephone wires, poles, microwave towers, etc, are expensive resources, and a company could hardly be expected to wire the entire United States without a way of protecting the massive investment required. Primary telephone service operated many years as a public utility—a sanctioned monopoly regulated not by competitive pressure but by government agencies such as the Federal Communication Commission (FCC). In return for the status of a utility, public telephone systems such as the Bell System were restricted from offering other services which might logically associate with their basic path service, but

Carrier Services—When To Break With Tradition

which could be offered in unfair competition with other businesses because of the government-sanctioned universality of their basic service and the corresponding public identification of the phone companies with any form of communication. Although recent issues of deregulation have dramatically changed the system, basic public telephone companies and the long lines portion of AT&T are still **regulated common carriers**.

Common carriers provide users with a path over which information can flow. The path need not take any particular form; copper wire, microwave terrestrial or satellite transmissions, or fiber optic paths are equally acceptable as long as they can carry information in some way. The hierarchy of paths and trunks in the AT&T system is shown in **Figure 7**. In general, basic path services transmit information without any knowledge of or sensitivity to its form as long as the transmission parameters of the media are met. For example, the dial telephone system can transmit voice or data, facsimile, etc, as long as the transmissions do not generate frequencies outside the **bandwidth** (frequency range) of the system, approximately 300 to 3000 Hertz (Hz) per second.

Suppliers of basic path services can be grouped as common carriers, specialized common carriers, and satellite carriers. The common carriers, independent telephone companies such as GTE and the AT&T System, generally provide direct consumer phone service. Specialized common carriers such as MCI or WUI provide services such as reduced-cost long-distance calling through supplemental facilities, but rely on the common carriers for their connection to the user. Satellite carriers provide an alternative to the common carriers for service between two points by using satellite transmission rather than the terrestrial microwave that encompasses most of the long-distance phone links.

Because basic path services are **basic**, the communication user who evaluates alternate sources of these services is primarily concerned with saving money. The Bell System operated for over 70 years on a complex structure of cross-subsidies between the long-distance portion, **AT&T Long Lines** (now AT&T Communications), and the local phone companies now called the **Bell Operating Companies (BOCs)**. A substantial portion of the fees paid for long-distance service were paid back to the operating companies to keep consumer phone bills down and to subsidize the installation and operation of the local loops, telephone instruments, and local offices. These subsidies have helped make AT&T phone services almost universally available, affordable, and of a quality ranking with the best in the world.

Cross-subsidies of elements of the phone system can lead to inequities, however. The fact that long-distance phone services are priced **artificially high** penalizes the user, business or individual, who makes many long-distance calls. This inequity created a demand for a form of lower-cost long-distance calling, which many companies were glad to fill. The most common service which these specialized common carriers offer is the alternative long-distance calling services, although many also offer leased line and even DDS data services. The specialized common carriers (SCCs) provide a backbone network of long-distance links which may be leased from AT&T Communications or another carrier, or owned by the SCC itself.

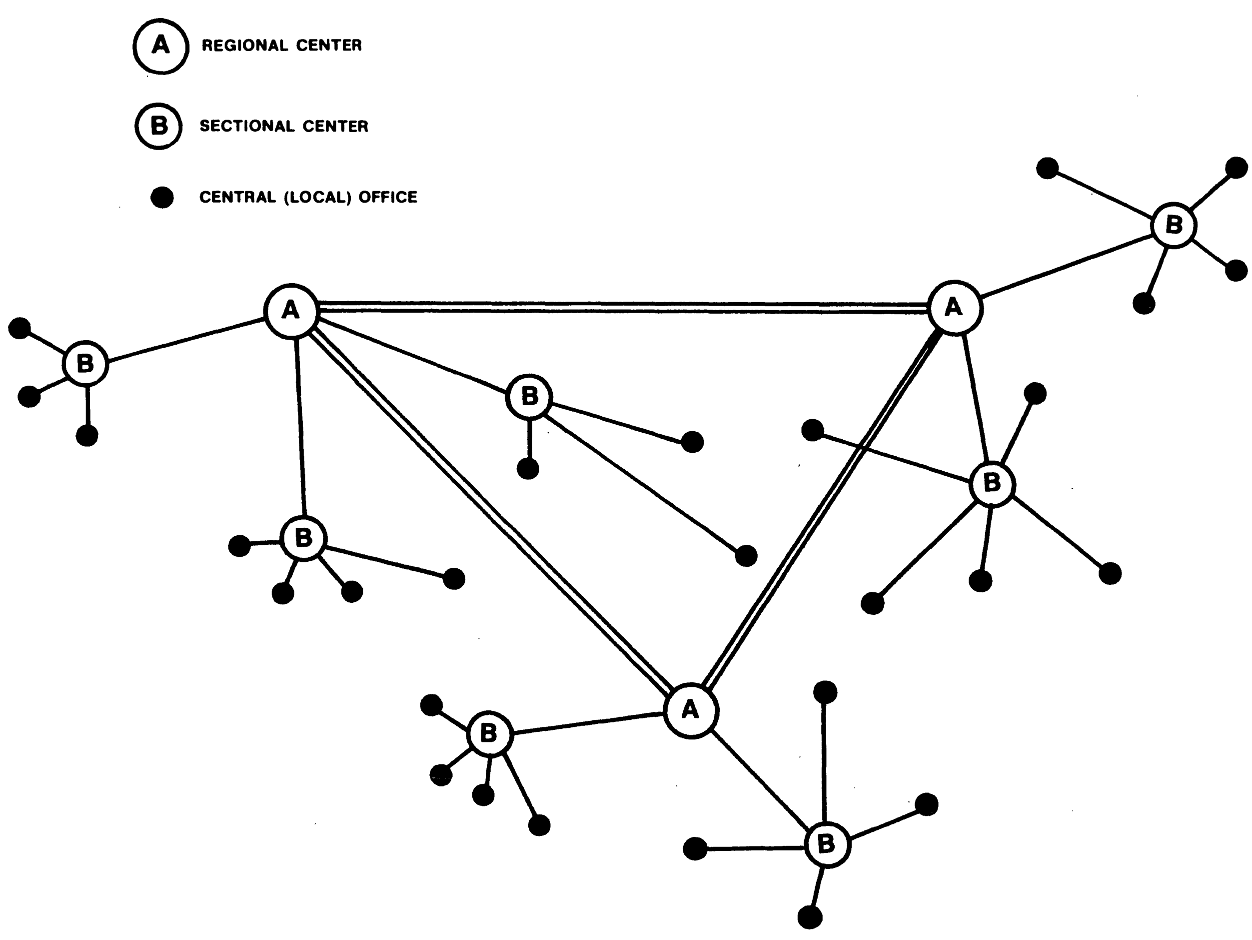


Figure 7 • telephone switching structure.

Carrier Services—When To Break With Tradition

The specialized carriers also design their networks to operate at a higher utilization than AT&T Communications. This translates to more calls per hour per line and allows the user to be charged less per call while yielding a high return on investment for the line. It also causes more busy signals to be encountered by users; experts have estimated that users of the specialized common carriers are three to seven times as likely to get a busy signal as a direct user of AT&T Communications.

Alternatives to basic dial-up services are available from the telephone company. The most common of these services is the **Wide Area Telephone Service, WATS**. WATS services are available in service zones at charges proportional to the coverage and the usage. Either an **originate** form of WATS or an **answer** or incoming call form, the **800 number**, is available. WATS circuits may be used for data communication, and may offer significant cost reductions over basic dial-up service, especially for companies who have many short data connections over a wide geography. If you plan to use WATS for data, chances are that the incoming form will suit you best. **Be aware that WATS does not permit call-in from locations in the same state/area as the termination of the 800 service.** If you expect to use an 800 number for intrastate and interstate dial-in data or voice, you will actually need a different number for the callers in your own state or dial area.

Dial-Up Services of Specialized Common Carriers

Understanding the potential benefits of use of specialized common carriers, as well as the potential problems, can be achieved best through an example of their use. Prospective users contact a representative of the carrier and are provided with a special access number for the geographies from which they intend to originate calls. They are also provided with an **authorization/billing code** which verifies their right to use the system and identifies the account to which the call is to be billed.

To use the system, the user calls the **local access number** (a local call in most cases, but charged as a message unit by the user's phone company). The access number may be busy, in which case the user must repeat the procedure. If the access number is free, the user enters the authorization code (five or more digits), followed by the called number, including area code. This entire process requires the entry of as many as **21 numbers**, one of the most frequent complaints of the users.

Let's leave the user for a moment to see what is happening with the carrier. The user's access number is a hunt group answered by a computer owned by the carrier. The computer reads the authorization code and verifies that it is valid, then reads the number to be called and seeks a route over one of the carrier trunks to that area. If the carrier does not offer service in the called area, the user is informed and the call disconnected with no charge except the message unit charge for the local access to the carrier's computer. Nearly all major metropolitan areas are covered by all specialized carriers.

The user, having called a served number, is now connected to the called party pretty much as usual, except via the trunk facilities of the common carrier. These may be the equivalent of the AT&T trunks, or even superior in quality. They may also be noisier, with more echo and distortion, or with satellite delays.

For a data communication user, dial-up services offered by a specialized common carrier can reduce costs by 50 percent or even more, but the use of the service must be evaluated:

- Does it serve all the areas to be called?
- Are the desired savings achieved in the time of day in which the calls are made, or must calling patterns be changed?
- If the calls are autodialed, can the auto-dial handle the longer number and the intermediate dial tone or **go-ahead** signal of the carrier's computer?
- Will the higher incidence of busy signals disrupt the computer/communication activity by holding up the start of a job?
- Will operations personnel use the system if they have authority to dial direct if the specialized carrier facility is busy?
- Is the application likely to fail to cost justify if the rates of the service increase? AT&T's divestiture of the Bell Operating Companies and its relationship with the specialized common carriers are likely to eventually affect the rate structure of the specialized carriers. If an application is economical only assuming a continued discount of up to 50 percent over AT&T rates, it should probably be closely examined.

Alternatives in Leased Circuits

Leased lines are the traditional vehicle of data communication. Voice-grade leased circuits are available nearly everywhere in the United States in either **two-wire half-duplex or four-wire full-duplex form**. AT&T Communications calls these lines **Type 3002 circuits**. Leased lines are **analog**, voice paths capable of being used as tie lines or for other analog applications as well as for the exchange of data. **Modems are required** for sending data on these links. Leased lines may also be ordered with **special conditioning** to improve the quality of the line and reduce data errors. Conditioning comes in two basic types, **C and D**, and each type affects a different characteristic of the line. The type of conditioning your line would require is a function of the type of modem used and the speed at which data is sent. It is best to follow manufacturer recommendations on conditioning, since failure to condition a line as required may cause **excessive errors**, and conditioning when it is not needed can add significantly to operating costs.

Leased lines are also available from the specialized common carriers, and leased circuits other than voice-grade are available from many sources. The basic alternatives to telephone company leased, voice-grade lines are:

- **Leased, sub-voice-grade service** from a telephone company or a specialized common carrier.
- **Leased voice-grade service** from a specialized common carrier.
- **Leased wide-band service** from either a telephone company or a specialized common carrier.
- **Digital transmission services** from either a telephone company or a specialized common carrier.
- **Satellite communication service**.

Leased sub-voice grade circuits. Although many users think of voice-grade lines as the basic grade of service, the telephone companies and some specialized carriers offer **sub-voice grade lines** capable of transmission at approximately 300 bps and below. These lines are known within AT&T Communications as the **1000-series**. For users who have an application demanding low-speed transmission, savings can be achieved through the use of sub-voice grade circuits. Where several slow devices are located in a single location it is generally **more economical** to use a conventional **voice-grade line and multiplexing**.

Leased voice-grade circuits from specialized common carriers. Most of the specialized common carriers providing alternative long-distance dialing services also provide voice-grade, leased-line service. The service areas in which this service is available match the service areas of the long-distance dialing services, and are **frequently less expensive** than their equivalent AT&T service. Savings, however, are **unlikely to approach** those achieved in long-distance dialing services, and the economic base of the service seems to be in question. At least two of the specialized common carriers have declared a moratorium on new leased-line installations pending a settlement of some of the relationship issues arising from the deregulation of the Bell System.

Wide-band analog transmission. Users with a very high-traffic volume between two points may want to consider a wide-band circuit as an **alternative** to voice-grade lines. These lines can be divided into **multiple voice-grade channels**, and the resulting multiple circuits allocated to voice, data, or other communication purpose. Wide-band analog service is easily adapted to any form of data and is generally available in more locations than the high-speed digital service. It is also almost universally **more expensive than the digital equivalent** since it is less efficient in its use of transmission facilities. The most common form of wide-band service is the **48-KHz AT&T Communications 8000-series** or its equivalent. If the user has no need to mix voice

Carrier Services—When To Break With Tradition

or other analog communication with data and can secure digital service in all locations, **digital transmission** would be **preferred**.

Digital transmission. Digital transmission of information is best known in the form of **AT&T's Dataphone Digital Service (DDS)**, now part of AT&T Communication's ACCUNET Digital Services, but it is also available from other carriers such as Western Union. DDS is based on a backbone of time-division multiplexed channels called **T-carriers**. There are various types of T-carriers, the **most common being the lowest level, T-1**. T-1 provides users with **1.544 million** bits per second (bps) of transmission capability, divided into **24 channels of 56K bps each**. The balance of the data rate is taken up by control information. Further subdivision is possible; AT&T Communications' DDS channels can be acquired at 2400, 4800, or 9600 bps in addition to the 56K-bps rate. DDS is a digital system and inherently **more immune to errors** than conventional analog transmission, but it is not entirely error-free and contains no error detection or correction features. Use of digital paths **eliminates** the need for **modems**, a significant cost savings, and digital circuits at 56K bps are **less expensive** than wide-band analog. Unfortunately, digital service is not available in all areas of the country. Digital may also be a problem if voice and data are to be mixed. Some customer-purchased T-1 products have the ability to **mix voice and data**, but the capability is increasingly rare as the line speed decreases to 9600 bps, and below this limit the intelligibility of speech is degraded. Voice data to be transmitted over digital links must be **digitalized** via a pulse-coding technique and restored to analog on the other end of the connection. Digital voice normally **requires a minimum of 24K bps of bandwidth** to retain speech quality, and **32K bps or 48K bps is generally needed** to **match the quality** of an analog line.

Satellite carriers. Most of the services described so far have been based on microwave transmission between towers on the surface of the earth. Terrestrial paths are becoming **increasingly expensive** as the cost of the equipment increases and the availability of space for the relay towers decreases. Satellite transmission provides a significant improvement in long-distance communication economics by eliminating the multiple microwave relays. It also introduces a unique problem in the propagation delay. Satellites used for communication are placed in what are called **geo-stationary orbits**, a condition in which the speed of the satellite in orbit keeps it located over a single point on the surface of the earth at all times. This equilibrium is reached at an altitude of **approxmiately 22,300** miles. A message from a satellite user is transmitted from the user to an earth station, up to the satellite, is received and retransmitted (called **transponded**) back to another earth station, and eventually received by the destination. **Eventually** may mean after a delay equal to the time it takes radio waves (or light) to travel the 45,000 or so miles (the path is a triangle, as shown in Figure 8), about **300 milliseconds**. Although this may not seem like a long time, data communication via terrestrial circuits never had to contend with such delays, and many of the older protocols are unable to operate efficiently over satellite connections. Some satellite carriers such as American Satellite, RCA/Cylix, and Satellite Business Systems have provided a form of **satellite delay compensation** to help **immunize the user** from the effects of the **path delay**. This compensation is very effective for the batch form of user data, since the information is collected in bulk for transmission and human intervention is not required. In interactive operations, the human review of the results of a transaction is delayed by the round trip of the message in addition to whatever other CPU-related delays may be present. Satellite users may attach to an earth station via a leased-line of some type or may elect to install their own earth station; propagation delay is not affected by the means of attachment.

Evaluating Basic Path Services

Path services have several common elements; they are all charged based on a combination of **distance and usage** in connect time, and they all provide a **guaranteed** pipeline of a specific capacity (the data rate) which the user is unlikely to fully utilize but which is available should peak volumes do so. As such, they generally impose the fewest restrictions on user operation and have the greatest tolerance for errors in estimating volumes or failures to predict growth accurately. The forgiving nature of the services, coupled with their wide availability, has made them the choice for the majority of users. A business wanting to use basic path services should select the particular service based on the normal criteria of **cost**, **availability**, **performance**, and **reliability**. The following questions will help in relating the needs of the user to the characteristics of basic path services:

1. **What forms of information will be carried by each circuit?** If the path is used strictly for data, digital transmission is an alternative. If voice, facsimile, or other more exotic information forms are to be sent, it may be necessary to use a strictly analog service.

2. **What areas must be served by the carrier?** Voice-grade service is available nearly everywhere. Sub-voice grade lines for data transmission, wide-band analog, and digital lines are **less likely** to be available. Satellite links generally have the lowest likelihood of availability unless the user is prepared to purchase an earth station. Obviously, the chosen method must serve the areas required.

3. **What capacity must each circuit possess?** Data lines can be measured in bits per second. For voice and other type lines, a count of the circuits required is the total capacity. If the capacity required for transfer between two points exceeds that of a single voice-grade circuit, it **may be desirable** to use wide-band facilities such as T-1 carrier. Mixed voice and data over T-1 circuits is possible with purchased equipment and may save a significant amount of money. Data capacity requirements over 9600 bps may justify 56K-bps digital facilities. It is **generally uneconomical** to use analog wide-band where **only data** is to be sent because of the **high cost of modems** and the **cost of the basic wide-band service**. Be sure that voice and other analog circuit requirements are verified—although there may be 20 **possible** simultaneous calls between City A and City B within a business, it is improbable that it will be economical to provide facilities to serve such a number. Many PBXs have the ability to provide records of the calls to a given point, and analysis of these can help define a level of service which dedicated facilities can economically support.

4. **What protocols are used for data transmission?** If satellite circuits are to be considered, the protocols used by any data users must be analyzed for **efficiency in long-delay paths**. IBM's popular Binary Synchronous protocol is **so inefficient in a satellite environment** that it may fail to operate altogether.

Basic Path Services—Who Needs Them & Who Needs Them Not

Although there is no substitute for a specific analysis of requirements, the following types of users are probably **best served** by basic path services of the types indicated:

- **Users who mix voice and data or voice and facsimile from a single workstation** probably should consider voice-grade circuits from either their telephone carrier or a specialized common carrier.

- **Dial-up data, voice, or facsimile users who make mostly local or short-distance calls** probably can best use voice telephone circuits provided by their telephone carrier or a specialized common carrier. If the number of long-distance calls is very small, such users would be best served by their existing telephone facilities.

- **Leased-line users with a few point-to-point or multipoint circuits using polled protocols** such as IBM bisync and having interactive data forms probably will be best served by leased voice-grade lines from their telephone carrier or a specialized common carrier.

- **Leased line users with a high volume of data traffic between two locations** should consider digital service at 56K bps or T-1 service. If the data is sent with a modern protocol such as SNA or X.25, or if the satellite carrier can provide delay compensation, a satellite path may save money.

- **Users with high volumes of voice, data, and/or facsimile traffic between two points** should consider wide-band analog circuits or, if the volume justifies it, T-1 carrier service.

The following users are probably **not best served** by basic path services:

Carrier Services—When To Break With Tradition

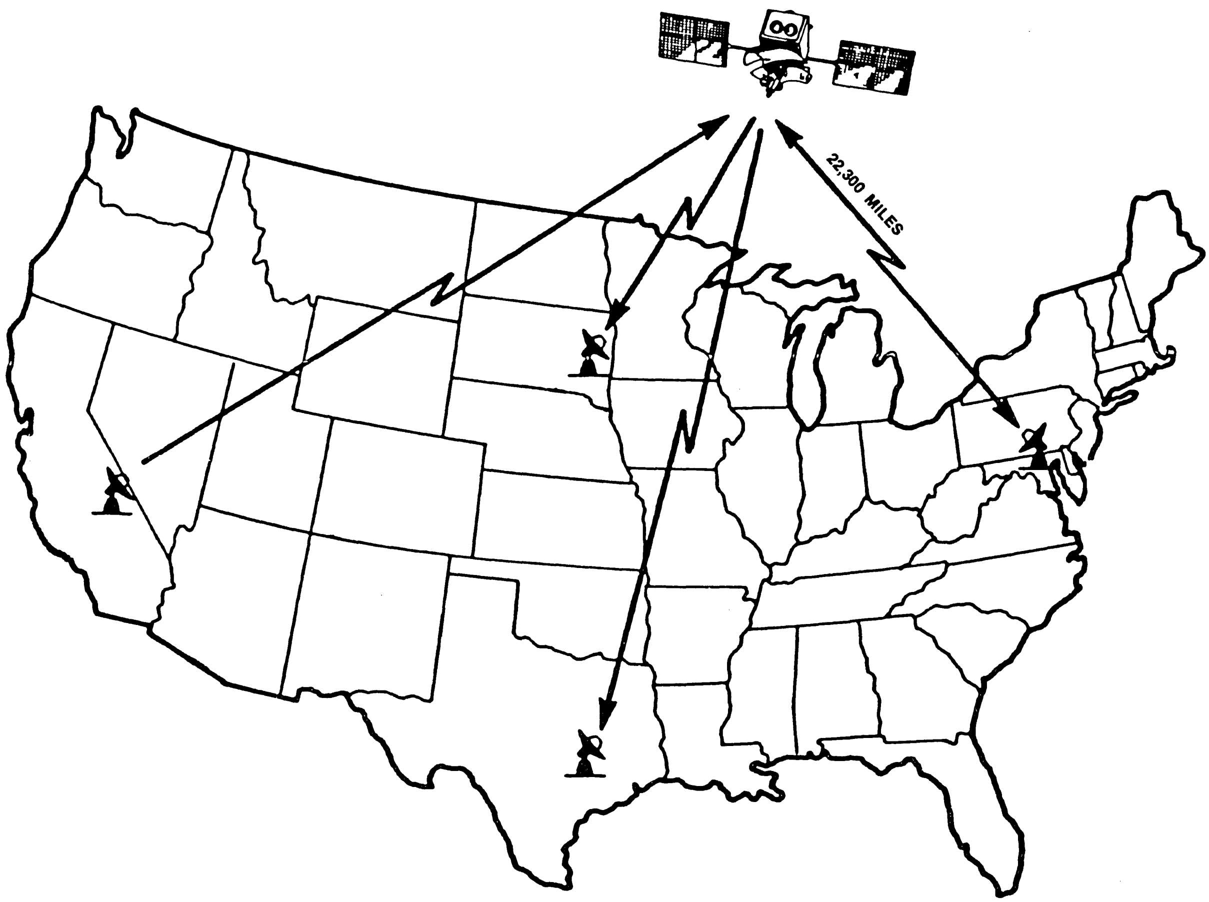


Figure 8 • satellite path geometry (drawing courtesy of RCA American Communications, Inc).

1. Users whose data connections consist of **relatively short but frequent calls to or from a wide geographical area**.

2. Users who require **access to international circuits** for data exchange.

3. Users who serve a data population with **many incompatible terminal types**.

4. Users who must transfer information to other users who are **often not available** for communication.

The following sections of this report describe carrier services intended for the users who fit into these categories.

