To help you understand the uses and benefits , this primer explains basic computer networking concepts and technology and also introduces computer networking terminology.

What Is a Computer Network?

On the most fundamental level, a computer network is an interconnected collection of devices that enables you to store, retrieve, and share information. Commonly connected devices include personal computers (PCs), minicomputers, mainframe computers, terminals, workstations, thin clients, printers, fax machines, pagers, and various data-storage devices. Recently, other types of devices have become network connectable, including interactive televisions, videophones, handheld devices, and navigational and environmental control systems. Eventually, networked devices everywhere will provide two-way access to a vast array of resources on a global computer network through the largest network of all, the Internet.

In today's business world a computer network is more than a collection of interconnected devices. For many businesses the computer network is the resource that enables them to gather, analyze, organize, and disseminate information that is essential to their profitability. The rise of intranets and extranets—business networks based on Internet technology—is an indication of the critical importance of computer networking to businesses. Intranets, extranets, and the Internet will be treated in more detail in a later section. For now, it is enough to understand that most businesses have installed intranets to collect, manage, and disseminate information more quickly and easily than ever before. They established intranets simply to remain competitive; now, the momentum continues, and extending the company network to the Internet is the next technological transformation of the traditional business.

What Are the Benefits of Computer Networking?

The most obvious benefit of computer networking is that you can store virtually any kind of information at, and retrieve it from, a central location on the network as well as access it from any connected computer. You can store, retrieve, and modify textual information such as letters and contracts, audio information such as voice messages, and visual images such as facsimiles, photographs, medical x-rays, and even video segments.

A network also enables you to combine the power and capabilities of diverse equipment and to provide a collaborative medium to combine the skills of different people—regardless of physical location. Computer networking enables people to share information and ideas easily, so they can work more efficiently and productively. Networks also improve commercial activities such as purchasing, selling, and customer service. Networks are making traditional business processes more efficient, more manageable, and less expensive.

Cost-Effective Resource Sharing

By networking your business computers you can reduce the amount of money you spend on hardware by sharing components and peripherals while also reducing the amount of time you spend managing your computer system.

Equipment sharing is extremely beneficial: when you share resources, you can buy equipment with features that you would not otherwise be able to afford as well as utilize the full potential of that equipment on your network. A properly designed network can result in both lower equipment costs and increased productivity.

Suppose that you had a number of unconnected computers. Employees using these computers would not be able to print unless you purchased a printer for each computer or unless users manually transferred files to computers with printers. In this scenario you would be choosing between hardware and labor expenses.

Networking the computers would give you other alternatives. Because all users could share any networked printer, you would not need to buy a printer for every computer. As a result, instead of buying numerous inexpensive, low-end printers that would sit idle most of the time, you could buy a few inexpensive printers and a few printers with high-end productivity features. The more powerful printers would be able to print more rapidly and with better quality than the less expensive ones. In addition, the more powerful printers might also be able to print in color and to sort, staple, or bind documents.

When you select the right mix of printers and assign each network user appropriate access to them, you have enough printing power to address the needs of all of your employees. Rather than leave expensive equipment idle, you provide your employees with the latest, most powerful productivity features—all for a significantly lower cost than if you were to purchase an inexpensive printer for each workstation on the network.

A network enables you to share any networkable equipment and realize the same benefits that you would enjoy from sharing printers. On a network, you can share e-mail systems, modems, facsimile machines, data storage devices such as hard disks and CD-ROM drives, data backup devices such as tape drives, and all network-enabled software. When you compare the costs associated with sharing these resources to the costs of purchasing them for each computer, the savings can be enormous.

A network also enables you to save money on software. Instead of buying separate copies of the same application for various machines, you can purchase one copy with enough user licenses for your network. In large businesses the amount of money saved on software is substantial.

Finally, you will also be able to reduce your administrative overhead. On a computer network, updates to software, changes in user information, and network security can all be accomplished from one location. With standalone computers you would be required to make these updates on each individual computer workstation.

Streamlined Business Processes

A well-designed computer network produces benefits on several fronts: within the company, between companies, and between companies and their customers. Within the company, networks enable businesses to streamline their internal business processes. Common tasks such as employee collaboration on projects, provisioning, and holding meetings can take less time and be much less expensive. For example, a managing editor, associate editors, writers, and artists may need to work together on a publication. With a computer network they can work on the same electronic files, each from their own computers, without copying or transferring files from a floppy disk. If the applications they are using feature basic integration with the network operating system (NOS), they can open, view, or print the same files simultaneously.

Provisioning, the process by which companies give new employees everything they need to get started (workstation, ID card, etc.), can be automated on a network. All the new employee's information can be entered into one terminal, and various departments such as properties, payroll, and security will receive that new information automatically. When an employee leaves the company, the process can be reversed just as easily.

Networks also make holding meetings more efficient. For example, collaboration software can search through a number of busy schedules to find time for a meeting—including the schedules of employees at different locations. The meeting can be held over the network through a teleconferencing session, thus eliminating the travel cost for those employees at remote sites. The attendees can simultaneously view and edit the same document and instantaneously view each other's changes as they are made. Moreover, they can do this without worrying about accidentally changing or deleting the work of others.

Freedom to Choose the Right Tool

A networking solution that enables data and resource sharing between different types or brands of hardware, operating systems, and communication protocols—an open networking environment—adds another dimension to the information-sharing capabilities inherent in computer networking. Open networking products enable you to work on the type of computer best suited to your job requirements without encountering compatibility problems. They also allow you to choose the system that best works in your environment without sacrificing interoperability with other companies' systems.