■ INTERACTIVE SWITCHED DATA SERVICES—PUBLIC PACKET CARRIERS

Public packet networks have been widely publicized recently as interest in the international standard protocol X.25 has mounted. There is no doubt that some form of **packet network is the most efficient way to transfer data** (Figure 2), and more users are looking seriously at public packet or other switched computer-based networks as a means of reducing transmission costs. Packet networks normally charge only by usage, measured by characters sent per second or by packets per second. Users of dial ports may also pay connect charges based on connect time. **In no case**, however, **is distance between users a factor**. This can be of significant benefit to users who have widely scattered terminals communicating with a single host. It may also aid companies who sell services accessed remotely by customer terminals or computers by eliminating the need to charge higher rates to distant customers.

Many communication suppliers who have in the past offered only basic path services or message services either have introduced or are planning public packet networks. AT&T Information Systems, the unregulated arm of AT&T, offers its **Net 1000 Service, a public packet network with significant value-added features including processing and storage**. The **IBM information network** is essentially a public packet network, although it is **based on IBM SNA** rather than on a packet protocol such as X.25. RCA, Western Union, and other companies have or are developing their own public networks based on packet switching.

Packet or other types of data switches differ from the basic path services in that the **packet switches transfer data in the form of data**. The telephone system may perform many of the same switching functions as a packet switch, but the telephone system is building a path by switching in circuit elements to provide continuous resources from source to destination. A packet or data switch accepts data from a user and routes it as data over shared facilities toward the destination. The **basic element** of data within the network is the **packet**, a data segment normally about 128 bytes in length. User **packets are transferred between** network computers called **nodes** which collect the data and hold it until space on a trunk line to another node is available. No part of a packet network can be said to be dedicated to any particular user. Thus, all users are in continuous competition for resources along their routes. This competition assures **high levels of resource utilization** and thus **lower per-user cost**, but it also ensures a **high level of network delay**. It is not uncommon for a packet to take over a second to pass from sender to receiver.

Users attach to packet networks in various ways. Basic asynchronous service can be obtained via a dial-in port which

Carrier Services—When To Break With Tradition

may either be public or private for the user. Other protocols may be connected to the network via leased lines or via a communication node leased or purchased by the user and dedicated to the user's attachment. These **engines** are normally required for attachment of a computer to the network but may be avoided if the computer is **capable of using X.25 and is certified** for operation with the network. **Figure 9** shows packet connection options.

Connection of protocols other than asynchronous or X.25 may cause problems because of the packet delay. If the protocol uses a block-by-block acknowledgement system for error detection and correction (such as that used by IBM bisync), the **propagation delay** of the network **may introduce delays of seconds** between blocks of a message, causing timeouts and increasing transmission times to unacceptable values. Any such protocol must be **converted** to a network-efficient protocol at the packet entry point to be economical. Users of these delay-sensitive protocols should be **sure** that the network is converting. **A test of performance is highly desirable.**

Another potential problem with packet usage is the connection protocol to the packet network. Most public packet networks would prefer that users employed the **international standard packet protocol X.25** for attachment. Use of X.25 allows the network to route the packets to their destination with a minimum of processing. If other protocols such as asynchronous or IBM bisync are used, the network must build packets from non-packet data at the sender end and create the original data stream at the destination. This **packet assembly/disassembly or PAD** process is expensive to the network and annoying to the user. Problems with user interaction arise from the need for the network to stop holding data to build packets when the user sends a logical end of message and awaits action from the opposite station. If the network does not forward a packet with the balance of the message because that packet is incomplete, the user will not see a response.

Discussions of PAD operation relate to costs as well as convenience. The PAD user may select from a group of **PAD parameters** a particular set called a **PAD profile** which operates to make the application as transparent to the network as possible. Improper selection of these parameters can result in **very short packets**. If the particular network used charges by the packet, the **cost** of communication **can more than double**.

Because packet switches are computer-based, many offer value-added services which take advantage of the power of the communication nodes to process information. Some of the services available are:

- **Enhanced billing and accounting information**, similar in form to that provided from a PBX.
- **User connection security** in the form of access passwords, restricted access to hosts by definition of closed user groups, or barring of incoming or outgoing connections.
- **Message storage** to allow two users to exchange information without being connected to the network simultaneously.
- **Temporary file storage** within the network for data transfer between computers with incompatible storage media.
- **Network resident processing power** available to the user on request for large jobs or special functions.

Evaluating Packet Switches & Packet Switch Services

Packet switches are cost savers. That is the conclusion of most of the satisfied users of packet-switch networks. Prospective users should therefore consider economic advantage to be the primary motivation for using packet switches. This means that evaluating the potential for cost reduction is the highest priority.

Packet switches reduce costs in two ways: **data-associated billing** and **distance independence**. Data-associated billing means that the charges for use of the network are related to the

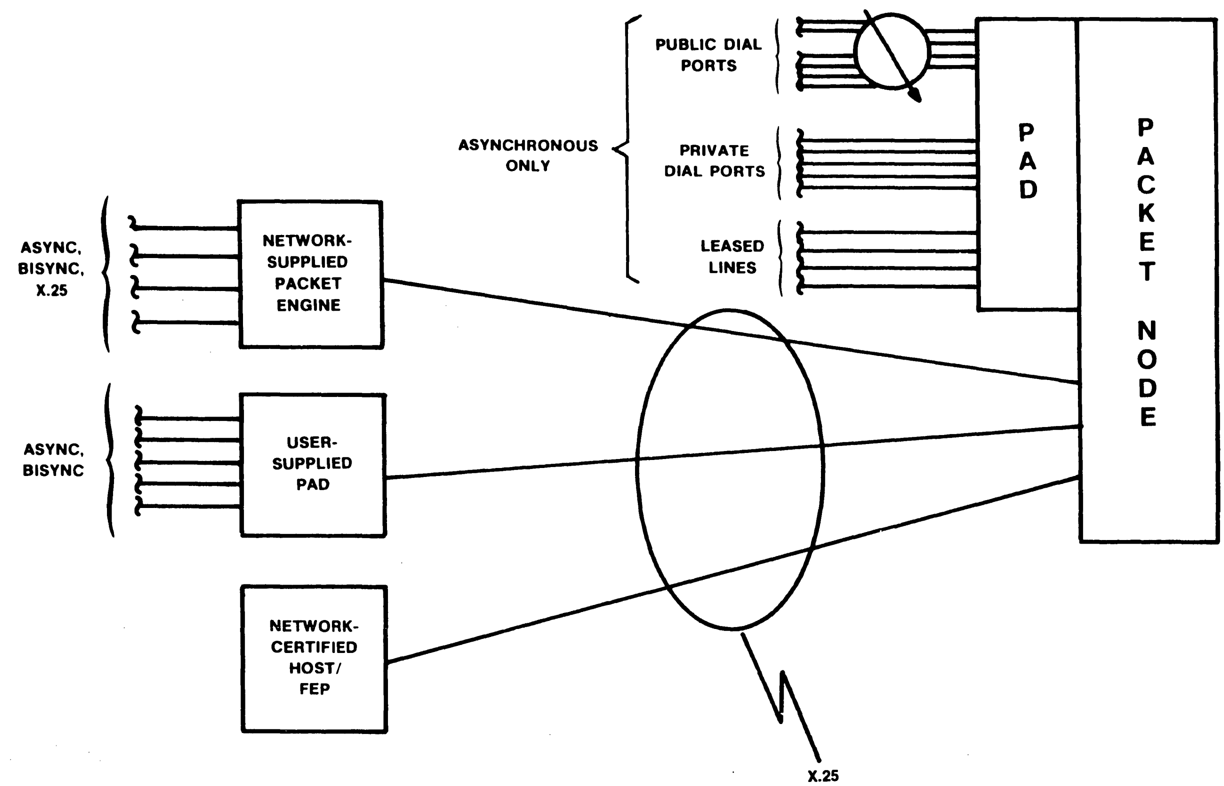


Figure 9 • connecting to a packet network.

Carrier Services—When To Break With Tradition

volume of data sent and not the time spent connected. Distance independence means that the costs are not proportional to the distance between users. These billing practices favor businesses whose **users are distributed nationally** and whose use of the network **does not involve massive exchanges of data**. It may also be advantageous for users whose primary terminal population is, for example, on the west coast and whose computers are too far to the east to consider packet networks. The overall cost of packet switching is lower than that of basic path communication such as dial or leased lines when the **distance between users exceeds about 1,000 miles**, but the exact value depends on the usage. Packet network suppliers publish schedules of rates for user review, and their representatives will quote specific situations if the data volumes and connection times can be described. There are some factors which will affect packet economics significantly:

• **Terminal attachment** to a packet network may be via public dial port, private dial port, or leased line. Such attachment is normally possible at a modest charge. Host computers, however, must attach either via a packet engine leased from the packet network or through a certified X.25 communication facility on the host computer or front-end processor. Either alternative is likely to be expensive, so—**don't neglect the cost of host attachment in evaluating packet switch benefits.**

• **Premiums charges** are assessed by many networks for synchronous protocols such as bisync or SNA. If you have several protocol options, select the most economical.

• Although **protocol conversion** may be available in some networks, all possible conversions are not supported by all suppliers. Be sure that the combination of protocols required for your connection is supported by the network or you may have to purchase your own protocol converter.

• **Study the effects** of adjusting the packet assembly/disassembly parameters on the cost of sending data as well as on the operational characteristics of the connection. Sometimes the cost of having **just the same response as before** isn't worth the cost.

Interactive Switched Data Services—Who Needs Them & Who Does Not

If an individual evaluation of the economics of packet- or interactive-switched networks is not possible, the following applications can be expected to provide reasonable economy:

• Low-speed asynchronous data collection from multiple terminals at distances of over 300 miles.

• Replacement of dial circuits for low-volume, low-speed usage at distances over 200 miles.

• Inquiry applications which require constant connection but little actual data exchange.

• Applications requiring exchange of information among users of different protocols **if the network can supply all conversions.**

The following are probably **not** good applications for interactive- or packet-switched networks:

• **Applications with high data volumes**, particularly if the information is exchanged 10 or more hours a day (leased lines will probably be less expensive).

• **Dial-up applications** where the number of **calls is very low** (the cost of host attachment to the packet switch will erase any cost advantage).

• **Leased-line applications** where the **distance is less than 500 miles** or dial-up where the distance is less than about 200 miles.

• **Synchronous protocol applications** where the network does not perform protocol conversion (propagation delay within the network will drastically reduce efficiency).

■ SPECIAL FORMAT NETWORKS—MORE VALUE ADDED

A variation on the interactive-switched network is the special format network. This network is custom designed for a specific purpose or protocol and may offer compatible users significant benefits in both cost of service and features and facilities.

Most communication via public networks is a **compromise** between the need to provide an **economical** means of exchanging information and the need to **adapt** to the **specific protocols** required for user attachment. In public networks such as packet switches the network attempts to adapt the data forms at entry so that efficient transmission is achieved. This adaptation may make it difficult to use the network with advanced and therefore complex protocols. For example, IBM has only recently supported X.25 networks as transmission media for SNA data. Special uses of networks such as **mailbox** message applications or **graphic display applications** may likewise be possible over conventional public packet facilities but could be more efficiently handled if the network were optimized to the specific use.

Protocol-sensitive networks are probably the most significant form of special-format networks. IBM's Information Network has a Network Services feature which permits SNA or bisync users to communicate between terminals and hosts, or host to host. Since the network is **designed** to operate with the SNA or bisync protocol there is **no performance degradation nor problem with timeouts**. If the user desires, IBM will obtain all lines, modems, and connecting devices as a part of the Extended Network Attachment option.

Another example of a protocol-specialized network is **Graphnet**, a network offering of Graphnet, Inc, a subsidiary of Graphic Scanning Inc. This network is designed for use with asynchronous devices operating at 1200 bps or less. Because it is targeted for low-volume async users, there is **no distinction** between host and terminal attachment. **Low-volume users** can thus achieve economic benefits restricted in packet networks to users whose volumes of data could justify the higher cost of host computer attachment. This network also offers a form of special-purpose network by recognizing that low-speed async users probably will not remain connected at all times, so data storage for later delivery is desired. Graphnet, also called the **Freedom Network**, is reportedly attracting an increasing number of microcomputer users who find the **combination of low speed and storage** of information a useful aid in the application of microcomputers to large businesses.

Another form of special-purpose network is the **video network**, used to transmit video information for applications such as **teleconferencing**. These networks provide users with a complete video system including voice and data communication, freeze-frame video for chart and slide presentations in addition to normal video for movement-oriented presentations, interactive voice, electronic mail, and high-speed facsimile. An example of this network is **Macomnet**, an offering of M/A-Com, Inc, and a sample configuration is shown in **Figure 10**.

One of the most significant problems associated with the special-purpose networks is locating one for the purpose the user intends. Because these **networks have a limited universe** of potential users they are often **not nationally advertised**. One of the first steps in any analysis of communication alternatives should therefore be investigating the existence of a special-purpose network. Leads on such networks can be obtained from the following sources:

• **Hardware vendors** for equipment which also serves the special purpose. For example, the best source for protocol-sensitive network information is the manufacturer who uses the protocol.

• **Controlled circulation publications** targeted to the industry or application.

• **General survey publications** describing a wide range of communication products and services.

• **Trade shows** which display products and services in the application or technical area involved.

• **Other users** who have similar product needs, contacted directly or at conventions or shows.

Evaluating Special Format Networks

Special format networks are more difficult to set general evaluation rules for. Once one or more have been identified, the user must determine first whether the services provided will **meet the needs** of the application and second whether the **costs are**

Carrier Services—When To Break With Tradition

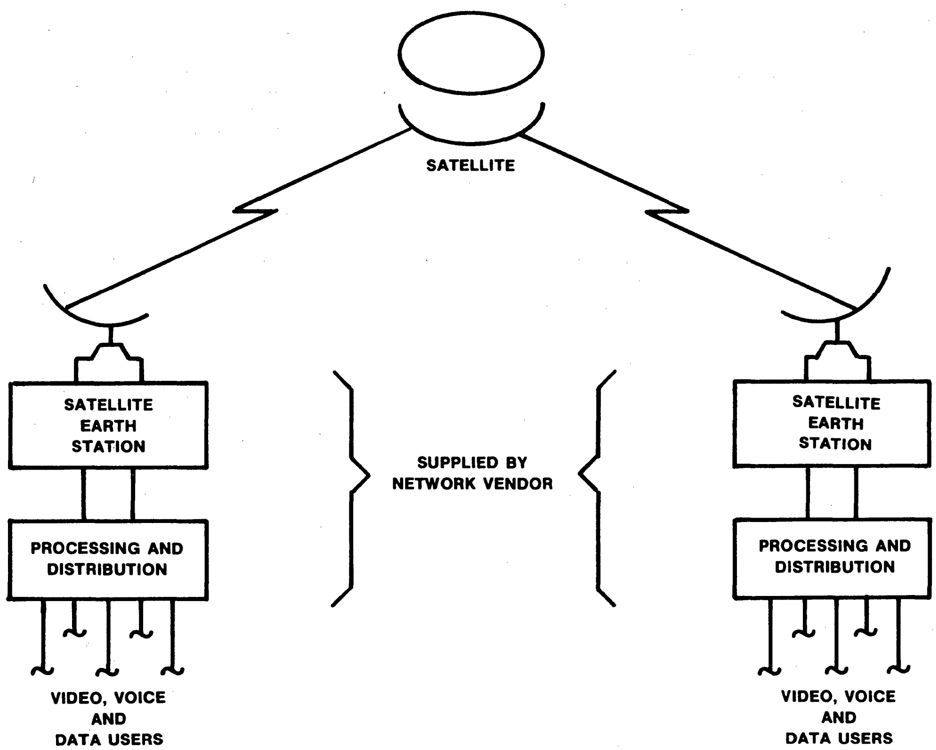


Figure 10 • a video-satellite network.

acceptable. With special format networks it is important to run a test with actual information if possible, or to see the network in operation under similar circumstances.

■ MESSAGE SERVICES—DATA MAIL INSTEAD OF DATA CALLS

Traditional data communication is interactive in nature. Both sender and receiver are connected to the transport facility at the same time, and the information received can be acted on and a response generated as a part of the same data **conversation**. A part of this interactive orientation probably comes from the dependency of data communication on the telephone system; a two-way path exists so why not use it? But all user information exchanges are not interactive in nature. The discussion of data forms earlier in this report defined **message-oriented services** such as electronic mail which did not require interactive dialog between sender and receiver.

Once the initial biases against one-way conversations have been overcome, it is easy to identify advantages to message services. First, the delivery of messages is almost always more economical because the network has great latitude in scheduling messages through its nodes. At any point where resources are scarce, messages can be stored and sent when the lines are free. This easy allocation of resources translates into lower costs. A second advantage is the ability to service communication between users who cannot connect simultaneously. Message service networks, as shown in **Figure 11**, resemble packet networks with storage capability.

Message services for data communication may require some specific application planning, and may be difficult to apply to existing applications. One of the tests for interactive applications discussed earlier was the test of the involvement of the communication in the **completion** of a transaction at a remote location rather than its involvement in the **recording** of the transaction. An example of this can be developed using inventory systems. If a customer walks up to a sales position and says, "I want to order a widgit," it may be possible to **take the order** without requiring a computer access as a part of the order. Does the company sell widgits? Is there any reasonable chance that one could **not** be ordered? If the company is a single product company or has relatively few products, chances are that widgits can be ordered and that no computer action is really needed to do so. On the other hand, if the company has thousands of products or if some specific delivery commitment is required ("I need a widgit in two weeks, do you have one?"), it may be necessary to access computer records to complete the sale. In the former case, a sales record of the transaction could be sent through a message system to be collected by the computer at any convenient time, and in the latter an interactive inquiry is probably required.

Another use of message services is **text communication**. Some of the data exchanged in a typical business has a specific format and meaning to a computer, a message containing stock number, price, and quantity ordered for example. Other information is intended to communicate ideas or information in more or less human form. This textual information may actually be best communicated in **message format** since the human receiver may not be able to assimilate it at computer speed. An obvious example of this is the result of an account status inquiry. If the report in lengthy, it will probably be needed in hard copy form for manual review. If that is so, **why not consider sending it as a message**? Having a wealth of information at your fingertips can be a heady experience, but it may have little or no business benefit. Studies have shown that complex inquiry results lead to a high incidence of repeat inquiries because the users forget information provided early in the response or cannot read the data at the display speed. If the output is printed rather than displayed on a CRT, it is certainly a candidate for message transmission.

Interactive inquiry and instant results have become conditioned into many applications which cannot justify the incremental cost, but the selling of a message alternative may be difficult. This is

Carrier Services—When To Break With Tradition

especially true where some form of interactive system was previously available. Some of this resistance can be eliminated by studying and perhaps changing the workflow associated with the items involved. Transactions handled on an interactive basis tend to be **reaction processed**; all of the functions associated with the transaction will be completed before the operator turns to other activity. This focus on a single event until it has been handled can be a benefit in some types of businesses, especially where direct customer interaction is taking place. But in other cases, functions are more efficiently handled if they are **function processed**, where items are collected and run through a common function in a batch. Posting items to a ledger is often most effective if it is done in **function** mode because setup for the job is reduced and the probability of error may be less because the worker **trains into** the job.

Message services also have special features available. Some networks allow users to send multiple copies of a document to different users by sending the text once and providing a list of addresses. Message services may also make it possible for multiple users to pick up a copy of a single document by providing **non-destructive delivery**. A user who lost a copy of such a message might be able to pick it up again without requiring that the originator resend it. Delivery of a message may also be scheduled for a specific time, even if the destination is online. This can prevent having an important and confidential document arrive at its destination in the middle of the night and sit in the printer, highly visible, until the following morning.

Although most message service results in delivery to a printer or hardcopy terminal, many computers have packages available for sending and receiving messages, including Telex/TWX. This permits the receipt of text information by a computer for storage and later editing or display. Message services administered by computers in multiple locations serve as the basis for electronic mail that is easily extended to suppliers and clients who are equipped with compatible facilities.

The Economics of Message Services

Most message services charge by the **character** or **word**, and a short message (about 60 characters) can be sent anywhere in the United States for about thirty cents. Connection to other countries is available at low rates as well. Sending a business report of 2,000 words to London via a message service would cost about fifty dollars, while the same message as data at dial rates would cost over a hundred. Access to message networks is similar to access to public packet networks, public dial ports and leased lines are available.

Do's & Don'ts of Message Services

Message services are not without their pitfalls, both in use and in estimating the savings. Here are some tips:

• **Find out what the charge basis is, by example if possible.** Some services charge by the time it takes to send the message to the network (so the faster, the better!), some limit you to a maximum number of words in that interval, and some charge by character. Analysis of the pricing strategy in light of the characteristics of your data can prevent costly errors.

• If you elect to dial in to the network, **be sure that the network will not disconnect you after one message**. Redialing can be frustrating, especially on a heavily used public port.

• If quick access to the network is necessary, **be sure that public port availability is high** or use a private port or leased line.

• **Don't purchase add-on features or services** unless they are independently cost justified. A major segment of savings can be eroded away by providing "full feature service" to business users who really don't need it and probably won't use it.

• **Take a close look at any special accounting services** available. They may make internal allocation of the costs of another resource, such as computer time, easier by identifying the users of a system and the level of use.

• **Beware of broadcasts!** If someone must enter each copy of a message to be sent, it deters mass mailings over the message service. If a single keystroke can dispatch a document to a thousand destinations, it probably will do so **all too often**. Message services make this **too easy** for some users, so **internal control may be necessary to keep costs in line.**

• **Some message networks charge for data left in the network beyond fixed limits.** Be sure that you don't **lose** a message and pay for its storage for months. This can be a problem in applications where a user can dial in and pick up messages. Some networks will notify you of these undelivered messages. Others will **simply bill you.**

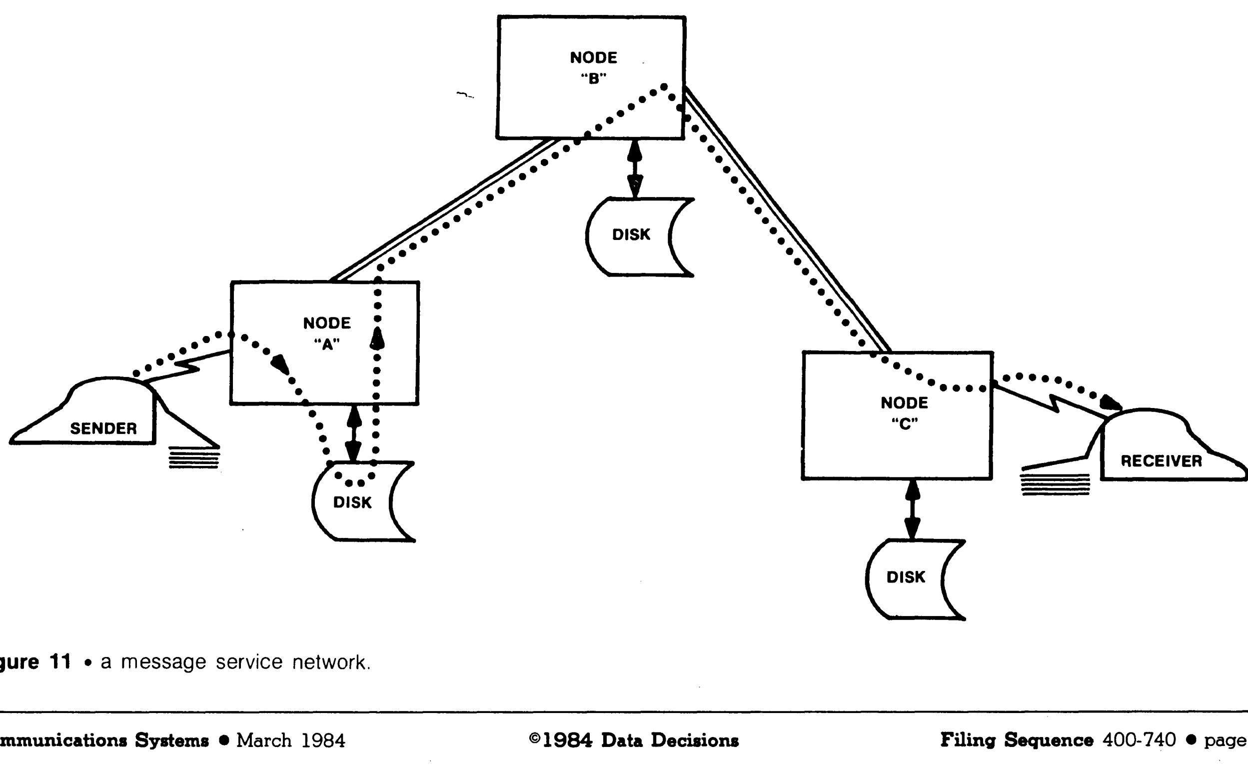


Figure 11 • a message service network.

Carrier Services—When To Break With Tradition

■ PROCESSING SERVICES—MY COMPUTER OR YOURS?

When communication networks provide such services as protocol conversion and store-and-forward messages, it is only a small step to making the computer power and storage used for these functions available for more generalized user application. Remote computing (online) services have long combined data communication and data processing, at least from the perspective of the user. The processing services of a communication network extend the remote-computing nature of timesharing to that of a **floating computer** which resides within the network and takes on many of the characteristics of a user's own host. More complex data storage, again a characteristic of remote computing systems, is also provided with these network processing systems.

Relying on a network for data communication is enough of a risk to many users that the thought of using one for file storage and data processing might seem slightly short of madness, but among the many companies who have taken the lead in the processing services market are the technological cream of American industry; **AT&T and IBM**. AT&T had announced some enhanced value-added network products to be marketed by its fully separate subsidiary Information Systems. Originally known as **Advanced Communication Service or ACS**, this network would have placed Information Systems in the **value-added public packet business**. In its evolution from that first announced form, ACS gained a new name (**Net 1000**) and considerable processing and data storage capability. **IBM's Information Network (IBM/IN)** has a kind of reverse evolution. As it was first proposed, IBM/IN was a remote processing and storage resource which supported program development for users whose own resources were committed or not yet delivered and served as a source of high-powered processing to smaller users who could not justify owning resources of that scale. IBM/IN was to be based primarily on IBM SNA, with accommodations for bisync users as well. The Information Network became a communication network with the addition of Network Services, and some specialists have suggested that the entry into data communication carrier service was prompted by AT&T Information System's inclusion of processing and information services into Net 1000.

It is unlikely that companies throughout the country will flock to disband their data processing facilities and embrace network-based resources. First, the **cost of large-scale use** of the storage and processing features of either network **would be greater than that of dedicated user facilities**. Second, the **risks of being isolated** from the computer and databases by a **network failure**, whether a real risk or a speculative one, will prevent some users from adopting the new services. But for the small user, the combination of processing, storage, and communication could be almost **revolutionary in its impact**.

When AT&T filed the initial request for FCC approval of the ACS service, it cited a **low-entry threshold** as one of the **prime objectives of the service**. IBM, in defining the Information Network, could hardly claim low-entry thresholds with a primarily SNA product, but did **stress** its role in helping professionals **develop, test, and implement applications, convert new systems, and balance peak-load computing requirements**. Despite the obvious differences in what each company would consider a **small** user, both networks are aimed at the user who is in need of supplementary computer power and storage, especially during the critical start-up phases of an application.

Communication applications are normally capital-intensive because both processing and transmission resources are required. When a new communication application, or even a conventional batch application, is being developed, it is often necessary to acquire the computer resources and a portion of the communication channels very early to support the development itself. This early acquisition of expensive equipment can place a project in the red for years, and in some cases eliminate the justification for an application which, on **operating resources alone** would be financially sound. Both Net 1000 and IBM/IN offer users an **alternative in network-based resources**. Programmers use the computer power and productivity aids of the network computers to develop an application **without significant investment in equipment**. Communication resources are allocated to the development as needed; testing, parallel, and initial low-volume production. At any point where the economics justify it, either the processing or the communication can be removed to user-owned facilities. If desired, parts of the application can remain on the network indefinitely.

AT&T Information Systems' Net 1000 is based on communication/processor nodes structured as shown in **Figure 12**. Each node consists of some number of DEC VAX 11/780 computers and IBM Series/1 computers, used as communication processors. The backbone of Net 1000 is **ACCUNET Packet Service** (formerly **Basic Packet Switching Service, BPSS**). This service has had its difficulties in negotiating the regulatory course. The initial offering was rejected by a staff report of the FCC on the grounds that it was obviously structured and priced only for the use of Information Systems. The full commission, while overruling the staff report per se, nevertheless rejected the application on a legal technicality. Information Systems then answered both the surface and underlying problems by correcting the filing data and repricing the service to **make access at lower costs and lower speeds more practical**.

The IBM Information Network is based on IBM communication controllers and mainframes located in various cities and acting as communication switches for the central processing point in Tampa. The **network is based on IBM's SNA**, but will also support attachment through IBM's older binary synchronous protocol. **IBM/IN has four distinct parts**: productivity and program development services, a data bank of current IBM product information, network services to allow the IN to be used as a communication network, and extended network attachment to extend IBM's communication responsibility to include the lines, modems, etc used to attach to IBM/IN.

Who can use a processing services network such as Net 1000 or IBM/IN? Nearly every user could in theory benefit from one or more of the services these networks offer. It is important **not** to think of this type of network strictly in terms of a **transmission alternative**. IBM/IN, for example, offers extensive graphics support which in itself could justify a user attachment, if the organization is a heavy user of color graphics services for presentations, documentation, etc. But IBM/IN is still **basically a network for IBM users**. Net 1000, on the other hand, has **protocol support for an estimated 80% of the communication user population**. Currently, Net 1000 supports asynchronous Teletype emulation or IBM 2741 devices, polled IBM 2740 asynchronous devices, contention bisync for the 2780/3780/3275, and polled bisync compatible with the 3270 family of terminals. SDLC/SNA support is to be available at a later date. Flexible interchange between this universe of devices is accomplished by defining three communication modes. **Common Mode** provides device-independent communication by defining a common subset of functions and features which are mutually supported by the unlike users. This form may restrict the use of features such as cursor positioning on CRT terminals. **Class-specific Mode** is used where the devices are compatible, and permits the full range of features which the devices support, including cursor control and special editing functions. **Transparent Mode** allows users to send information without process intervention by Net 1000 except to recognize control requests such as Interrupt and Disconnect.

Using remote processing capabilities to their best advantage requires that application design be conducted to exploit these opportunities. Although good modular programming practices would normally dictate that the transaction editing, transaction processing, and response generation functions of a communication application be separate modules, users often allow **expediency** to **muddy the boundaries** of the actual program tasks during implementation. The **separation of communication tasks from processing tasks is necessary if the user wishes to have an orderly path of evolution** from a development system supported completely on the network's resources to a system supported in whole or in part on the user's own equipment.

Applying Network Processing & Storage Services to an Application

Both data processing and data storage facilities within a network can be classified as interactive or background. Interactive

Carrier Services—When To Break With Tradition

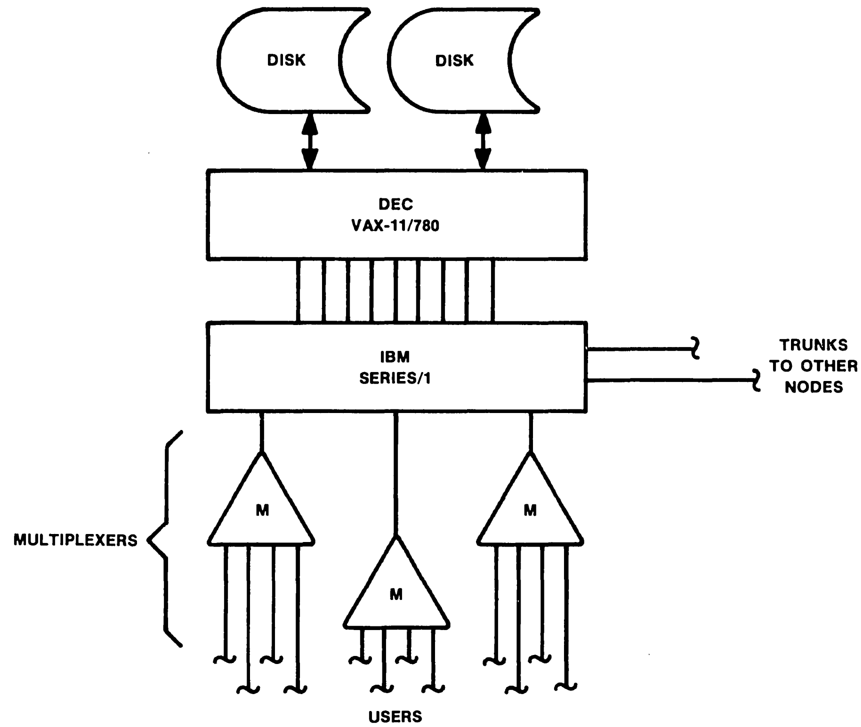


Figure 12 • Net/1000 node.

facilities are available at the time when data is being exchanged with the network and may respond to or be used directly by the communicating devices. For example, interactive programs can edit a transaction as it is entered, and interactive storage can hold the completed data for delivery or processing. Background storage would normally be used to hold master files or databases used for inquiry, while background processing would be used to update a master file with accumulated transactions. The following points can help identify opportunities to use batch or interactive services:

• **Applications** which could otherwise be considered message application, but which require simple data entry editing to assure correct processing of the collected data, may be **candidates for interactive processing and storage services**. The user defines a program to receive and edit the data, and that program communicates with the user and writes the result to network storage. The destination user may then connect and collect the edited information.

• **Background storage** may be used in conjunction with the transmission facilities of the network to **supply help lists** to remote users without loading the user computer facilities or the transmission facilities with the data.

• **Background storage and processing** can be used to **back up** the application programs of a user where critical processing tasks might be interrupted by user facilities failure, or to handle overflow processing when facilities are busy.

• **Interactive or batch storage and processing** can be used to **convert** the format of data for exchange between unlike host computers. A database file might be unloaded into the network and processed by a background program to convert data formats, then reloaded onto a different database system running on a different processor.

• **Background processing** can be used to **stand in** for a host computer by executing a more primitive form of processing.

Credit authorizations, for example, could be handled from a file of basic limit information if the main computer credit system were down or fully occupied with other tasks.

Figure 13 shows an integrated application using network-based processing and storage.

The Don'ts of Processing Services Applications

There are some specific pitfalls in the use of network-supplied storage or processing services:

1. **Don't get tied into expensive alternate storage or processing facilities.** Very few users would benefit from the indefinite use of network facilities to support an application. If you elect to use such facilities, regard the step as a transition to use of your own facilities and **provide a plan for migration to private facilities, properly supported by the application design, at the very start of the development of the application.**

2. **Be sure that the data and program security features of the network protect sensitive information.** A simple error in the definition of the access rights of a file can make an entire credit list available to any user who can request credit authorization, or even allow those inquiry users to update the list!

3. **Try to analyze the costs of transmission facilities and processing facilities separately.** If the use of a processing-services network increases transmission costs over other alternatives, be sure that the services it provides are cost justified. Also, don't tie in communication use of such a network and database or inquiry use. If you need the IBM/IN for graphics support, for example, you don't necessarily have to use it for all your data transmission as well.

4. **Be wary of applications which barely cost-justify in operation on a processing-services network but which require specialized programming to utilize the facilities of the network.** If rate structures change, the application may actually lose money, and the relative youth of these types of networks makes them unusually vulnerable to such changes.

Carrier Services—When To Break With Tradition

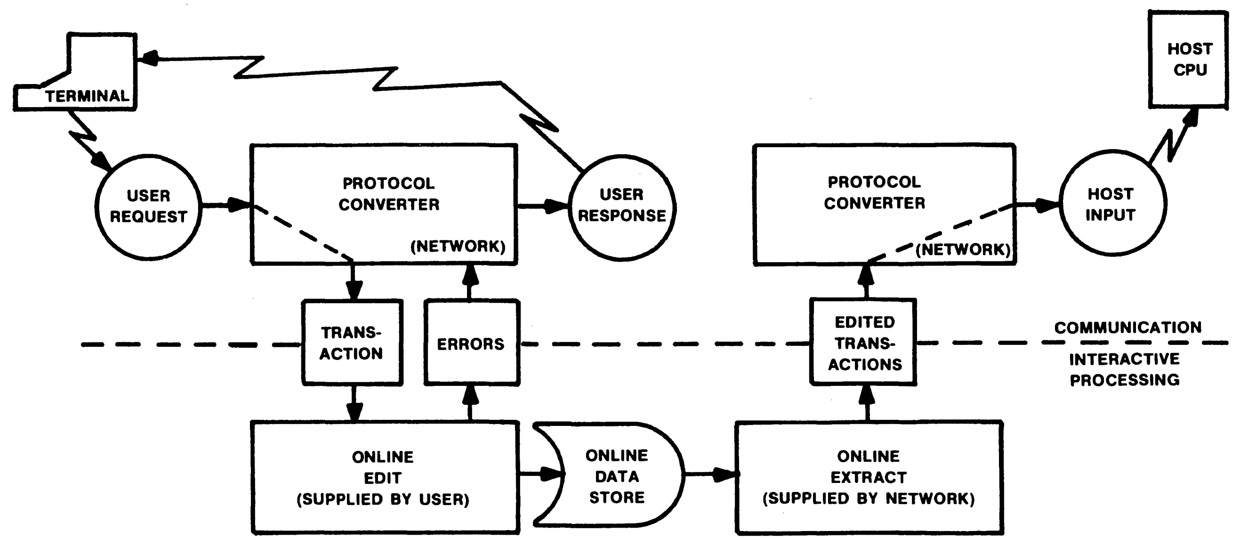


Figure 13 • an example of using network processing services.

■ SELECTING A TYPE OF COMMUNICATION SERVICE

Evaluating the alternatives for data transmission is becoming more difficult as the options increase and regulatory and economic uncertainties continue. Organizations with staff personnel expert in communication applications and facilities should undertake a study to **match the characteristics of any new application to the technical, economic, and support parameters of each communication alternative**. Smaller users should either enlist the support of a communication consulting organization, allocate a long period of vendor study and application analysis, or take the relatively safe route of using basic path services provided by either the telephone company or another common carrier. **If your organization does not have the expertise to evaluate the application and alternatives effectively, don't try to find the leading edge of communication technology.** There is no shame in admitting that a primarily non-computer business would not have the background to perform a communication analysis.

Users who plan to undertake communication studies without experienced people on staff should consider the following guide to evaluating communication services:

1. **Determine** the parameters of the application, as described earlier in this document.

2. **Digest** the information into several charts and tables and take this data to your computer vendor for advice.

3. **Take** the proposals and advice of the vendor, together with the original data and a technical description of your equipment, to representative vendors for each type of transmission service, starting with the telephone company. Get comments, routes, proposed equipment, and above all **cost estimates** from each.

4. **Analyze** the costs and approaches of the previous step to select one or more preferred transmission technologies. Refine your description of needs on the basis of the technology you prefer and ask for quotes from all vendors in the field.

5. **Select** the vendor with the best Cost/Benefit match for your application.

Collecting & Reporting the Characteristics of Your Application

Communication vendors cannot properly quote your application without statistics on it. They will require that you provide transaction information, response time requirements, and geography of the network. They'll also need to know about any equipment already identified for use in the application, including its operating speed and protocol. Here is a suggested format for the information:

1. **Network map.** Draw a map showing each location served by your proposed network (**don't show the lines you think are needed—that's the vendor's job**). Also show the path of any communication **resources you already have** and indicate how that resource is **presently used**. If you expect to use it in the new application, indicate how. Give each data source or process point a unique identification (a code based on a structure of **T** for terminal and **C** for computer, plus a unique number is satisfactory). **Figure 14** shows a sample map.

2. **Traffic chart.** Provide a chart which shows the traffic from each identified point to all other points with which it exchanges data. **Figure 15** shows a suggested format. Be sure to note the following information:

- peak and average transaction volumes
- message structure of each type of transaction (a five-character request yielding a 200-character display, all 200 characters sent back as an update, for example)
- distribution of transactions by hour, if the volume is time-variable
- response time requirements identified for the transaction
- expected growth of volume over the life of the system

3. **Equipment chart.** Indicate the make, model, and options of

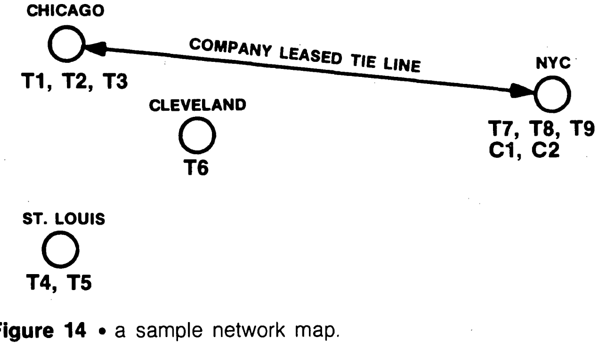


Figure 14 • a sample network map.

Carrier Services—When To Break With Tradition

each communication device (terminal, CPU, front-end processor) by **map device key**.

4. **Application overview.** Describe the operation of the system in **simple terms**. For example: T-1 and T-2 are used to collect inventory data from the warehouse and transmit it to computer C-11. Each time an item is removed from the warehouse, one of the terminals is used to record the item code, authorization order number, date, and time. Current stock status for the item is returned to the terminal.

Presenting Your Application to Vendors

A key point to remember in dealing with either your own computer vendor or prospective vendors of communication services is that the **probability of your getting valuable and accurate input from them is proportional to the ease with which they can evaluate your application**. One major corporation recently issued a **Request For Proposal** in the form of raw traffic volumes, clerical procedures, and statistics about the current manual operation. There were no responses. The vendors contacted reported that the effort required to make a proposal under those conditions was out of proportion to the potential profit on the contract finally awarded. **Vendors do want your business, but there are limits to their ability to provide consultation and design recommendations without direct compensation.** To get the best value from each of your contacts with vendors, apply these rules:

1. **Select the vendors to contact carefully.** Use a research service, vendor survey, references, or other sources to build your vendor list. Unless you are **sure** that a given vendor is a prospective supplier of the type of service you desire, call a representative and check.

2. **Give responding vendors as much latitude as possible** in the format of the response, especially if the value of the service you are purchasing is relatively small in proportion to the business typically handled by the vendor. Don't expect a major vendor to respond to a fifty page RFP for five thousand dollars a year worth of transmission services. If you do issue a formal RFP, spend some time in design of the request and of the expected response. The ideal is to identify key issues for specific attention and define a general format for a presentation of the vendor's own approach to the problem. Be especially wary of fixed form response requests if you are evaluating a wide scope of possible solutions; your format may unintentionally favor one alternative.

3. **Give vendors a chance to present their proposals individually.** Group vendor conferences are barriers to an effective dialog because vendors will fear theft of their system concepts by others. Many users feel that these group meetings assure a free exchange of ideas to the benfit of the user, but more often they result in suspicious silence. Use vendor conferences only when many vendors are involved and the contract price is high enough to assure their attendance.

4. **Encourage innovation on the part of the vendor.** Let any new suggestions or variations in approach be proposed in a positive atmosphere. This is especially true in the preliminary dialog with your own computer vendor and the dialog with a "target" vendor representing each major alternative technology. You are seeking alternative solutions, not endorsements to established concepts.

5. **Make each vendor provide prices in a way which relates to the application.** Don't accept raw price data, and make sure that any assumptions which the vendor makes in the calculations are identified. Otherwise you may find that the prices you have relate only to operation of the system between midnight and six in the morning.

6. **Get historical price trends.** If a service has increased in cost by fifteen percent for the last four years, it will probably continue to do so. The price trends should be factored into future-year operating costs.

Analyzing Alternative Transmission Technologies

Once you have prices and proposals from at least target vendors in each category, you can evaluate the costs and benefits of the transmission technologies represented. Here it is necessary to be hard-headed; if you specified that a response time of three seconds was required and all proposals meet that requirement, don't weigh heavily in favor of a vendor who offers two-second responses. Will that extra second save money or improve productivity or sales? If not, forget it.

Here are the steps recommended in evaluating proposals:

1. **Review the proposed solutions** and discard any which do not meet one or more specific requirements identified for the system.

2. **Make up a chart** which lists the vendor, the type of solution proposed, and the cost. See if there are any patterns in the responses. For example, do all of the public packet vendors propose a higher-than-average price? If any technology is universally out of the average price, it is probably best to discard it.

3. **Rank the conforming vendors** in increasing order of cost. Start with the least expensive and answer the following questions:

- are there any elements of the solution which seem undesirable?

	TO C-1 COMPUTER NYC					TO C-2 COMPU		
	TRANS PEAK	TRANS AVER	PEAK PERIOD	CHAR/ TRANS	RESPONSE CLASS	TRANS PEAK	TRANS AVER	PEAK PERIO
T1	22	8	8-10 AM	220	CRIT	0	0	
T2	10	2	8-10 AM	100	ROUTINE	0	0	
T3	0	0		0		18		
T4	30	18	10 AM - 2 PM	41	CRIT	0		

Figure 15 • a sample traffic chart.

Carrier Services—When To Break With Tradition

• are there any elements of the solution which seem uniquely beneficial?

• is the vendor's reputation for support and honesty satisfactory?

• what elements, if any, do **you** think in the vendor's proposal might justify the cost differential between this vendor's bid and the lower bids?

4. **Now, review the results** of the evaluation so far. Some vendors may be dropped based on identified weak points in their solution which are not offset by cost advantages. Others may have enhanced services which do not justify increased costs. Try to cut the list in half at this point.

5. **Select the lowest remaining three vendors** if they all represent the same transmission technology. Otherwise select the lowest two of each technology which seems satisfactory. Use this new list as the basis for individual vendor conferences, contract talks, and eventual selection.

Evaluating Support—A Most Important Intangible

One of the most **difficult issues to evaluate** in selection of a communication transmission technology or a specific communication carrier is that of **support**. Support issues include vendor assistance in identifying problems, probable frequency and length of network failures, assistance in installation and optimization of facilities, and costs for service.

Users have traditionally justified selection of the telephone company as a communication vendor on the **single point of contact theory**; if all of the network products and services come from one source, then only one call gets the correct agency to fix the problem. Users entertain themselves at conventions and seminars by trading stories of the time that the entire XYZ system was down for three days because the wall outlet was broken, and nobody would test it out because it wasn't their equipment. There is some truth in the assertion that the chances of two vendors agreeing on which is causing a given problem is small, but there is increasing truth to the more modern complaint that even a single vendor will refuse to admit responsibility either. **Users must ultimately accept the fact that some familiarity with their own network and its vulnerabilities is the best insurance that problems will be promptly identified and corrected.**

Does anyone really offer single point of service any more? Yes, with qualifications. Many of the network suppliers will take the responsibility of ordering the lines and equipment needed to attach to them and will accept problem calls on the circuits as a whole. The question is whether they will be more effective than the user in identifying the problem and contacting the responsible party. Before you accept support of this type as the key issue in selecting a vendor, talk to users of the vendor's service and find out if there are any real advantages. What about the telephone company in its traditional role as the sole supplier? Deregulation and divestiture will undoubtedly affect the ability of AT&T Communications to offer users the **free assistance** which some areas have experienced in the past. In general, the tariffed services should increasingly be regarded as commodities with little service or assistance bundled with them, and the free-market unregulated areas are the best place to obtain effective single-point-of-contact service.

A Checklist for Selecting the Best Technology

Users who wish to get a preliminary idea of the best technology for their application can do so through the following checklist:

1. **What protocols will you use to communicate?**
   - Asynchronous
   - X.25
   - SNA
   - Bisync

2. **Select the data form which best describes your application.** (See the earlier section of this report on **Need** if you have any questions.)
   - Interactive
   - Batch
   - Mail

3. **What is the average distance which separates your users from the processing point?**
   - Less than 200 miles
   - Between 200 and 500 miles
   - Between 500 and 1,000 miles
   - Over 1,000 miles

4. **What is the average number of characters per hour which you will send to a terminal or to a computer, whichever is larger?**
   - Less than 6,000 per hour
   - Between 6,000 per hour and 60,000 per hour
   - Over 60,000 per hour

Now, review your answers. If you plan to use bisync or SNA as a protocol, you probably should look at a **special-format network** such as IBM's Information Network or AT&T Information Systems' Net 1000 in any case where the following suggestions mention any form of network **other than** basic path service. If you use X.25, you should **consider public packet under all circumstances—it may pay with relatively high volumes and low network mileage. If you selected the mail** form of data in the second question, you should probably **consider a message network** unless your volume is high (60,000+ characters per hour) or your distance is low (less than 200 miles). In that case, basic path service will probably be better. If you selected **batch** as your transmission method, **public network services** such as packet switches **may be best unless** your **volume is over 60,000** characters per hour or your **distance is under 500 miles. In interactive applications, distances over 500 miles will almost always indicate an opportunity to save money with public packet if the response times can conform.**

A last issue in selection is one of **flexibility**. If you have an application which has little room for adaptation and have few communication specialists on staff, **look carefully at any alternatives to terrestrial basic path services.**

■ FUTURE TRENDS IN CARRIER SERVICES

The primary question in the minds of both communication users and vendors is the effect of the break-up of the Bell System and the deregulation of AT&T Information Systems. The only point that can be made with assurance on this topic is that the pricing and features offered by communication carriers through the mid-1980s will be **subject to unusual variations** as the shock waves in the industry subside. There are some points on which many experts agree:

• Prices for services will tend to move to more accurately reflect the cost. This may cause problems for some of the specialized common carriers who are relying exclusively on quirks in the tariffs and subsidies to price their services under those of AT&T Communications.

• The fragmentation of the Bell System with the resultant eradication of end-to-end service, dichotomy of responsibilities among AT&T and BOCs, pricing uncertainty for services, and diversity of product offerings among AT&T and BOCs emphasizes the necessity of users to seek viable alternatives.

• The requirement that the BOCs return customer-premisis equipment to AT&T effective January 1, 1984 has sparked the sales of telephone instruments to subscribers in many states. Modems and other more exotic facilities will follow. The resulting independence of the BOCs from AT&T equipment opens them as a market for other companies producing AT&T-compatible or equivalent modems and other phone and switching equipment. This will further undercut the traditional **all-AT&T** communication position.

• The entry of AT&T Information Systems into the value-added network arena, and its competition with IBM there and elsewhere, is likely to legitimatize alternatives to basic path services in the minds of users. With significant influences such as AT&T and IBM

Carrier Services--When To Break With Tradition

promoting the advantages of these networks, application design to retain compatibility with packet concepts will advance, and public data networks could become a significant force in data communication by the end of the 80s.

Digital Termination Systems & Local Loop Alternatives

The total body of data communication literature dealing with the so-called **last mile local loop issues** would probably stretch several of these **last miles**. There is no doubt that this last extension of the telephone system is the most difficult to address with modern technology. Recent interest in **Digital Termination Systems (DTS)** which use broadcast microwave techniques to transmit wide-band information, be it voice or data, to users is reflected in the thirty-plus applications for such service pending with the FCC. **DTS and cellular radio** offer potential for eliminating the maze of wiring that fills conduits beneath the streets in most cities, and the frustration of users trying to get new service or have service restored after the inevitable back-hoe cuts the cable. But the capital expenses required to make such concepts widely available are huge, and economic conditions are less than favorable. Users in metropolitan areas should **track the progress of these technologies**, but should not wait for them.

Communication and data processing share many common elements. As computers became more complex and the features became more difficult to understand and to employ effectively, programs were developed to use these features within higher-level structures such as special languages or application packages. This eliminated the need to seek out more and more esoteric programmers. It is probable that this phenomena will be repeated in data communication. The easiest path to market for complex transmission features and services is in a **protocol-independent integrated network**. Users of such a network can be **immunized** against satellite delay, bird flocks in the microwave beam, and back-hoes in the roadway without even being aware that a problem exists. It is possible that these types of networks represent the safe haven for users that the traditional and now very much changed AT&T has represented so effectively for so long.

• END

Communication Test Equipment— A Solution To Increased Productivity

An Introduction to Communication Test Equipment, How it Can Resolve Your Network Problems & Eliminate Excessive Downtime & Guidelines to Select the Test Equipment That Best Serves Your Needs

■ INTRODUCTION

Data communication is the link between the user of information and its processing/storage facility; failure of the communication path is likely to cause a major disruption in a business and loss of productivity. Many user networks are designed for partial immunity to such failures, either through the use of high-reliability components or through redundancy. Such measures may dramatically increase costs, and may actually fail to provide sufficient improvement in uptime to justify the expense. Many users feel that a combination of proper testing and diagnostic procedures coupled with selective sparing of equipment and well executed maintenance agreements is a better guarantee of network availability than redundancy. For smaller users, it may be the only economically feasible solution. There is a considerable amount of specialized test equipment available today, and users could easily spend their entire communication budget on just a few pieces. Even a selection of seemingly affordable test equipment can fail to maintain network availability if its use requires technical skills not available within the user organization, or if it is not integrated into a master strategy of fault detection, isolation, and correction. This report explores the subject of network failures and the use of test equipment to prevent, isolate, and correct network faults to achieve maximum uptime.

■ NETWORK FAILURES—SOURCES & IMPACTS

Communication failures tend to be defined in terms of almost mathematical relationships between the phase of signals at various frequencies, line noise, amplitude distortion, and so forth. A user considers a communication failure in a simpler light; **the failure to communicate**. To a user, particularly the night shift supervisor of the communication center, a mathematically exact description of the failure is secondary to getting it fixed! This attitude is understandable in light of the statistic which indicates that the primary cause of the loss of communication between points lies in the user-controlled equipment and the interface cable. Communication lines represent the second largest problem, and communication equipment such as modems and multiplexers are third.

If you have just spent a considerable sum of money to back up a communication circuit, these facts may make you pause. Increasing reliability through the development of networks which are tolerant of failures seems logical in that it takes the problem diagnosis procedure, often a complex task which many users are a little apprehensive of, out of the critical solution path. Increasing network reliability as a tool in increasing uptime is normally a good idea, but like most it can be carried too far. Just how far to take reliability improving measures is a function of the **risk of failure** against which you are protecting yourself. Not the type of risk but the **chance**, the probability of failure. Problems in a communication connection fall into the following groups:

- Operator or procedure errors
- End-point failures of equipment, terminals, cables, etc
- Environmental/power failures
- Communication line failures
- Communication equipment failures

In any communication center each of the above faults has a probability of occurence. Proper development of a high reliability network begins with addressing the risks in order of their probability and probable impact on the network. It makes no sense to provide duplicate communication links against a line failure which is experienced only once every 3 months when a power loss in the computer room is occuring every 10 days. Each major area defined above has an average interval of failure, called the **mean time between failures**, and an average interval for failure recovery called the **mean time to repair**. **MTBF** and **MTTR** measurements, even if they are approximate, are the first steps in planning for a more reliable network.

■ FAULT TOLERANCE VERSUS FAULT RECOVERY

Most communication budgets do not contain an indefinite amount of money to be spent to assure network availability. Users can spend what they have on spares and backup equipment, on test and diagnostic equipment, or a combination of the two. How much should be allocated for each? Each expenditure must be related to the **improvement in MTBF/MTTR** which will result.

Figure 1 illustrates a basic communication network consisting of 3 computer centers and the communication link which join them. The communication operations manager has reviewed the statistics on the failure rate of the network and derived the following:

1. **The computer system** at location "B" fails due to a software fault about once every 6 days and requires a half hour to restart.

2. **The terminal room** at location "C" is subject to overheating in the summer, requiring that it be closed down. This happens about once a month during the May-September time frame.

3. **The communication link** "B"-"C" demonstrates undefined failures about once a week with no indication that external or end-user equipment is at fault.

Unless the communication manager has responsibility for the computers and terminals as well, this scenario indicates that there are significant problems in other areas of the business. First, without provoking an interdepartmental conflict, the communication manager should define the tolerance of the business to the first 2 problems. If management feels that **they're just things we have to live with**, then they have established **a level of acceptable risk**. The communication manager is responsible for reducing risks which **can be controlled** such that they do not significantly increase the total risk of failure.

Given these facts, the communication manager can address the problem of the line between locations "B" and "C." In the first place, a failure a week is almost certainly excessive even in the presence of other higher probability failures. Rather than moving to correct the problem by backing the line up, the first step should be to complain to the local phone company. If this does not bring the problem to an acceptable level, dial backup or some other measure may be needed. **Figure 1** also shows various fault-tolerant versions of the basic network, introducing backup lines, alternate routes, or dial backup. Selection of the best alternative depends on the cost/improvement relationship. If a line failure clears within 3 to 5 minutes, for example, is it worth the expense to provide dial backup? Manual reconnection of the link via dial backup could take minutes itself if inexperienced personnel were involved.

Some users embrace fault tolerance as a means of eliminating costly and complex network testing and diagnostic equipment,

Communication Test Equipment—A Solution To Increased Productivity

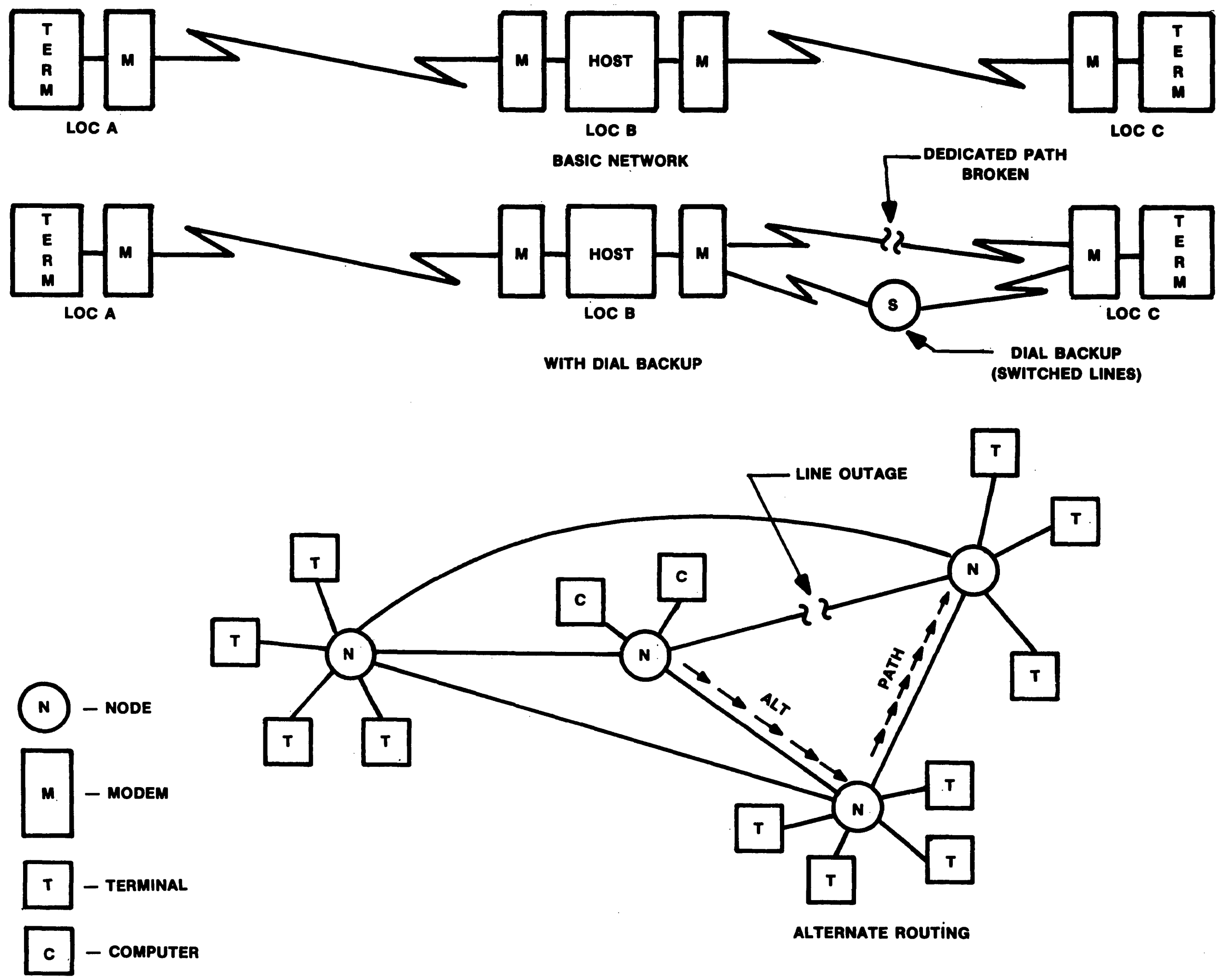


Figure 1 • networks and fault tolerance.

but most find that the application of fallback and recovery procedures must begin with a recognition of the problem, or at least that a problem exists. The least expensive source of such data is the operation itself; most system failures will result in some symptom visible to the operators or users; loss of connection, increased response time, etc. More accurate information may be available from the equipment being used. Most modems display the important EIA control signal states through front panel LED indicators, and some will indicate circuit quality through a digital display or one or more LED indicators. A label affixed to a modem indicating the proper state of the indicator LEDs can be a valuable aid in detecting a problem if the operating personnel are asked to check the lights at intervals, such as shift changes or when any unusual operating conditions are noted.

But is all of this enough for problem solving? If, in our network of Figure 1, the operator of the location "C" center notes a lack of response from location "B," it could be caused by the failure of the troublesome "B"—"C" link or by the failure of the system "B" software. Dial backup will not restore the operation of the computer at location "B." If the total network is to function properly again, the operator at "C" must first have a fallback procedure to cover both the possibility of communication failure and facilities failure, and second be able to tell which has occurred so that proper action can be taken.

Suppose that the operator of the location "B" system trips over the interface cable and disconnects the modem to location "C"? While a dual set of modems and cables would certainly solve the problem by providing an alternate path, it would be cheaper and easier just to plug the cable in again. Some communication problems such as this are correctable by the user, and in such cases the time required to isolate and repair the fault should be considered. It may not be worth providing alternate paths or procedures.

Communication failures can be grouped by cause, as shown previously. They can also be grouped by the level effort which might be expended to correct them:

1. **Uncorrectable within the communication system.** These problems are environmental or user equipment failures which cannot be managed through communication redundancy and cannot be corrected in the network even if identified.

2. **Correctable by a combination of communication and non-communication procedures.** Failure of a data center computer can be corrected if a backup computer at another location exits, and if the communication circuits can be switched to that location.

3. **Correctable through intervention with an outside agency or**

Communication Test Equipment—A Solution To Increased Productivity

by duplicate communication facilities. Line problems can be cured or bypassed.

4. **Correctable by substitution of customer-premises equipment.** Modem failures can be repaired by substituting a spare modem.

5. **Repairable at the customer site.** More complex equipment such as multiplexers may be repaired on-site if maintenance response is adequate.

6. **Procedural.** Tripping over cables, leaving switches in the wrong state, switching off terminals, and other operational problems can be corrected without repair if they can be identified.

A network should be analyzed by defining the types of failures which can occur, categorizing them as above, indicating their probability and repair time, and estimating their impact on the business. Once this has been done, network uptime can be assured through application of procedures to identify and isolate the problem, apply solutions through such redundancy as is built into the network, and/or initiate repair procedures. Although fault correction, like alternate facilities backup, is an option in dealing with a problem, the application of **problem detection and isolation procedures is the first step in all problem management procedures**, and some test and diagnostic equipment for support of that task is all but **inevitable**.

■ BASIC PROCEDURES FOR PROBLEM DETECTION & CORRECTION

A national maintenance organization reports that most user attempts to correct problems fail in the first critical step—**the problem report**. The accuracy of that first report and the correct level of detail support the mobilization of the proper resources to continue diagnosis and move to correction. Lack of attention here often causes the familiar maintenance wild goose chases. Problem reporting should be a formalized procedure, both in the information given and in the point of reporting.

Here are some rules for establishing a proper procedure for problem reporting:

- Each person who might make a report should have a **single** number to call, and should call it whenever an error is suspected. You do not want a terminal operator to guess whether the problem is communication or computer related and possibly start the wrong organization in motion.
- The caller should provide his or her own name, the location of his or her workstation, and a number where he or she can be reached.
- The exact conditions leading to the report should be described according to the time sequence in which they were noticed. Tell what you experienced or did first, then lead to the later reactions. Don't editorialize or guess at causes, and don't report events out of sequence.
- The trouble line personnel should be trained to walk the caller through a checklist to verify current conditions. The checklist will depend on the type of workstation, making accurate identification of operator and workstation important. The checklist should be designed to both gather information and identify possible operational errors such as disconnecting the data cables or turning the terminal to local mode.
- A written problem report should be taken on each call, and the report should be logged for tracking. It's easy to forget a single problem if something else comes up later.

Once a written report is available it may be desirable to notify key people of the existence of a problem. Some users set a threshold of some sort for alerting higher level personnel to the difficulties, but most organizations want the responsible management to be alerted that a fault has occurred through their own chain of command rather than by having system users call. In parallel with this notification, the problem should be analyzed based on the symptoms taken. If desired, the reporter of the problem can be held on the line for this stage. The analysis format will vary with the type of system, but the following guidelines will generally apply:

1. **The primary question is the nature of the problem**, whether it is data processing equipment, human error, or communication facilities. Are other operators nearby still online? Are system users in other areas also reporting failures? What are the conditions of the indicators, if any, on the operator's equipment or on the local modem? If the operator reports that all other users are running normally and the area is served by a single communication line, then the likelihood of a communication failure is minimal. This level of inquiry should result in a first-level deduction on the major system which has failed. Make that analysis based on the symptoms and move on.

2. **Test each guess about the problem** by projecting the observable consequences of a fault of the type you have guessed, then check to see if these projections are in fact true. If you think that a local cluster controller has failed, for example, you should be able to predict which stations are operating and which are not. You might also expect a host error message. Do you have them? Be sure to follow this logical confirmation strategy, because problem diagnosis will get considerably off track if the initial assumptions prove false.

3. **If a theory checks with observable facts**, apply an accurate test. Put a line in loopback test. Key some data with the terminal off-line. Switch a device to a spare line or port.

4. **Save the most complex testing for the last phase of diagnosis**, the checkout of a component already strongly suspected of being at fault. There is a tendency to attempt to diagnose critical problems by moving immediately to complex tests. **Avoid it** unless your operations personnel are extremely technical. Focus on detail tests early in the process can blind people to the overall patterns.

The basic approach outlined above is sometimes called **intuitive** because it relies on a stream of successive deductions which seem based almost on technical intuition. While this strategy will normally produce results quickly where personnel are well qualified, many users have no communication or system personnel. Even organizations with technical support must occasionally contend with absences, so it is valuable to conduct a systematic approach to problems when intuition has failed. There are 2 such approaches; sequential testing and halving and bracketing.

□ Sequential Testing to Solutions

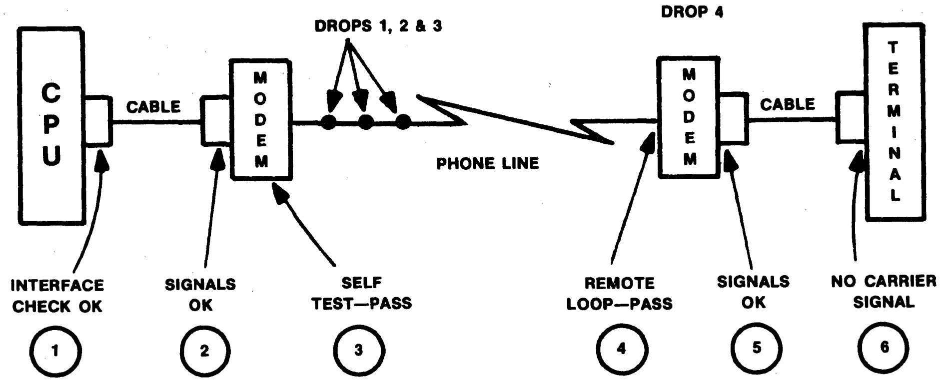
As inelegant as it sounds, the concept of sequential testing as a problem solving technique is long established and very effective. Personnel using sequential testing will start at a known point of proper operation (the CPU end of the connection is usually best) and **sequential test** outward a component at a time, verifying proper operation **to that point** by the appropriate test procedures. If there is a fault in the circuit, the failure will be identified by a sudden loss of communication at a **sequential test point**. **Figure 2** shows an example of sequential testing. The host computer is interacting with a multidrop line, and the line has failed between drops 3 and 4. Since drop 4 is at the end of the line, a line failure cannot be assumed as readily as it might if several terminals at the end of such a line were failing. The sequential testing procedure starts with the host and verifies that it is operating, then moves drop by drop, possibly through modem loopback testing, until finally the failure between 3 and 4 is noted. If the modem at location 4 self-tests correctly, a break in the line can be assumed.

Sequential testing is an easy, systematic procedure which appeals to non-technical personnel. Why would anyone not want to use it? If our Figure 2 circuit had 60 drops on the line, it would take an average of 30 loop tests to find the problem. Try running 30 loop tests with the vice-presidents of several user organizations watching you carefully. At times of high tension, a faster mode of solution is required.

□ Halving & Bracketing

In programming terms, the sequential test method of diagnosis is called a **sequential search**. The problem with sequential searches is obvious if there are a lot of items to be checked (60 drops is a lot), but there is another programming technique for searching which also applies to fault isolation and works better for large numbers—the **binary search**.

Communication Test Equipment—A Solution To Increased Productivity



CONCLUSION: FAULTY TERMINAL CABLE

Figure 2 • example of sequential testing.

In a binary search, start in the middle of the circuit and see if you have a proper path. If you do, the fault is **beyond** that point and if you do not, it is between you and the host. In either case, you move in the direction of the fault by **half** the distance and check again. This process of **halving and bracketing** continues until you isolate the failure to a single element.

Figure 3 shows a 60-drop line with a break between drop 37 and 38. The first test at drop 30 indicates that the failure is beyond that point, so halve the distance and check at drop 45. It fails. Now the error is between 30 and 45, so drop back 8 (always round up) and try at 37. It works, so you know the fault is between 37 and 45. You move up 4 to 41 and try, and that fails so you have isolated it to the 37 to 41 range. You move 2 back to 39 and that fails, so you move back to 38 and it fails also. The failure is thus between 37 and 38. Sound complex? It is a little more difficult than sequential testing but, it took only 6 checks to find the failure in 60 drops. with a little math you can prove that the largest number of checks it can take for any given length is a power-of-two relationship: 16 takes 4 checks, 32 takes 5, 64 takes 6 and so forth. In applying it, use the kind of chart shown in Figure 3 so you don't lose track of the range of possible failures.

□ Summing Up The Basics Of Fault Isolation

1. Get a good report of the basic symptoms to isolate the point of failure.

2. Use an approach to further isolate which is appropriate to the skill level of the people available.

3. Don't get into detailed testing until you have the fault isolated.

4. Track the problem and keep key people informed; they'll call and bother the testing personnel otherwise.

■ THE DATA COMMUNICATION PATH AS A SYSTEM

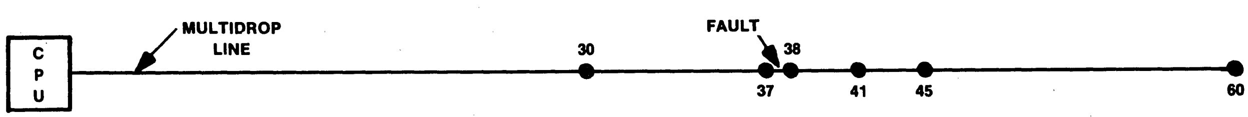
The failure isolation procedure treats the data communication path itself as a system, which it is. **Figure 4** shows the elements often found in a digital data path between a terminal and computer. Your particular connection may not have all these elements, but future requirements could add them to any circuit. The basic elements of the communication path are:

• **The digital interface.** RS-232C, RS-449, V.35 or whatever, there is some type of standard connection between the business equipment (computer, terminal, etc) and the rest of the path. This interface performs 2 functions: it transfers the data in each direction and it allows the equipment to control the circuit. A failure of the interface is a sure interruption of the path.

• **The local connection.** This is the path between the digital interface and any concentration equipment which might be in the circuit. In some systems, this path may be standard cable and in others a modem pair and the connecting wires.

• **The concentration equipment.** If any form of line sharing equipment such as multiplexers is used in the circuit, they form the bridge between the local connection and the trunk.

• **The trunk connection.** This is the main circuit, usually a dial or leased phone line and a pair of modems. It may consist of a local loop to the telephone central office, a long distance trunk, and another local loop at the destination. It may also be a digital link, a satellite path, a value added network, or other circuit.



TEST 1 — AT DROP 30 - LINE IS OK — FAILURE IS BETWEEN 30 AND 60

TEST 2 — AT DROP 45 - FAILURE — BREAK IS BETWEEN 30 AND 45

TEST 3 — AT DROP 37 - LINE IS OK — FAILURE IS BETWEEN 37 AND 45

TEST 4 — AT DROP 41 - FAILURE — FAILURE IS BETWEEN 37 AND 41

TEST 5 — AT DROP 39 - FAILURE — FAILURE IS BETWEEN 37 AND 39

TEST 6 — AT DROP 38 - FAILURE — FAILURE IS ISOLATED TO LINE BETWEEN 37 AND 38

Figure 3 • failure isolation using halving & bracketing.

Communication Test Equipment—A Solution To Increased Productivity

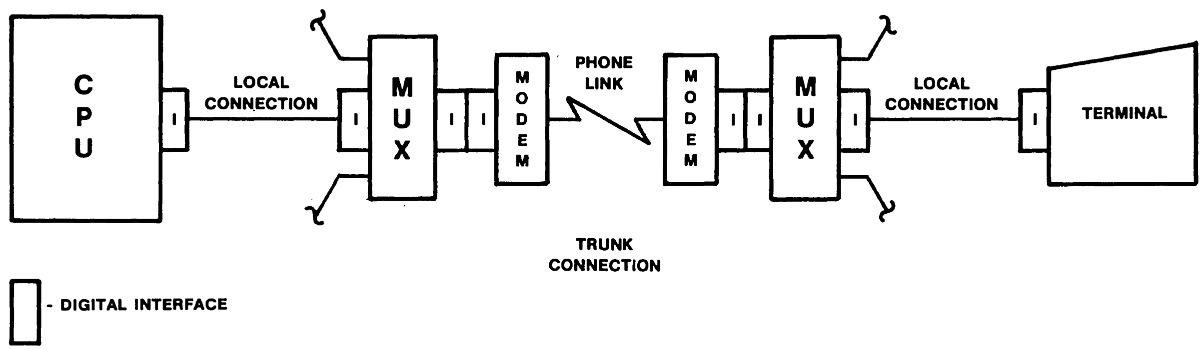


Figure 4 • the digital path as a system.

Knowing the elements of your communication system is important when the principles of risk analysis are applied. The highest probability of failure is in elements of the circuit which humans control or interact with; the end-point equipment and that communication equipment local to the end points. Local paths have the next highest failure rate, and the main trunk lines the lowest. Another important fact is that **the boundaries between elements are the best test points.** It's difficult to get into the middle of a telephone line (unless it is multidropped), but it is easy to get to the connection between the digital path and the analog path via the modem.

Each element of the communication system has a set of test procedures and equipment which helps evaluate its function and detect faults. Very large users may want and need equipment for analyzing the operation of all their communication elements, but the average user with a finite budget must make compromises. What do you really need? Can you get by with the test facilities integrated into your equipment? One way to decide is to review the use, advantages, and disadvantages of the popular test equipment.

■ USING TEST FACILITIES INTEGRATED WITH COMMUNICATION EQUIPMENT

The most common and often most valuable pieces of test equipment are also the most often overlooked—those which are built into other equipment already a part of the network. Nearly all computers and many terminals have some form of indication of the status of a connection, and most modems offer a display of EIA control signals. Test capabilities may be built into all of these components directly (the host computer may have a diagnostic test capability, the terminal loopback settings, and the modem a bit error rate test) or it may be possible to invoke limited testing by taking advantage of operating characteristics of the devices. A terminal with a **local** mode of operation may be tested off-line to the communication path by switching to **local** mode and keying to see if characters appear.

Applying these integrated device tests to your problem solving can save both time and money, but is is important to remember that the objective of network diagnosis is the systematic solution of the problem and not the exercising of all the capabilities of testing built into each component. Don't let the fact that a given device has built-in test capability cause you to jump to it and begin testing as soon as a failure occurs. Local device tests should be employed, but only when your test methodology has brought you to that point in the network. If you are running a **sequential test** and the modem is the next element, by all means use its **loopback and bit error rate (BERT)** test capabilities. **Don't skip other untested elements to get to the modem.**

The CPU is normally a good place to begin testing, especially when sequential testing is performed. The communication package on the host may not provide detailed test capabilities, but it is likely at least to inform you of a failure. Check the operator console of the CPU for the following messages:

• **"DROP nn NO RESPONSE"** or similar messages indicate that a specific device on a multidrop line is not active. If there are other drops on the line and they are not likewise reported, you can assume that the basic communication path is sound and that the problem lies in the unique devices and connections serving the failed drop. If several drops are reported down, look for a **common thread**. Are they at the end of the multidrop line? Are they in areas sharing power, cabling, or other resources?

• **"ERROR THRESHOLD EXCEEDED LINE nn"** or messages of this type indicate that the communication protocol used for the connection permitted error detection and retransmission to attempt correction. The attempts made to correct a bad message were not able to do so, or an excessive number of bad messages which could be corrected were noted. This type of message indicates a **soft failure**, dreaded by most communication troubleshooters because of the difficulty in isolating the problem. This type of message is often repeated if the situation persists, so check to see how often it is being logged. An increasing number of reports per unit time means a component is degrading. If the process is fast enough (rate of message doubles every 2 minutes, for example) it may actually be best to wait a few minutes until the problem becomes **hard**, meaning that it persists and can be more easily isolated.

• **"LINE nn NO CARRIER"** or **"LINE nn CONNECTION LOST"** messages normally mean that the host has detected control signal status which indicates that the path is not operational. This is an indication of a hard failure affecting the entire circuit.

• **"BUFFER OVERFLOW PORT nn"** and similar messages indicate that the data from the line has overrun the system's storage capacity. This condition could be caused by an extreme overload of the system or by a terminal device which is **streaming** data because of a malfunction.

Whatever the communication message presented, be sure to check the documentation for the exact meaning and suggestions on the cause of the problem. Many large computer systems have excellent documentation on failure messages, including suggestions for diagnosing the problem.

□ Built-In Modem Tests

The temptation to run to the modem on all communication problems is understandable, and if your experience in diagnosis makes the **intuitive** method of problem solving effective you may want to do so. Because the modem is the bridge between the digital world of the business device being served and the analog world of the phone connection, it often provides information quickly on the major system at fault. Modems normally display some test and diagnostic information, and the higher the speed (and cost) of the unit, the more sophisticated these test facilities are likely to be. **Figure 5** shows the front panel of a modem with translations of the display LEDs and the switches. If the option is available on the modem, be sure to perform a lamp test before relying on the condition of the LEDs as an indication of operation. Here are some of the valuable test features:

• **EIA control signal indicators.** It cannot be overstated that the state of the interface control signals, sometimes called RS-232C

Communication Test Equipment—A Solution To Increased Productivity

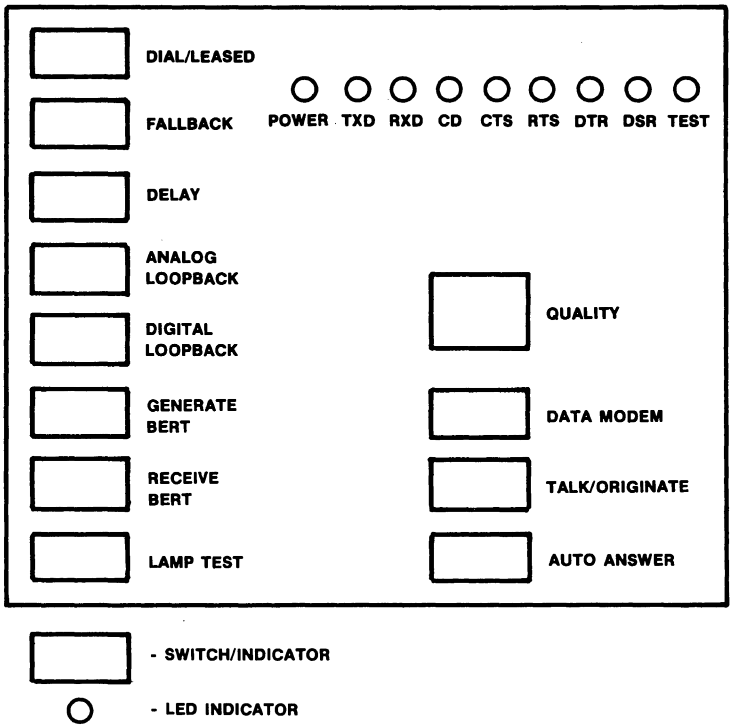


Figure 5 • typical modem front panel.

control signals or EIA signals, are the **most valuable indications** of the overall status of the connection. Most connections will not operate properly without the Data Terminal Ready and Data Set Ready (called DTR or TR and DSR or MR on various systems) indicators lit. The next section will discuss the analysis of the control signals in detail, but there are a few additional points which relate to the modem indicators of these signals. First, be sure to record the normal state of the control signals. **Figure 6** shows the format of a label which can be affixed to the modem and which indicates normal values. Having such a label will reduce the number of times an operator will give your technical specialists the benefit of their long experience at reading signal values by such terms as "the lights are all wrong" or "I think that little one on the left is usually kind of glowing." Second, be aware that the modem sees the signals **at its location**. A cable fault can cause a difference in control signal states between the modem and the host. Third, many modems are optioned to maintain the level of certain control signals at all times, including when there is no connection. Thus, you may see Data Terminal Ready illuminated even when the host computer is not connected to the modem. Be sure that your label indicates when a control signal is forced into a certain state.

• **Modem operating indicators.** As Figure 5 shows, there are often LED indicators on modems which display some aspect of the modem's operation not directly related to the control signals. On the popular AT&T 212-compatible modems used for timesharing and other interactive terminal applications a front-panel indicator can be used to tell the user that the modem is operating in 212 mode (at 1200 bps) or in 103/113 mode (at 300 bps). Many agonized efforts to diagnose the problem "my display is moving real slowly" can be shortened by checking that indicator. Other indicators tell that the modem has fallen back to a lower speed due to line conditions (another nice thing to know, if people are complaining of the response times) or that it is in test mode. Again, the normal state of these indicators should be posted on the modem.

• **Line quality indicators.** When dial-up connections are used, or where leased line quality is relatively poor, it is often very useful to have an indicator of the quality of the connection. This may consist of a digital display of relative quality (be sure you check to find out if high numbers are good or bad) or of one or more LEDs to indicate quality. Most of these displays measure the ability of a modem to **train** to the transmissions of the other, so don't assume that the line is necessarily at fault if you see a bad reading.

• **Loopback testing.** Most modems can perform loopback tests, some of which check the local modem (a self-test), some the line and remote modem (remote analog loopback) and some the entire path (remote digital loopback). **Figure 7** shows some of the loopback options. Loopback tests are usually controlled by 2 switches, one which selects digital or analog loopback and one which causes the modem to generate a test signal (self-test). Some modems can also be looped back for external data transmission. The last example in **Figure 7** shows a terminal performing a data loopback to test the link and its own operation. It is normally best to begin modem testing with a self-test of the modem, then proceed to remote analog loopback and remote digital loopback. You may need the cooperation of an operator at the opposite modem for end-to-end tests. Check your modem documentation for specific test procedures.

• **Bit error rate testing (BERT).** High-speed modems sometimes have a built-in bit error rate test, sometimes as a part of the self-test function. This feature allows the user to check the modem logic for proper bit recognition either locally or end-to-end. If the feature is available, it functions as a self-test when applied to the local modem alone and can be used with digital loopback when performed as an end-to-end test.

□ Terminal Testing Aids

Most of the inexpensive asynchronous terminals are not designed to provide a great deal of test capability. Part of the reason is that most asynchronous protocols do not support error detection and retransmission and most are used on point-to-point circuits where a failure to communicate can be the fault of only 2 parties, the host

Communication Test Equipment—A Solution To Increased Productivity

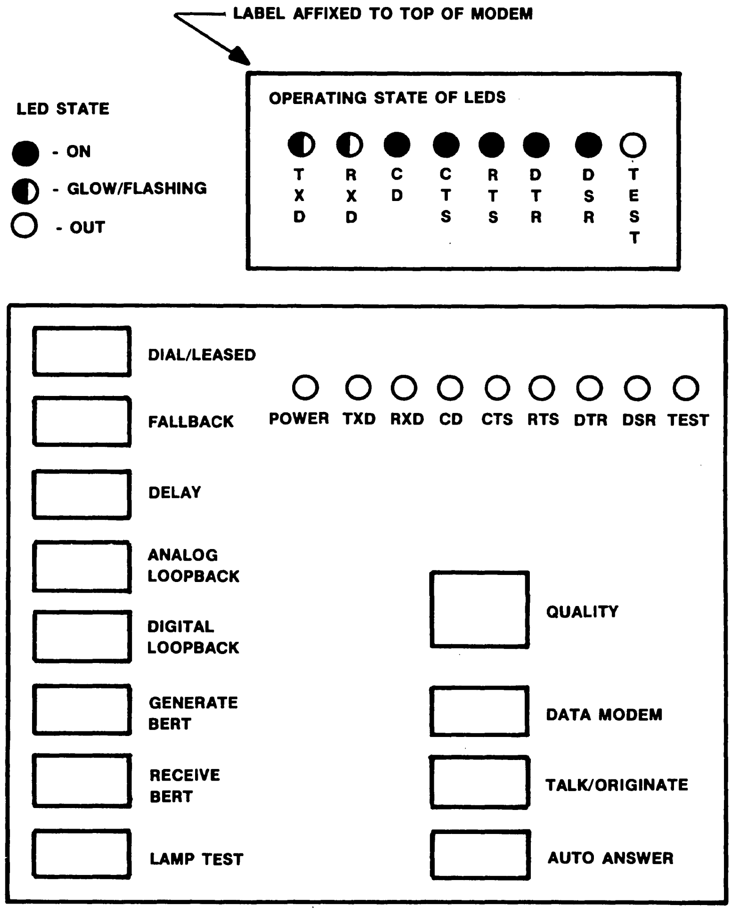


Figure 6 • labeling a modem for indicator states.

or the terminal. More expensive synchronous terminals may provide more sophisticated test aids. Some terminal features to look for:

• **Power on light.** As strange as it seems, the power on indicator may be the best terminal test aid. The majority of communication problems with remote terminals can be traced to operational problems, and not turning the terminal on or accidently turning it off or unplugging it are high on the list of probabilities. Power problems are also detected this way. Don't assume that because the lights are on the terminal power is available, most businesses have separate circuits.

• **Activity indicators.** Synchronous terminals such as the IBM 3270 normally have an LED to indicate that they are online and being polled. The state of these indicators can give you the terminal's side of the picture. If a terminal is not being polled after a failure it may have been dropped from the CPU poll list, an operational issue rather than a continued failure.

• **Local or half-duplex operation.** Most terminals can be made to operate off-line by setting a switch to **local** or **half-duplex** (HDX). On asynchronous terminals this provides a path for the local echo of keyed characters so that they are displayed on the screen. A quick check in local or half-duplex mode can verify that the terminal electronics are functioning. Synchronous terminals and more complex batch terminals also normally have a means of local testing. Consult your user's manuals on the details.

Network control and management systems may be the best way to conduct communication testing in large networks. These systems perform network surveillance, alert an operator to a detected failure, and provide an operator console or panel from which each modem can be tested. Most produce alarm reports on abnormal conditions. The individual tests which these systems support are covered in this document at various points, the modem tests being described above. Details on the operation and use of these very sophisticated and expensive systems should be taken from their technical documentation. With a few exceptions which will be noted, they can be regarded as a means of making the step-by-step diagnosis of problems in large networks practical by supporting it from a central point rather than a vehicle to support new sets of tests.

■ TESTING THE INTERFACE—THE BREAKOUT BOX

The digital interface is the gateway to the communication path, attaching the business machine to the rest of the circuit. Problems here will almost certainly prevent communication altogether; the equipment to test this interface is among the least expensive of communication test equipment, yet few users have such a tester.

The device most often used to test the interface is normally called an interface monitor or breakout box. It is normally quite small and consists of a pair of digital connectors (the DB-25 for the popular EIA RS-232C standard interface, for example) and a set of DIP switches, plug pins, and LEDs. **Figure 8** shows a typical front panel for such a unit. Using the box is very simple. The cable attached to a modem or terminal is removed and inserted into the correct connector on the breakout box. The connection marked **terminal** must be connected to the terminal or computer, and the one marked **modem** must be connected to the modem. **Be sure to connect this correctly or the signal status will be misleading.** If

Communication Test Equipment—A Solution To Increased Productivity

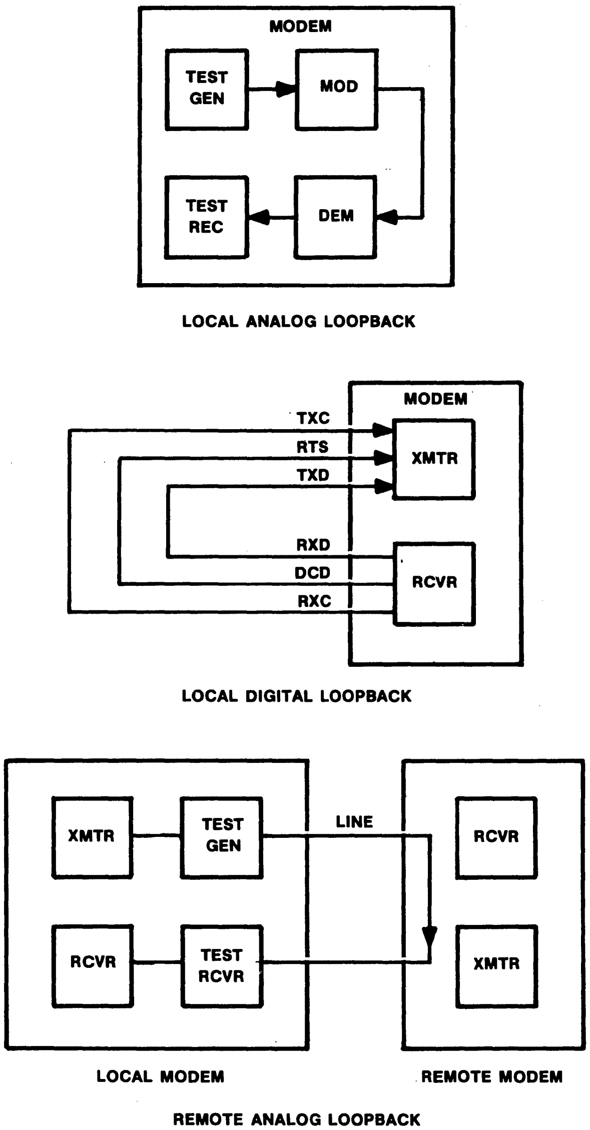


Figure 7 • modem loopback configurations.

your cables follow the standards it will be impossible to connect this wrong—you'll have 2 male ends or 2 female ends together.

Connected in the circuit, the breakout box displays the status of the control and data leads on the interface via the LED indicators. Since RS-232C (or CCITT V.24/V.28) is the most common standard, the examples given here will apply to that interface. RS-232C will normally use the following interface leads for the reasons given:

- **Pin 2—Transmit Data (TD)** is information flowing from the computer or terminal to the modem.
- **Pin 3—Receive Data (RD)**.is information flowing from the modem to the terminal or computer.
- **Pin 4—Request-to-Send (RTS)** is a command from the terminal or computer that the modem prepare to transmit data.
- **Pin 5—Clear-to-Send (CTS)** is an indication that the modem is ready to accept terminal or computer data to transmit.
- **Pin 6—Data Set Ready (DSR) or Modem Ready (MR)** is a signal to the computer that the modem is operating and connected to the phone line.
- **Pin 8—Carrier Detect (CD)** is also called Receive Line Signal Detect (RLSD), is an indication from the modem that there is a transmit carrier from the opposite modem in the circuit, i.e. the other station has requested transmission.
- **Pin 20—Data Terminal Ready (DTR)** is a signal to the modem that the terminal or computer is active and connected to the interface. In an automatic answer environment, DTR enables the modem to answer an incoming call.
- **Pin 22—Ring Indicator (RI)** is a signal from the modem that an incoming call has been detected. If DTR is present from the computer or terminal, the call will be answered.

Other interface leads have meaning in some applications, but many interface monitors will not display their status. If you have an unusual application (secondary channel usage, for example), you must be sure that the monitor you select will display all the control leads you might use.

Using the monitor means interpreting the proper state of the control signals, which requires that you understand a little about the interface signals used by your equipment. RS-232C is a standard, but its observation by some vendors is loose to say the least. If you can, be sure to consult your equipment reference manuals. If you cannot, here are some general rules:

- **The idle state of a dial-up line** when no call is in progress is usually DTR on (lit) and all other indicators off.
- **The idle state of a leased line** (no data being exchanged) will usually be DTR and DSR on. If the line is full-duplex, RTS, CTS, and CD will also be on.
- **During actual data transmission** from the computer or terminal through the monitor to the modem, DTR, DSR, RTS, and CTS will usually be on. TD will glow or flash as the data is sent.
- **During reception of data** (from the other station) DRT, DSR, and CD will normally be on, and RD will glow or flash as the data is received.
- **RI will light for a brief flash** when an incoming call is received at a dial-up modem.

If you are going to use a breakout box to its full advantage, you'll have to understand the way in which your own equipment uses the interface. One way is to observe normal operation and to record the proper state of the LEDs in a manual or on a label attached near the test point. The above rules, however, will generally apply. For example, if you connect a breakout box between a terminal and modem and the DTR light does not light you can generally deduce a terminal failure or cable problem. Modems just do not operate with non-ready business machines. A failure of the modem to present DSR means a modem or modem/line connection problem. If a computer or terminal is not raising RTS and all other indicators are proper, it is not attempting to send data and could be malfunctioning. If RTS is present without CTS, it indicates a modem or cable problem. Lack of carrier (CD off) when you know that the other station is attempting to send may indicate a failure of either modem or of the communication line.

Breakout boxes can also be used to interrupt the signal path for test purposes or to create a new path by patching or bridging 2 signals together. A common use of these techniques is to determine the combination of bridge connections needed to create a **null modem** cable; a cable whose interface signal leads have been arranged to let a terminal and computer directly connect without an intervening modem. If you do decide to use patching and switching, be sure you understand where the switches and patches and LEDs connect into the circuit. **Figure 9** shows a typical breakout box design. Note that the LED displays are taken from the terminal side of the interface, so they will always show the state of the interface as seen by the terminal. Also remember who generates each signal. Turning off DTR at the switch may not turn off the indicator, since the indicator monitors the terminal side of the connection.

Communication Test Equipment—A Solution To Increased Productivity

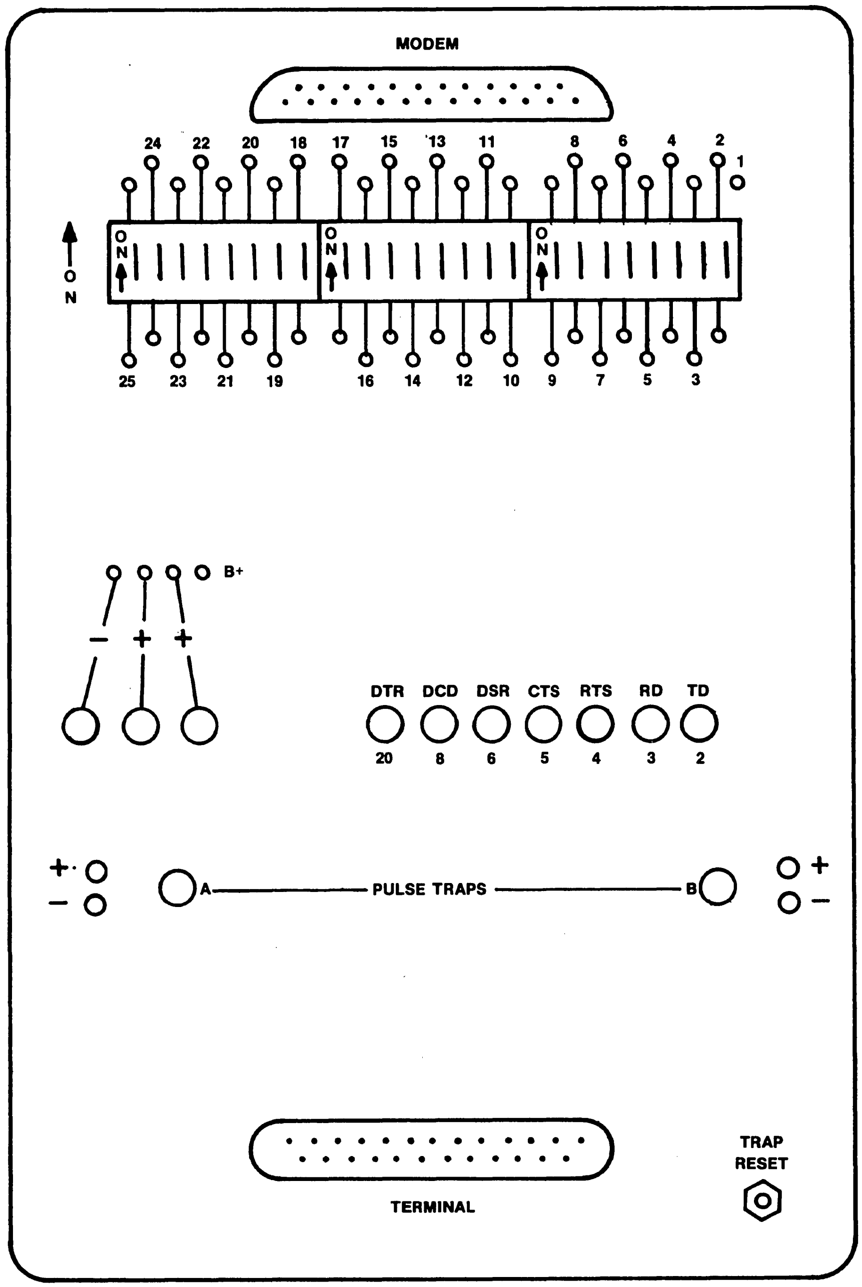


Figure 8 • the breakout box.

Most breakout boxes contain several auxiliary LEDs which can be patched into the circuit to detect signals on leads for which permanent monitoring LEDs are not assigned. Some breakout boxes, such as the type shown in Figure 8, have the provision to detect pulses on interface circuits short enough to be invisible with standard monitoring. This is done by latching the LED on if a pulse is detected. There is a reset button to release such latched displays.

■ DIGITAL TESTS OF THE LINE

If the interface checks out and communication is still not proper, or if the interface test suggests a line or modem problem, the user has several choices in how to proceed with testing. The line can be tested in its voice or analog form, or the line can be tested using digital signals. Analog testing of a line is normally more complex, and in any case it should not be attempted until the proper operation of both modems has been established. Digital testing allows a quick check of the data path as it appears to the computer or terminal. Don't waste your time with digital testing if there is no data path between the users of the circuit. Digital testing detects soft failures which cause data errors, not total failures. There are 2 basic forms of digital testing, the **bit-error rate test** and the simulation or **FOX test**.

Communication Test Equipment—A Solution To Increased Productivity

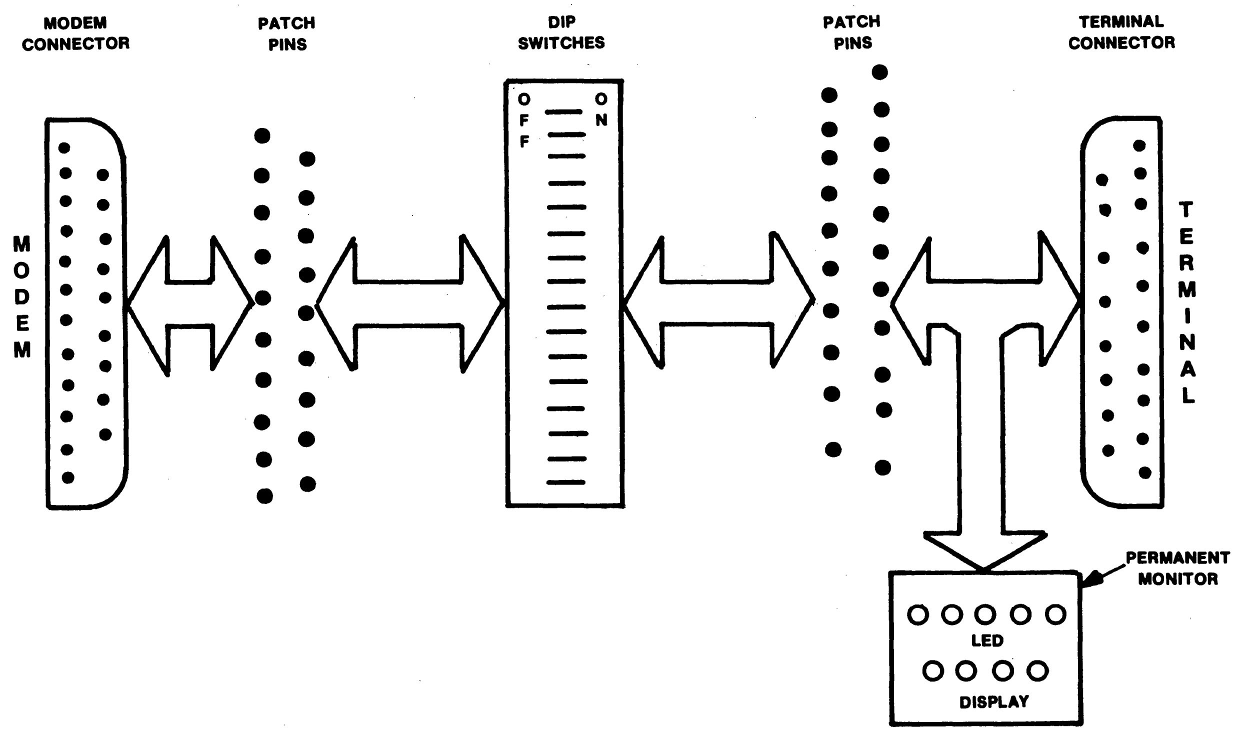


Figure 9 • inside a typical breakout box.

Bit-error rate testing is performed by looping the modem transmit path back to the receive side and injecting a test signal on the transmit side whose bit pattern is known. The most common such signal is the CCITT 511-bit test pattern. The pattern is checked for errors, and an excessive number of errors indicates a soft failure in the path (a hard failure would prevent the test from being run at all). Modem loopback can be performed remotely or locally, and in some cases a second unit can be used to receive and check the pattern, eliminating the loopback. **Figure 10** shows the application of a BERT tester in checking a pair of modems and the connecting line. Note that the modems can be either synchronous or asynchronous. Most BERT testers will operate through the entire range of data communication line speeds.

In operation as shown in **Figure 10**, BERT testing provides a measure of the quality of the transmission paths, including the modem. By operating the local modem in analog loopback first, then performing a line loopback at the remote modem, and finally a digital remote loopback each element of the path can be checked. Many users begin with the remote digital loopback test and shorten the test path only if an objectionable level of errors are detected. What level is objectionable? Most users consider that a single error on a 511-bit pattern is unacceptable, but you may want to check the error rate on a line which is operating acceptably as a benchmark.

BERT testers, for asynchronous use, tend to have standard features. If you plan to use such a tester with synchronous links, be sure that the tester will support external clocking if needed or will supply clock as required. You may also want to look for the

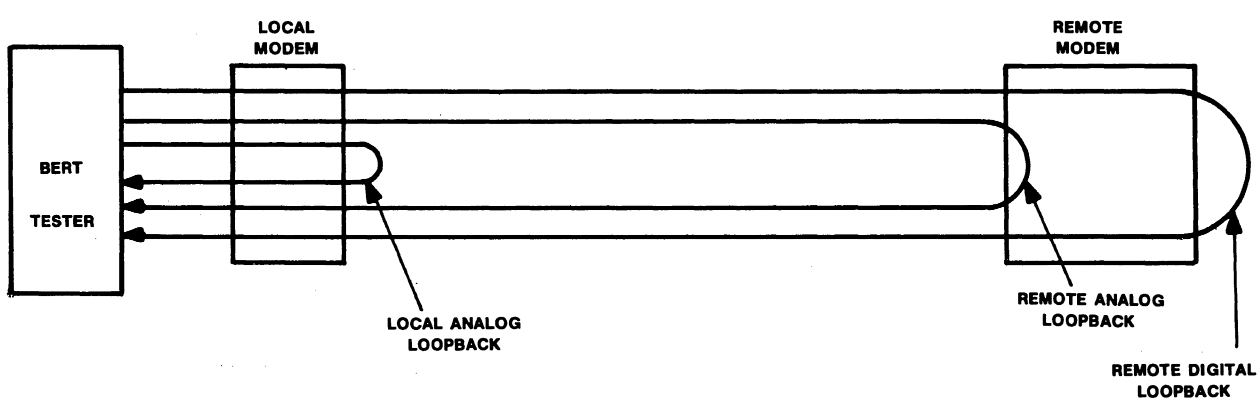


Figure 10 • bit error rate testing.

Communication Test Equipment—A Solution To Increased Productivity

following features:

- Force error switch to generate and inject an error on command, or at a random interval to a basic bit error rate (1 in 1,000, for example).
- Built in EIA interface monitor.
- Longer bit error patterns (2047 bits is common).

Simulation Testing

Sometimes bit error rate testing will prove that the communication path is operational but not uncover the cause of the failure. BERT testers cannot generally operate to the end-point business machine. For quick tests of the total path to a remote terminal it may be convenient to use a piece of test equipment to emulate the computer. By the same token, a tester which can emulate a terminal device can be very useful in checking out new communication software. Such a device may be called **a terminal tester, a fox box, or a simulation tester. Figure 11** shows such a device connected to test a terminal and a computer In the terminal tester mode the simulation device can be a very simple unit which generates an asynchronous string of characters such as the **quick brown fox message** at the speed specified. The characters will appear on the terminal, verifying the communication path. Similar logic can be used to simulate a terminal to the host computer, providing that the computer will accept this type of message.

Some asynchronous applications and nearly all synchronous applications need a more complex form of testing since the communication protocol between host and terminal is more difficult. Synchronous testers are normally customized for the type of device they can check, the IBM 3270 terminal tester is a popular example. These testers can be configured to recognize an address and to respond to a poll with data. Many will also acknowledge a select message from the host.

Because simulation devices simulate the operation of the terminal or computer, they replace one or the other in the connection and normally operate straight through to the other end. They are primarily useful for conducting a quick test of total path integrity rather than for the systematic checking of the circuit elements. In more complex applications, they are used with a type of test equipment covered in the following section, the **data line monitor**.

Many users solve the problem of whether to buy a BERT tester or a **fox box** by buying either, both, or neither. Some combination units are available for those who actually need both capabilities, but most users will find that their requirements dictate one or the other. Users with low transmission speeds and no desire to check on the phone company should consider the **fox box** or another type of simulator. Users with high data speeds and a history of problems with the quality of the phone circuits may want to get a tester which can provide a direct measure of the error rate of their digital path. Before you buy a BERT tester, however, make sure that one is not built into your modems.

■ PROTOCOL TESTING WITH DATA LINE MONITORS

A data line monitor presents a window on the transmission path to allow the user to observe and analyze the actual transmission for protocol or software errors without interfering with the transmission. It is perhaps one of the most complex but beneficial pieces of data communication test equipment, and also one of the most difficult to use properly. Although the units are much more expensive than the other types of testers described so far, they are also the most versatile and most useful. A large installation with multiple terminal types and multiple computers will probably find the data line monitor indispensible, and any user who must program computers for customer interfaces should purchase such a unit as soon as the requirement is known. The line monitor is not normally used to diagnose total failures of a line, but to determine what data is flowing on the line as an aid to the detection and correction of more subtle problems such as high error rates, improper protocol, improper terminal addressing, etc. To fully utilize a data line monitor the user must know how normal operation of the line would appear and how the line protocol would respond to unusual conditions. The manufacturers' information in the communication device manuals will normally provide a description of both correct operation and the results, line errors, or other failures.

The data line monitor consists of a CRT display connected to a line interface; it monitors all EIA control signals and both the transmit and receive data paths. **Figure 12** shows both a block diagram of the device and its typical method of connection to a line. As the figure shows, the monitor is logically attached via a "T" in the line, although the physical connection normally consists of removing the terminal or computer cable from the modem and plugging into the monitor, then running another cable from monitor to modem. Like the interface monitor, the proper operation of the line monitor is dependent on correct cabling.

Once connected into an operating circuit, the monitor displays the data exchanged as a series of lines similar to that of a CRT

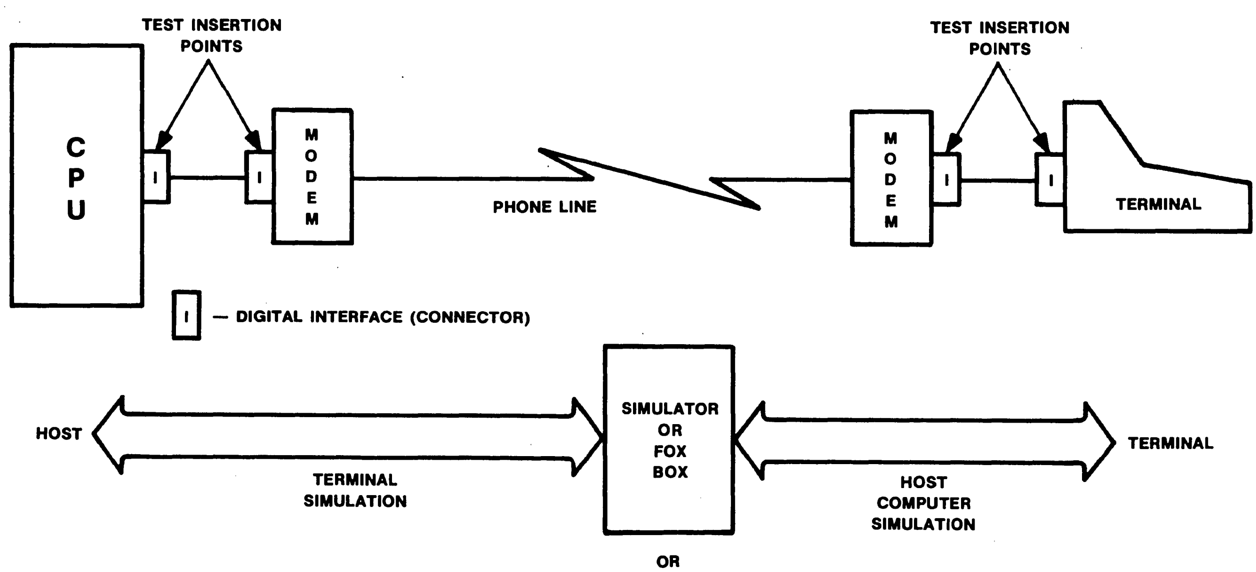


Figure 11 • use of simulator/fox testers.

Communication Test Equipment—A Solution To Increased Productivity

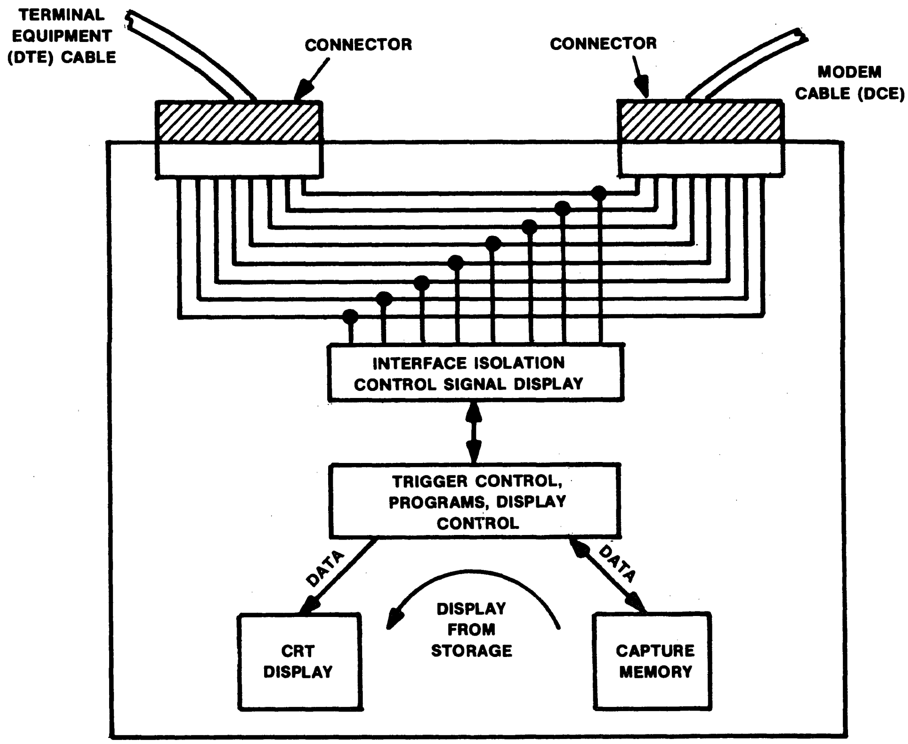


Figure 12 • internal configuration of a data line monitor.

terminal. Idle line time is normally indicated by a special graphic symbol or a series of dots ( ). The data may be displayed in a hexadecimal code (the letter "A" is equivalent to hex 41 in the ASCII code set, for example) or decoded into characters. But because line monitors read **data**, they must be present to the proper communication parameters, including line speed, protocol, character type, code set, and in general, all of the parameters that a terminal would need to establish for proper operation on a line. These line monitor settings are usually established through displayed menus or sometimes via what may seem a bewildering array of front-panel controls.

There are many types of data line monitors with many different schemes for setting the relevant line parameters, so a description of any specific method will be of limited value. There are some general groupings of functions, features, and controls which can be defined, and the operation of the data line monitor in various applications can be demonstrated using these general groupings.

☐ Using a Data Line Monitor for Asynchronous Monitoring

The easiest function to set up on a data line monitor is the review of asynchronous data. This application serves to demonstrate the basic **input and display control** features of the unit.

First, all line monitors provide selection for the communication parameters of the data to be monitored, either **SEND** (from the terminal or computer direction) **RECEIVE** (from the modem direction), or both. When both are displayed, you may have a set of options like HDX-2, HDX-4, or FDX. In half-duplex, 2-wire mode, the control signal Request-to-Send controls the display of transmitted data and received data. In HDX-4 mode, it is assumed that Request-to-Send is normally present and both data paths are monitored simultaneously. This mode will cause overlap and garbling of the display if the stations transmit simultaneously. In that case, the line is running full-duplex and the FDX display mode must be used. Most data monitors underline or highlight the received data (from the modem) so that it can be distinguished from transmitted data. Most asynchronous devices run full-duplex when directly connected or over full-duplex modems. Some exceptions are block-mode terminals used on such systems as the HP-3000.

The **character type** must be selected to identify the data as asynchronous, character synchronous, or bit-synchronous (SDLC and HDLC are bit-synchronous protocols). If the protocol is asynchronous or synchronous, it is also necessary to select the number of bits (including parity) per character. Some monitors also require that parity type or mark/space (odd, even, none, zero, or one) be specified. These units may flag a bad parity character on the display. If a bit protocol is used, it is necessary to define the line coding as normal (DIR) or NRZI. For asynchronous operation, you can normally assume an 8-bit character. One start bit will be assumed, and 1 or 2 stop bits are usually acceptable. Some units may require that you select the number of stop bits.

The **code set** must be defined if you want the displayed characters to be understood. Most line monitors will display the data in hexadecimal form as standard and decode it to alphanumerics optionally. Look for the ability to display in hex, ASCII, or EBCDIC. The latter is the standard with the large IBM mainframes while ASCII is most common in minicomputers or microcomputers. Hex display is much more difficult to read (the word **NAME** would appear in hexadecimal as "4E 41 4D 45" in ASCII), but is applicable to any code set and will help reading protocols such as SDLC which use individual bits as meaningful data.

The proper parameter settings can usually be taken from the terminal or modem settings. Once the data line monitor is properly set up (correct operating parameters are selected), information is displayed on the screen as it is monitored. If the source device is a terminal, the characters will be widely spaced to pace the speed of the operator's keying. The responses from the customer will normally be tightly spaced indicating transmission of a block of information. In some applications, the computer will echo each character keyed to cause it to display on the terminal,

Communication Test Equipment—A Solution To Increased Productivity

resulting in the transmission of each keyed character back in the other direction.

□ Viewing Synchronous Data

The setup for synchronous monitoring of a line is initially the same as that for asynchronous, except for the character type selection. In addition to this, however, it is necessary to tell the monitor what type of sync pattern will be used at the start of each message. This is normally done by entering the hex value of the code, something which many users will not know. IBM bisync, probably the most common of the synchronous protocols, will use a sync pattern of hex 1616 (2 bytes of value hex 16 each) if the code set is ASCII, and 3232 if it is EBCDIC. It is also necessary to tell the system when to break synchronization. This is normally an all-bits-on character, hex "FF."

■ SPECIAL FEATURES OF DATA LINE MONITORS

The majority of users will find that the basic monitoring functions described above will satisfy their needs, but there are times when special features and capabilities can enhance the usefulness of a line monitor and reduce the time needed to diagnose and correct a problem. Some of these features are:

• **Integral EIA breakout panel.** Nearly all data line monitors provide a full display of the EIA control signals. Many will also provide the switching and patching functions of the **breakout box** discussed earlier. A line monitor is too large and too expensive to be considered as a substitute for a breakout box, but the availability of the features will make some diagnosis easier.

• **Automatic setup.** Some units will sense and interpret the line protocol and parameters and set the speed, character configuration, and other parameters automatically. This can be a big help in installations which have a mix of lines with various speeds and protocols. Be sure that the automatic feature can be overridden and that its range of selection is adequate for your use or it may be a **hindrance** instead of a help.

• **Marking data.** Most monitors allow you to highlight characters with bad parity, block check sequences which do not check properly, or characters which arrive while the Carrier Detect control signal is present. This is selected by a **marker** control and results in the selected element displayed in a high-visibility mode, usually reverse video (white backgrounds with black character image).

• **Triggers.** A trigger is the result of a test for a condition which requires a specific action when it occurs. The trigger may trap a data stream, highlight a character, turn the display on or off, increment a counter, start a timer, etc. Triggers are programmable functions limited to more sophisticated units, and it is necessary to understand the details of the communication protocol in order to use them. For example, a trigger can be used to suppress the display of information on the screen except when a certain drop on a multidrop line is addressed. In this case, a trigger would be set to enable the display when the drop was addressed in a poll or select message, and another to disable the display upon receipt of an EOT. The best data monitors will support trigger matches of multiple characters, and will allow the user to insert one or more **don't care** or **mask characters** into the string if non-contiguous characters are to be checked. Line monitors which include the ability to **generate** data interactively (sometimes called **emulation mode**), the trigger forces the data monitor to recognize a unique event in a protocol and to respond to it properly by triggering a specific emulation function.

• **Timers.** Most communication protocols require responses to requests or commands within a **specific time interval**, and many evaluations of terminal or computer performance require the timing of communication events. Triggers can be used to **start or stop timers** so that the **interval** between events such as command and response **can be measured**. The trigger starts a timer when the terminal sends data to the host (by matching the header identifying the terminal) and stops it when a message is sent to that terminal (by matching the select sequence).

• **Counters.** Like timers, counters are used in conjunction with triggers to tally the occurrences of a specific event. A counter and timer may be combined to provide a measure of interval and rate. For example, the total time elapsed during message transmission can be determined by a timer which starts at the beginning of a message and stops at the end of a message character. A counter can count data characters within the interval, allowing the message data rate to be calculated. A counter can also be used to calculate the polling rate of a system by measuring the number of polls during an interval of time.

• **Emulation or Interactive Mode Operation.** An increasing number of data line monitors provide an interactive mode to transmit messages rather than remain totally passive in a connection. An interactive line monitor can perform the functions of a **fox box** or other type of terminal emulator, but in a more complex and sophisticated manner. Transmission of data is normally associated with the use of a trigger. When the test condition is encountered, it triggers the monitor to transmit a predefined message, entered by the user or programmed by the vendor. This feature can emulate a terminal by causing the monitor to respond to a poll, but its applications extend to the simulation of unusual conditions such as bad block handling, timeouts, invalid frames, protocol errors, and others. If a variety of incoming messages to which unique replies were required were to be handled, multiple messages and triggers would satisfy the application.

• **Data storage and replay.** Anyone who has ever used a data line monitor has seen conditions where extensive setup time to create a specific problem for review was wasted when the event occurred at a time when no one was watching the screen. All data line monitors have the ability to trap and store information in capture memory or on a diskette or tape cassette. The captured information can be displayed for careful analysis; small events which might otherwise be difficult to detect can easily be observed. Information can be captured by a trigger on some models, allowing the user to selectively monitor the line and record only information pertinent to a particular problem.

• **Stored programs.** The entry of trigger, timer, counter, and transmit specifications normally requires a monitor with a menu-driven parameter structure, but even this interface will not address the problem faced by users who must re-enter complex applications many times. For these users, most programmable units have the ability to store information on a diskette or cassette tape. The user can then load a previously defined program from a menu and run it at will, making it possible to conduct complex monitoring or emulation sessions with personnel technically unqualified to set the unit up properly. **Figure 13** shows a tape program load menu.

• **Remote operation.** Some data line monitors are designed to collect information remotely rather than to display to an operator. These units consist only of the monitor control and storage modules, and may be combined with a local data line monitor

```
** PROGRAM TAPE DIRECTORY **

PROGRAM #: 10

TEST ID: 3705 EMULATION/3276-SNA

TAPE NAME: EXAMPLE 100 USER PROGRAMS

STATUS: PROGRAM LOADED

001: ECHO TEST FOR HP

002: TERMINAL-KEYBOARD EMULATION

003: BISYNC TEST & ERROR PROMPTS

004: SDLC/NRZI MONITOR

005: NETWORK PERFORMANCE TEST
```

Figure 13 • displayed tape loading of programs.

Communication Test Equipment—A Solution To Increased Productivity

having conventional display capabilities. **Figure 14** shows a remote data collection device used to monitor a 56K-bps path, then dump the collected data over a dial circuit at a lower speed so that it can be monitored. This can be very useful in applications where a network switching hub is not manned or has no on-site technical specialists.

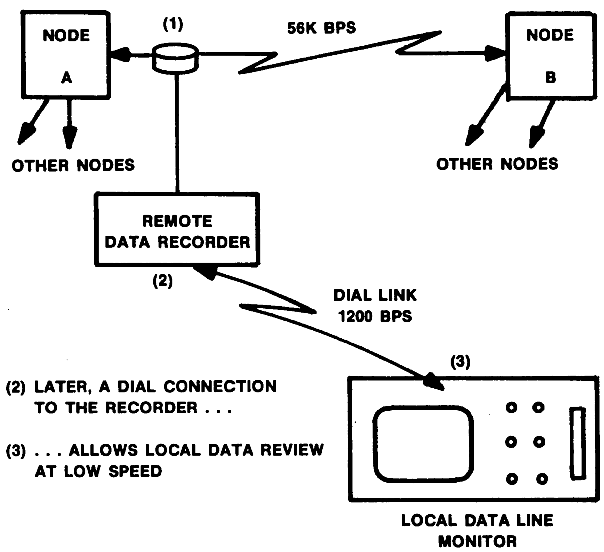


Figure 14 • example of remote data recorder application.

• **Custom protocol displays.** When a monitor is used with complex **layered protocols** such as IBM SNA or CCITT X.25/X.75, conventional displays in either hexadecimal mode or character mode are of little value because the fields in the control headers of these protocols are often only a few bits long. For such protocols a custom display can make the decoding of line activity considerably easier and reduce the probability of a misjudgement which can send an entire diagnostic effort down a wrong track. **Figure 15** shows the custom display associated with monitoring an X.25 link. Note that the packet and frame information are separately available.

☐ Selecting a Data Line Monitor

The proper data line monitor for your installation is a function of the complexity of the application and the degree of sophistication of the technical personnel available. Most asynchronous data centers need not spend the over $10,000 price tag of the more complex units, because the cost of these units is often escalated by features such as the ability to monitor X.25 or SDLC. Here are some guidelines on desirable features divided by the type of monitor application:

Basic Monitoring:

• **Basic monitors must be easy to use.** Screen menus are preferable to switches for establishing communication parameters and control functions. Are the displays easily understood? Is the documentation clear and does it cover all of the functions of the unit?

• **Look for protocol flexibility.** Can you monitor synchronous and asynchronous lines? How about X.25 and SNA? Are all the character size and parity combinations available? Can you select automatic recognition of speed, character type, and protocol? Are high monitor speeds supported? Is there an integral EIA Interface monitor (breakout panel) with LED display? Is there an integral or external recording medium?

• **Portability is often an asset with monitors.** Can the unit be hand carried and shipped easily for field use? Is it relatively insensitive to variations in line voltage and temperature?

DTE						DCE					
ADDR	TYPE	R	S	P/F		ADDR	TYPE	R	S	P/F	
01	*INFO	2	4	0	G	01	RR	5		0	G
03	RR	3		0	G	03	*INFO	5	2	0	G
01	*INFO	3	5	0	G	01	RR	6		0	G
03	RR	4		0	G	03	*INFO	6	3	0	G
01	*INFO	4	6	0	G	01	RR	7		0	G
						03	*INFO	7	4	0	B

DTE						DCE					
LCN	TYPE	R	S	QDM		LCN	TYPE	R	S	QDM	
000	*RSTRT		00 00		G	000	RSTRT	C			G
001	*CALL		E E		G	001	CALL	ACC			G
001	*DATA	0	0		G						
001	*DATA	0	1	Q	G	001	RR	2			G
001	*DATA	0	2		G						

Figure 15 • custom displays of layered protocols.

• **Small displays can be hard to read** and may not show an entire sequence. How large is the display, both in terms of area and in terms of characters displayed? Is it adequate for your needs?

Protocol Analysis:

• **Triggers are the key to analysis.** How many triggers are offered? What can the triggers select on; strings of data, interface control signals, timeouts of timers? Can either send or receive direction, or both directions, be tested and displayed independently? What can the triggers activate? Timers? Counters? Trapping or recording enable and disable for visual analysis on the screen? Is there a convenient way to enter the trigger information, display its state, and record it?

• **Custom displays may be needed for complex protocols.** Can the unit display the layered protocols in a special format? Are there different displays for each layer in the protocol?

• **Data recording is normally required in protocol analysis.** What method of recording data is available? How long can

Communication Test Equipment—A Solution To Increased Productivity

recording take place at the maximum link speed? Can triggers stop and start recording? Can **unidentified** data be recorded or must the character structure and protocol be identified to the monitor first? Will the triggers and other programming functions be available during a replay?

Terminal Emulation & Message Transmission:

• **What initiates a message transmission** is a major factor in the usefulness of the feature. Will a trigger initiate transmission of a message? Can the operator force message transmission manually?

• **Message entry and storage features will vary.** Is there a full keyboard? Can it be used interactively or must messages be stored for later transmission? What is the maximum length of a single message? The total length of all messages? How many different messages are supported? Will the system calculate the BCC for synchronous messages? Will it provide data framing?

• **Program storage is a requirement for these units.** Is the unit user programmable or does it provide vendor programs? How are programs stored? How are they rerun? Can multiple programs be placed on the same diskette or cassette? Can data share space with the programs?

• **Sample programs are needed for these units**, since learning the complex features is difficult even with manuals. Does the unit come with program samples? What kinds? Are the sample programs discussed in a training document?

Be sure to consider the frequency of use and the future plans of the organization in the selection. A monitor which will be used primarily in the heat of a full-scale communication crisis must be very user friendly (easy to operate), and one being purchased in an organization planning dramatic communication growth must be highly flexible.

■ ANALOG LINE TESTING

The average small user should consider the analog telephone circuit as something which works or does not, and testing to establish a failure in this case need not go further than the determination that the link itself is the problem. Larger users with many lines may find that a more detailed analysis of the conditions on a line suspected of failure can make the phone company more responsive. Whatever your business size, however, you should realize that since you cannot repair telephone circuits yourself any testing of the circuit is a gathering of ammunition in a power play to assure the fastest possible response from someone who can. Before you play hard ball with an organization as experienced in telecommunication as the phone company, you had better be prepared to learn the details of the analog path, how to test it, and what constitutes a failure.

A basic telephone channel is assumed to have a useable frequency response (bandwidth) of 300 hertz (cycles per second, abbreviated "Hz") to 3000 Hz. Signal loss, crosstalk, impulse noise, signal delay, etc are all controlled on the path to a specification provided by AT&T or other carriers. When you lease a phone circuit you are entitled to a line which meets the basic specification of the circuit, classified by AT&T as a **3002 channel**. You may also elect to pay a premium charge for a **conditioned channel** which exceeds the basic 3002 specification by reducing one or more of the controlled interferences to clean communication. Conditioning is classified as **Type C or Type D**. Type C conditioning controls the attenuation distortion and the envelope delay distortion of a line, and Type D the nonlinear distortion and signal-to-noise ratio. Both types of conditioning are available in grades, and for each type and grade there is a new specification which you are entitled to expect to be met by your path. **Figure 16** shows an example of the effects of conditioning on a line characteristic.

When a line fails completely, there is little equivocation possible on the part of its supplier, and further testing is both unnecessary and probably impossible. But when a line degrades in one or more of the basic operating parameters such as frequency response the result may be a path which will carry voice and low-speed data but cause excessive errors at higher speeds. The line's supplier may feel that the problem is not in the channel but in some piece of user equipment which by coincidence it did not supply. How do you break the deadlock? First, it may be possible to use some of the modem indicators to help. If your modem has a line quality indicator, it should indicate a poor line in the presence of any operating degradation which actually affects its operation. If you can show the carrier service personnel that the modems operate properly in self-test or local loopback but report poor circuit quality, it may accept the fact that the line is bad, particularly if other lines are operating and reporting good quality. If not, you may need to perform some level of analog testing to track the parameters of the line and indicate where they deviate from specification.

One form of analog line tester contains a cathode-ray tube (CRT) or numeric display and a set of operating controls, and can be used to monitor a test signal and report its operating parameters. The parameters measured are distortion, c-notched noise, phase jitter, dropouts, gain hits, and other line disturbances. A typical use of such a device is shown in **Figure 17**. Note that there is signal generation and measurement involved, a condition which is typical of analog testing.

If you feel it necessary to provide your own telephone circuit testing capability, don't dissipate its value by establishing an adversary relationship with the phone company. Address inquiries to the proper levels with courtesy and restraint, something which may be difficult with the high degree of internal pressure which usually accompanies a communication failure. And be sure that you do your homework. Nothing will reduce your credibility quite as fast as having thousands of dollars worth of test equipment which you misread or apply inappropriately. Here are some guidelines for employing analog measurement and test equipment:

1. **Don't use analog test gear unless you have a connection.** A hard failure will prevent testing. Loss of connection on a leased line may be hard to distinguish from a very bad line, but in general the modems will show some flicker of the Carrier control signal or the line quality indicator will show a bad circuit.

2. **Don't bother with analog tests on dial-up circuits.** You will never be able to prove or do anything constructive with the result because for each dial connection, a different path is connected.

3. **Develop a systematic test plan for analog testing of a line** and apply that plan each time you test. It's best to have a worksheet on which to record the readings you get. Having something down on paper looks best, and it makes it possible for you to spot degradation in certain values.

4. **Double check before you accuse.** Have another operator run the tests, or run them a second time with the same personnel. Check the test setup carefully. Also, be sure that the values are reasonable. If you get a frequency response curve that indicates that the line fails to meet specification within the normal bandwidth by a large margin, check again.

5. **Follow procedures.** If you are going to use analog testing, find out what to do with the results. Normally, you will want to notify the business office of a problem, of the fact that you have independently tested the circuit and found it to be out of specification in certain areas, and that you will file for a rebate after 2 hours of downtime.

If this sounds complex and confusing, it generally is. Most users will find that using analog line testers is a form of trying to beat a team with the home court advantage. Unless you have a lot of communication experience within your organization you would probably be better off just to report problems and spend your test dollars elsewhere.

■ MODEM TESTING & MODEM NETWORK CONTROL—THE EYE PATTERN

If you are a typical user of communication channels you may feel discouraged by the complexity of analog line measurement and testing, but unwilling to accept the phone company's good will as the only alternative. In fact, it is not. Many modem manufacturers offer network control and management capabilities which perform network surveillance for abnormal conditions and the line through the analysis of modem operation. Some of these systems are very sophisticated and will provide an actual display of the line responses. **Figure 18** shows such a display. Other systems may display the characteristics of the line in a different way—**the eye pattern**.

Communication Test Equipment—A Solution To Increased Productivity

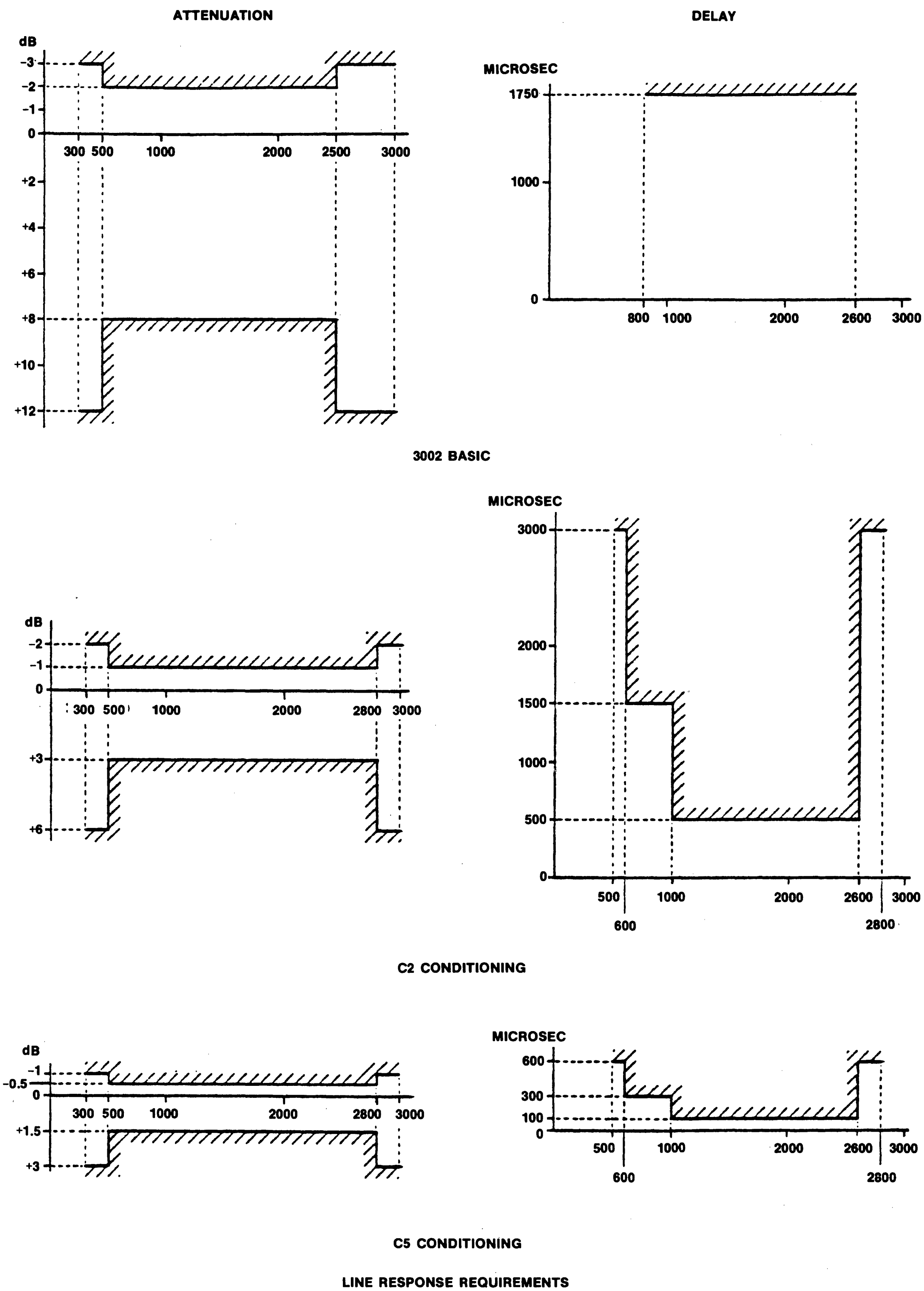


Figure 16 • effects of line conditioning.

Communication Test Equipment—A Solution To Increased Productivity

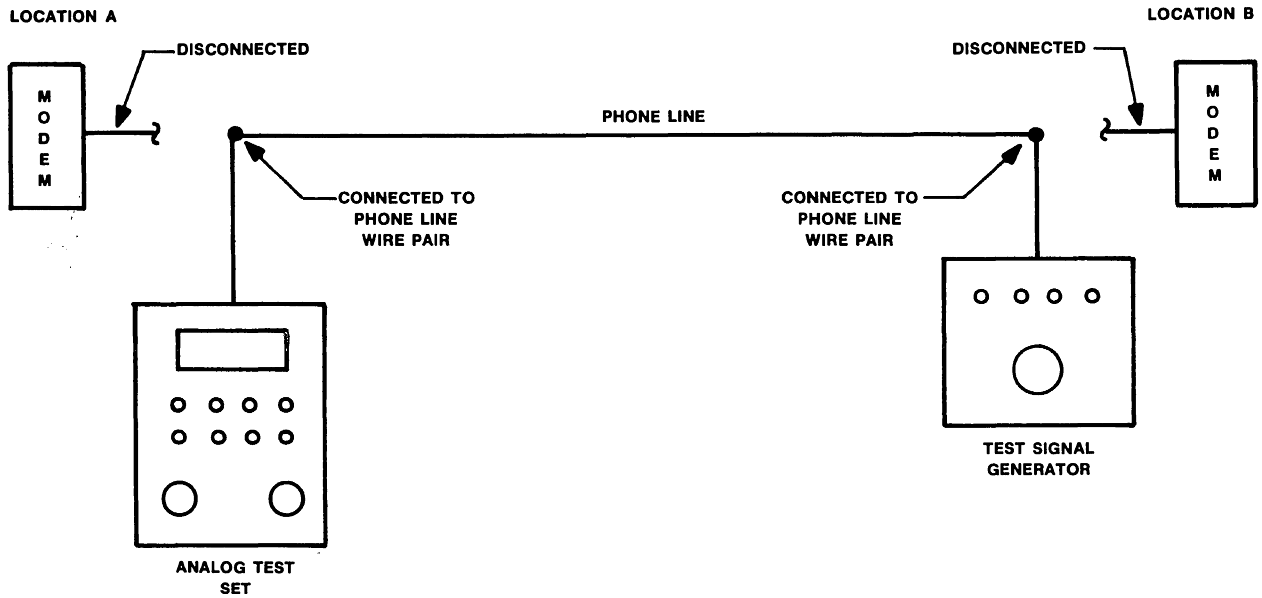


Figure 17 • using analog test equipment to determine phone line parameters.

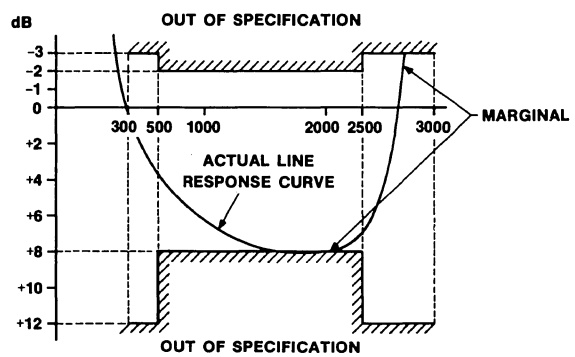


Figure 18 • line parameter display—3002 attenuation vs specification.

Modern modems usually employ a form of modulation which represents groups of bits by a single signal value. For example, a 4800-bps modem normally groups bits into threes, so there are 8 possible signal values, corresponding to the bit combinations 000, 001 ..110, 111. A modem will operate correctly if it can unambiguously interpret the signal on the line at any moment as one of these possible values. But distortion or other line problems interfere with the signal and may cause it to **slide** into another range, producing a bit pattern error. A measure of the quality of the line can thus be made by comparing how well the signal values actually received on a line match the **ideal levels**. One popular way of displaying this information is the **eye pattern**, a radial display of line signal energy. An eye pattern is formed by treating the display as a **pie** whose slices are of the same size and number 1 for each possible signal value. Our 4800-bps modem example would have a **pie of 8 slices**. The pattern is plotted on the screen as shown in **Figure 19**, and each time the signal is sampled on the line a dot is plotted at its position in the pattern according to its value. The distance from the center point is proportional to the amplitude of the signal and the direction from the vertical represents its phase. The center of each **slice** represents the ideal location of each signal plot, but in practice the line will show a group of dots around this ideal point, as shown in **Figure 19**. If the dot representing a signal value sample moves close to and eventually over the line to the next **slice** it will be interpreted as **that** bit combination; a line error.

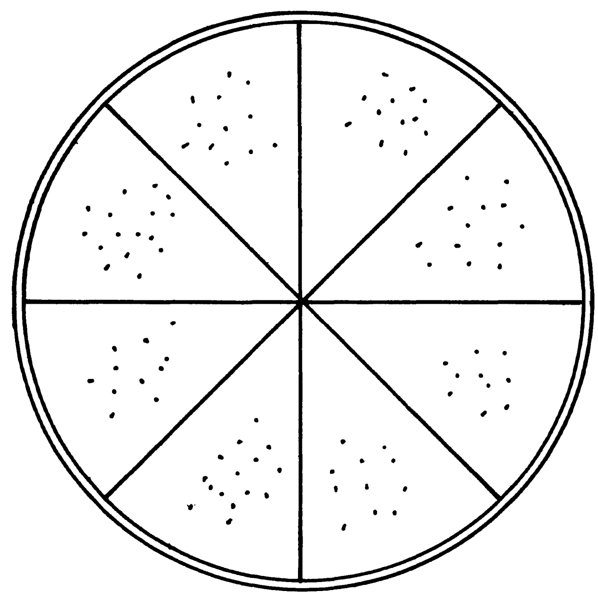


Figure 19 • an eye pattern displayed on CRT.

Given the eye pattern concept, it is possible to deduce the effects of various types of noise on the line. **Figure 20** shows some of these effects. If amplitude distortion is present on the line, signal amplitude and thus the distance between the dots and the center

Communication Test Equipment—A Solution To Increased Productivity

of the display will vary. This will cause the pattern to **smear** along a line from the center of the display to the ideal location. Severe amplitude distortion will cause the signal to fall outside the allowed range of amplitudes and be missed. Phase jitter on the line will not affect the amplitude, but will cause the pattern to smear at a tangent to the circular pattern of the display as the phase of each signal is altered. If enough jitter is encountered the samples will cross the **slice line** and be misinterpreted. A high rate of random noise will cause the amplitude and phase to vary roughly in proportion and thus expand the cluster within each slice, again offering the possibility of the sample being lost or moving to an incorrect area of interpretation.

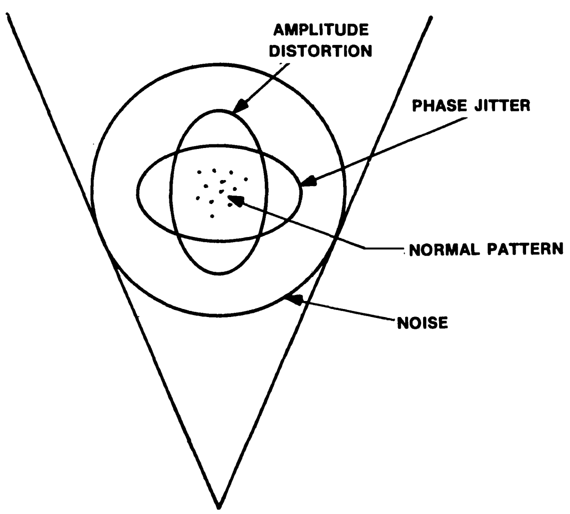


Figure 20 • effects of line problems on the eye pattern.

The eye pattern is one of the most useful indirect tests of the line because it is easily interpreted and because it shows line problems in terms of their effect on transmission, the bottom line of analysis to most users. Some manufacturers offer eye patterns as part of network control and management systems, others provide a modem interface for an oscilloscope to display the patterns. There are also some high-cost analog testing devices which will display an eye pattern. Users who would like to consider the use of eye patterns in line diagnostics should check with their modem vendors.

■ SELECTING THE TEST EQUIPMENT FOR YOUR NETWORK

Even if your organization has an unlimited test equipment budget it is doubtful if you would want to buy all of the equipment covered in this report. With practical financial restrictions on the amount of equipment and personnel skill restrictions on the type and complexity, you may find that selecting the right test equipment is very difficult. There may be significant differences in network reliability riding on your choices. You may also want to consider multiple units of certain types if you have multiple locations and extra diagnostic personnel, further complicating the problem. The best starting point in selecting equipment is to ask the diagnostic specialists to recommend and justify the type of equipment that would be most useful. Try to evaluate the type of problem the recommended equipment would solve against the questions, **how often will the problem occur**? and **how long will I be down with and without this gear**?

If you do not have adequately skilled personnel to help select the equipment (you must plan to train them to use it in this case!), here are some recommendations:

1. **Every diagnostic technician should have an EIA breakout box.** They should all be the same type, the same model, and there should be common checks and procedures defined for their use.

2. **Each computer site should have a BERT tester and/or fox box for circuit testing.** If the modem has BERT capabilities, get a fox box or terminal simulator. You may have to break this rule if you are an SNA operation because fox testers for SNA are very expensive.

3. **If you have additional test funds to spend, consider a data line monitor.** These units start at less than $3,000 and can be very helpful in diagnosing soft errors. Select a model whose complexity and features match the skill levels of the test personnel.

4. **If you have a large number of circuits and require analog line quality measurements**, try to get a modem system which provides them. The capacity for remote testing, backup, and monitoring of these systems will normally be needed in operations who feel they need technical interaction with telephone circuit service personnel.

5. **Spend money to upgrade the previous types of test equipment** before you buy any specialized analog gear unless you have private circuits.

6. **Allocate at least 10 percent of your budget for test equipment to training** the people who use it and the development of diagnostic procedures. A good test plan supported by a pair of breakout boxes will solve a problem faster than a pair of $20,000 line monitors and no organized approach to the problem.

Every user network is to some degree unique, with its own technical characteristics, reliability requirements, and business vulnerabilities. Know your own network, and select test equipment which supports your operation properly, or your expenses in this area will never pay dividends and you lose in productivity.

• **END**

Facsimile Terminals
Evaluation & Selection

■ INTRODUCTION

The concept of facsimile transmission was spawned more than a century ago, but remained relatively dormant until the age of electrical and electronic technology, that its application became feasible. Early usage was dedicated to Western Union telegraph, photographs for wire services, and weather maps. Facsimile has since spread to other dedicated applications such as police mug shots and finger prints, however, it never really became popular for general business applications. Why? Because it had been an expensive alternative to sending information. The relatively low rate of transmission per page (as much as six minutes) resulted in substantial telephone usage charges where facsimile transmission is conducted over long distances. More sophisticated equipment speeds up the per-page transmission rate, but the user is penalized by increased equipment charges. Therefore, high-speed equipment is applicable only to high-volume usage for operating economy.

Facsimile is an extremely useful and flexible means of information transmission. It can handle virtually all forms of information: graphics, photographs, typed copy, etc. And the copy received at the destination is identical to the original, hence the name facsimile. Recent improvements in facsimile technology and the attention gained by electronic mail as an attractive alternative to postal service, has renewed interest in facsimile for general business applications. Many companys are realizing it's the only practical solution to their needs. Over the past two years, IBM and Rapicom have introduced products which allow facsimile to be extended to electronic mail and store-and-forward applications. Rapicom's R-5000 with the SAF-PAK option will store up to 60 pages (8.5x11-inch) and distribute them up to 100 different locations. This product will also accept and store pages from similarly equipped R-5000s and distribute them to their indicated destinations. All of this is done without aid of a computer.

IBM's new Scanmaster Model 8815 allows a document to be routed to multiple locations, but requires a host computer to store the transmitted pages and provide the routing services. Scanmaster can also run IBM's Displaywriter, 8100 Distributed Office Support systems, and the 5520 Administrative System. In this environment, Scanmaster employs the services of IBM's Distributed Office Support System (DIOSS/370) contained in the host's front end to handle the inputs and route the documents.

Rapicom also provides a link to the host processor through its Intelligent I interface. This piece of firmware/hardware allows Rapicom's R-3100 and R-3300 transceivers to interact with a host and furnishes asynchronous, HDLC, and IBM 2780/3780 protocol support. Thus many different mainframes, including DEC's PDP-11 and compatible IBM systems, to be interfaced. Intelligent I also accepts ASCII and EBCDIC file data, allowing the facsimile unit to act like a printer.

In the coming year, expect to see transceivers with multilink facilities, and software which will support selective routing of messages to different locations. This will further extend the appeal of facsimile as an electronic mail medium.

In choosing a facsimile terminal, users must consider such facilities as scan-image resolution, document feeding capabilities, compatibility with CCITT standards and other facsimile equipment, data transmission speeds, etc. This tutorial provides a basic understanding of facsimile technology and advises the prospective user in evaluating and selecting terminals to service specific needs.

■ THE CONCEPT OF FACSIMILE COMMUNICATION

A facsimile terminal is essentially a copier equipped to transmit and receive graphic images. It scans a source document, and converts the scanned image into an electrical signal that is transmitted over a communication medium (typically a telephone line) to a remote facsimile terminal, which converts the electrical signal to a graphic image reproduced on paper identical to the source document. The input document can contain virtually any form of information, including alphanumeric data, photographs, charts, and graphs.

Facsimile terminals are classified as send-only terminals, receive-only terminals, and transceivers. To communicate with one another, sending and receiving terminals must employ the same scanning technique, the same rate of scanning, and have the same electrical signal characteristics. While facsimile terminals are often only compatible with identical terminals, many are now being designed to communicate with models from other vendors.

Encoding Techniques • The information scanned from the document is converted into an analog or digital signal. With analog conversion, the scanning mechanism produces an analog signal corresponding to data and blank spaces on the document. The digital conversion process translates the data (considered black spaces) and blanks (considered white) into binary ones and zeroes, respectively. This process is called run-length encoding. Some analog machines will ignore white spaces while scanning, thus speeding overall processing. Run-length encoding is generally associated with data compression. Using data compression, the terminal analyzes the data for repetitive sequences of white and black spaces, and assigns a code to each sequence. Run-length encoding is classified as either one- or two-dimensional. One-dimensional encoding compresses data in a horizontal direction only; two-dimensional encoding also compresses data in a vertical direction. Terminals that use data compression techniques, more effectively use the communications facility by transmitting virtually no redundant data. Most digital facsimile terminals that compress data use a modified Huffman code to encode compressed data. This code is the one adapted by the CCITT Group 3 Standards Committee, and is described under Terminal Compatibility.

Data Communication • Facsimile terminals are connected to a communication medium through a modem, which converts the facsimile signal to a form suitable for transmission, and converts the received signal to one suitable to produce a graphic image. AM (Amplitude Modulation) or FM (Frequency Modulation) modulation techniques may be used.

Many terminals are equipped with an integral modem suitable for connection to a variety of transmission facilities. For transmission at speeds up to 9600 bps, users generally have the option of using the Direct Distance Dialing (DDD) switched network, or they choose to rent private voice-grade telephone lines, which may or may not be conditioned.

If the DDD is used, make sure that the integral modem either includes its own data access arrangement (DAA), or that it has been certified by the FCC for direct connection to the DDD network. DAAs are required by law to protect the telephone network from power surges which can interfere with normal telephone service. While they are not particularly expensive to buy or lease, they do represent an added expense. DAAs are not required on private lines.

Low-volume facsimile terminals operating at speeds of 1200 bps or lower can use an acoustic coupler in lieu of a modem for connection to a conventional telephone. While acoustic couplers generally provide acceptable communications service, they are susceptible to background noise, which can introduce errors.

High-volume facsimile terminals that transmit at speeds ranging from 19.2K to 56K bps require either wideband service or AT&T Communication's Dataphone Digital Circuits (DDC). Wideband

Facsimile Terminals

links and the DDC are relatively expensive and are not always available, but they do provide a high level of service for volume applications. The DDC does not use a conventional modem for connection to the network, but requires a Data Service Unit (DSU) or Channel Service Unit (CSU) rented from the carrier.

Compatibility • Manufacturers have traditionally produced facsimile equipment which was only compatible with their own products. While part of the reason for such exclusivity was based on proprietary techniques that vendors carefully guarded, marketing considerations played a very strong role in compatibility decisions.

The expansion of business applications for facsimile, especially electronic-mail prompted the need to communicate with other manufacturers equipment. Equipment compatibility has become an important element in most marketing plans. The Consultative Committee for International Telegraph and Telephone (CCITT), has established recommendations for protocol and transmission techniques, scan direction, scan-line size, scans per millimeter, phasing, synchronization, and modulation techniques, so that machines from different manufacturers can communicate with one another. These standards, officially adapted in 1981, are divided into three groups and include analog AM/FM, and digital equipment.

• CCITT Group I defines equipment that uses an FM analog modulation technique and transmits a standard 8.5x11-inch document in six minutes. The recommendation also allows a four-minute transmission as an option.

• CCITT Group II defines equipment that uses an AM analog modulation technique and transmits a standard document in three minutes. A two-minute transmit speed is optional.

• CCITT Group III defines equipment that uses digital encoding techniques and transmits a standard page in a minute or less. The standard data compression technique is a modified Huffman code, which is a one-dimensional horizontal scheme. As an option, a two-dimensional scheme called modified READ may be implemented.

In addition to the three standards, a fourth (called Group IV) is being formulated that will define high-resolution digital equipment operating in the range of 56K bps. The resolution will be at least 240x240 lines per inch.

■ SELECTION CONSIDERATIONS

The selection of facsimile terminals is not an easy task. The user must first consider his current and future information handling needs. Consideration must be given to the type or class of information to be handled. Alpha-numeric data does not require high-resolution terminals, unless it is fine print. The user can reduce equipment cost by selecting a low-resolution terminal if it adequately serves his needs. Highly detailed documents require high-resolution equipment, which drives up equipment cost and extends per-page transmission time, resulting in additional connect charges if using the DDD network. Applications that typically handle several documents at one time (sequentially) require a document loader to eliminate the time spent to manually load the machine. Unattended sites require an auto-answer feature and automatic paper loading. These are also requirements for "after-hours" transmission to an unoccupied location. Data compression cuts communication costs by eliminating the transmission of redundant information. The transmission characteristics of the terminal, especially transmission speed, determine its compatibility with competitive terminals and the document transmit time, which is important if the DDD network is used. All these factors should be carefully considered before making a buying decision.

Resolution • image resolution is defined as the scan density, or number of stops per unit distance, typically specified by vendors in lines per inch. Both horizontal and vertical resolution is specified. Generally, horizontal resolution is greater than vertical, but quite often the resolution of both are equivalent.

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. Very high resolution is required for photographs or data with several shades or tones of gray. For example, a scanner with a horizontal resolution of 96 lpi 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.

The user should be advised that price is proportional to resolution, and transmission speed is inversly proportional to resolution. The greater the resolution, the more expensive the terminal and the slower the transmission speed, unless the scanning rate is increased to compensate for increased resolution.

Most terminals provide a variable scan rate which is either automatically or manually set. Some vendor literature will list the ranges, but quite often only the maximum rates are presented. Keep in mind that high-resolution terminals are more expensive than those with lower resolution; so don't overbuy. Scanners are described under Scanning/Printer Technology.

Choosing the right printing technique is just as important as selecting the correct scanner. For straight text printing almost any unit will do. The cost of the unit and its paper will be the determining factor. For high resolution, more sophisticated (and costly) reproduction techniques are required. Printing techniques are described under Scanner/Printer Technology.

Document Handling • terminals feed documents automatically or manually. Automatic feeding is generally handled by fixed-bed devices that employ stack/roller feeders, or drum feeders. Manual feeding is done by hand, and ranges from something as simple as feeding a pre-cut sheet into a feeding mechanism, to attaching the sheet to a drum. The key advantage of automatic feeding is that it theoretically allows the unit to run unattended. If that is an important consideration (or an absolute must), select a roll-fed unit for the receiving site. These units are less likely to jam.

Do not overlook the document-handling capability of the unit. While all will handle standard 8.5x11-inch documents, others will accommodate larger and smaller dimensions in both length and width. Also check the scanning width of the unit. A few products which handle 8.5-inch wide paper have a scanning width of only 8 inches.

The flexibility to set scanning margins is a definite advantage in document handling. Limiting the scanned area to the information portion of the document increases operating performance and transmission efficiency.

Transmission Characteristics • when selecting facsimile equipment, keep in mind that the speed at which a document is transmitted depends on the speed of the scanning device and the data transmission speed of the terminal. High rates for both equate to reduced communication cost. With digital devices, consider the range of transmission speeds supported. Most will communicate at speeds up to 4800 bps, while many others will run considerably faster.

Terminals conforming to CCITT recommendations will have their scan rates and transmit speeds established at specific settings. This is done to allow multivendor products which might ordinarily be incompatible to communicate with one another. This is further explained under compatibility. Thus, if your organization practices a multivendor policy, be sure the prospective facsimile equipment conforms to appropriate CCITT standards before you place an order.

Data Communication Costs • as was mentioned under Data Communication, a modem converts the output of the terminal into a form suitable for transmission. If the terminal lacks a modem, expect to spend from one to several thousand dollars per terminal (depending on the speed of the modem) to add this device. For low-speed transmission up to 1200 bps, an acoustic coupler might be a wise choice.

The transmission medium itself can be the DDD network, dedicated telephone lines, microwave, and satellite links. Of these choices, most users will go with the DDD network or dedicated lines.

If the DDD network is chosen, an obvious cost is the charge for the call. While those figures are set by the carrier and, are therefore, out of control of the user, the connect time can be reduced by choosing a device with fast scan and data communication rates.

Facsimile Terminals

Another cost associated with dial network connection is the Direct Access Arrangement (DAA). DAAs are available from several vendors, and the costs vary. Many facsimile vendors who supply integral modems also supply the DAA as part of the package. Others supply modems which have been FCC-certified for connection to the public network eliminating the need for a DAA.

Dedicated telephone lines are usually chosen when the costs associated with the DDD network become excessive and/or a conditioned telephone line is required. Conditioning is a service provided by the carrier to improve line quality, where intolerable electrical conditions prevent error-free communications. Dedicated lines handle higher data rates than the dial-up network, and they do not require a DAA.

■ SCANNER/PRINTER TECHNOLOGY

The techniques used for scanning and printing are diverse. Some are relatively simple, while others are extremely complex and more expensive.

The following provides a brief discussion of the more commonly used scanning and printing techniques employed by facsimile equipment manufacturers.

Rotating Cylinder Scanner • the rotating cylinder scanner employs a cylinder (drum) and a photocell scan head. The document is attached to the drum manually, or it can be inserted into a feed slot and held in place by vacuum. The scan head can travel along a threaded rod that parallels the length of the document, or the scan head can be fixed with the cylinder moving laterally.

The lighting technique employed to illuminate the document varies from vendor to vendor. Some provide a lamp which covers a large portion of the document, while others use a light which concentrates on the area being scanned. In any event, the reflected light from the document is focused on the scanner photocell, which generates an analog signal. Generally, the entire document is scanned top to bottom. To eliminate the scanning and transmission of blank spaces associated with the document's top, bottom, and edge, some units provide adjustable "scanner stops" which are set by the user.

Flying-Spot Scanner • flying-spot scanning also employs a photocell to generate the analog signal, which represents data read from the input document. Three components make up a typical flying-spot terminal: a cathod-ray tube (CRT), a focusing lens, and the photocell. The document to be read is placed on a flat-bed platen, and an electron beam from the CRT scans each line much like the electron beam in a TV kinescope scans its phosphor screen. The beam provides the light source which is reflected off the document, through the lens, and into the photocell. Flying-spot scanners are often used to produce a high resolution image.

Solid-State Scanner • this is another scanning technique which generates an analog signal via photodiodes. However, instead of using a single photocell, a stationary array of photocells are employed. The scanning operation is accomplished by roller-feeding the document through an illuminated area passing the photodiode array. Reflected light from the document is focused through a line onto the photodiode array.

The photodiode array technique is much more sophisticated than other scanning methods that employ a photodiode. The photodiodes are arranged in a matrix which equates to one lateral scan line. Each diode acts independently to form an element of the scan line. At the receiver, these elements are reassembled to form the transmitted line.

Some terminals now on the market have charged-coupled diodes (CCD) in place of photodiodes. These devices function like an array of photodiodes, but are more effective.

Oscillating Scan Head • with the oscillating scan head, the document is positioned on a semi-cylinder platen that moves laterally beneath a pivoted scan head. The scan head oscillates or rotates within the semi-cylinder to capture the image. Each revolution/oscillative accounts for one scan line. Some vendors have replaced the scan head with an oscillating mirror that reflects light from the document, through an aperture, and onto a photocell.

Rotating Lens Turret Scanner • this technique also requires the document to be positioned on a moving semi-cylinder platen. Here, however, the optics consist of a stationary mirror surrounded by a rotating turret containing a number of lenses. Light reflected from the document passes through each lens and onto the mirror, which focuses it through an aperture and onto a photocell.

Rotating Helical Aperture Scanner • a rotating helical aperture also employs a mirror, lens, and photocell arrangement. The document is roller-fed over a flat-bed platen where it is illuminated by a light source. The reflected light from each scan line on the moving document is reflected by a mirror, through a focusing lens and fixed-slot aperture, and onto two rotating helical apertures.

In the dual aperture approach, each aperture rotates indefinately at different speeds. Light from the fixed aperture is passed through the second which reduces the scan line to a series of sequential picture elements that are sent to a photocell.

Laser Scanner • laser scanning employs a low-power laser light source to form a beam that scans the document. The movement of the beam is controlled by a rotating galvanometer which directs the beam a line at a time across the document. The reflected light is directed to a photocell that converts it to an electrical signal.

Electrolytic Printing • electrolytic printing employs paper treated with an electrolyte, and a stylus which passes the signal current through the paper to produce an image. The paper is roll-fed past the stylus, and changes color depending on the intensity of current passing through the stylus.

Electrothermal Printing • electrothermal printing uses a chemically treated paper coated with an electrosensitive material. An image is produced by passing a signal current through a stylus which burns off the electrosensitive layer, exposing the dark paper. On drum-fed machines, the stylus moves parallel to the rotating cylinder. On roll-fed printers, the stylus is positioned on a revolving belt that prints one line per revolution.

Dielectric Printing • dielectric printing uses a specially treated paper which is charge-sensitive. This paper is generally roller-fed past an electrode array where an electrical charge is applied on a line-by-line basis to form a latent image. The charged paper is passed through a toner; the toner adheres to the charged image, and heat fuses the toner to the paper creating the printed document.

Electrostatic Printing • this technique is quite similar to the dielectric printing process, except the image to be printed is first transferred to an intermediate light-sensitive drum which reproduces the image as an electrostatic charge. Printing is accomplished by applying toner to the drum, which adheres to the charged image. Plain paper is passed over the drum and the toner "image" is transferred to the paper. The electrostatic technique is very similar to that used by office photocopiers, and uses plain paper in place of more expensive treated paper.

Photographic Printing • with photographic printing, photopaper is wrapped around a cylinder in a light-tight box. The paper is then exposed to a light source that is controlled (modulated) by the incoming signal. The developing and printing process is quite similar to that of conventional photography.

Ink Jet Printing • ink jet printing directs tiny droplets of ink sprayed on plain paper. The images themselves are formed by infinitesimal dots.

• END