The opposite of the open networking environment is the proprietary or homogeneous environment in which only one vendor's products are used. Proprietary environments tend to be most successful in small companies that do not require a wide range of functions from their network. Medium- and large-sized companies, however, find that one computing platform is often more appropriate for a particular task than another. In an open environment you can combine many kinds of workstations and systems to take advantage of the strengths of each. For example, Novell network users can use IBM personal computers (PCs) running any version of Windows or DOS, Macintosh computers running a version of the Macintosh operating system (OS), Sun workstations running the UNIX OS, and other types of computers all on the same network. You can use the computer equipment best suited to the work you do and your equipment will still be compatible with other systems. Most important, it will be compatible with systems in other companies.

Powerful, Flexible Collaboration between Companies

When two or more companies connect selected portions of their networks, they can streamline business processes that normally occupy inordinate amounts of time and effort and that often become weak points in their productivity. For example, a manufacturing company that grants its suppliers access to the inventory control database on its network can drastically cut down on the time it takes to order parts and supplies. The network could be configured to alert suppliers immediately when the manufacturer needed a new shipment, the purchase order could be automatically generated, and authorization could be granted electronically—all over the network.

Improved Customer Relations

The most obvious way in which networks connect businesses to customers is through the electronic store front—a Web site where customers can search for and order products and services over the Internet. Many customers enjoy the convenience of shopping at home, and many businesses enjoy the expense saved over maintaining several physical "brick and mortar" stores. But networks provide customers with more benefits than simple convenience: they also make it easier for businesses to customize services for each customer and to respond more quickly to customer concerns.

Networks speed the flow and analysis of data so that businesses can determine which products their customers want most at each of their physical stores, for example, or so they can catalog and analyze customer complaints and make necessary improvements faster and more efficiently. Companies that maximize the capacities of their networks gather, analyze, and disseminate critical marketing information quickly, which can give them an advantage over their competitors.

Secure Management of Sensitive Information

Another significant advantage of computer networking is the ability to protect access to network resources and files. A network that is properly designed has extremely powerful security features that enable you to control who will have access to sensitive data, equipment, and other resources. This control can be exercised over both your own employees and those outside your company who access your system over the Internet.

Worldwide, Instantaneous Access to Information

If you choose a networking platform that offers a full suite of products—including robust directory services—and one that supports open standards, you will be able to securely connect heterogeneous computing equipment located at geographically separated sites into one cohesive network. As a result, you will be able to disseminate critical information to multiple locations anywhere in the world, almost instantaneously.

When you implement a business intranet, you can create or update information and make it accessible to all company employees easily and immediately. With Web publishing tools and a World Wide Web server running on your intranet you can create or change any information, and you can have that information automatically and instantaneously published on your Web server.

With access to your business's intranet and Web server, your employees will be able to access any new or updated information from anywhere in the world within a few seconds after it is published. The Internet provides the low-cost backbone for global access to your intranet. Web browsers and other intranet tools make it easy for even a novice computer user to access the information and intranet resources they need.

Integrated, flexible information sharing, instantaneous information updating and access, lower equipment costs, flexible use of computing power, secure management of sensitive information—these are the benefits of computer networking. With a properly designed and implemented network, you increase efficiency, productivity, and profitability.

The remainder of this primer is divided into sections designed to explain the fundamentals of computer networking as well as define the various technologies with which it is associated. The following topics will be explained (in this order) in the corresponding sections:

- Application Software—Introduces computer applications and their function both on standalone computers and in a network.
- Desktop Operating System—Explains the role of the desktop operating system as the link between the application, the computer hardware, and the rest of the network.
- Data Transmission—Details how information must be converted into electronic data and then transmitted from one computer to another through the various levels of the Open Systems Interconnection (OSI) model.
- Hardware Technology—Defines the hardware required to connect computers on a network.
- Network Operating System—Explains how the network operating system serves as the control center of the entire network.
- Network Topologies—Explains the configuration options of the various types of computer networks.
- Internetworking—Explains how networks can be expanded, combined, or partitioned.
- Real-World Networking—Examines the implementation of an actual versus a theoretical network.
- Important LAN and WAN High-Speed Technologies—Explains several technologies used in both local area network (LAN) and wide area network (WAN) environments that provide high-speed data transfer.
- Internet Technology—Explains how the Internet has affected modern computer networking and how Internet technologies are now being used in business networks.
- Network Management—Explains the complex nature of network management, including extended sections on network security and directory services.

These sections are arranged to guide you from the most fundamental aspects of computer networking (the user interface) to the more complex (high-speed technologies and network management). Each section builds upon the information discussed in previous sections.

Application Software

Applications are software packages that you use to do your work. For example, a word processor is an application with which you create and modify documents such as business letters and mailing lists. Applications work at the highest level of computer networking. You use applications as an interface through which you can access the resources of the computer as well as resources on the network to which your computer is connected. Commonly used application software includes word processing, accounting, spreadsheet, database, and e-mail programs. You may even use customized or one-of-a-kind applications built specifically for your company.

One important issue to consider when selecting application software is its degree of network and intranetwork integration. Not all applications are designed for network use. To effectively use network and intranet services, application software must be well suited to the computer network environment. The level of network integration built into any application determines how well you are able to collaborate with others, whether you can access network services, and how easy the application is to manage across the network.

Applications cannot function by themselves: they require resources provided by the computer hardware such as memory, storage devices such as hard disks, and peripheral devices (printers, fax machines, modems, etc.). For example, while using an application you might need to store documents on the hard disk in the computer on which the application is running. However, applications do not have the capacity to run hardware. An operating system, on the other hand, is software that controls the hardware, and therefore acts as an intermediary between applications and hardware. If you need to store or “save” a document on the hard drive, you would employ the application’s conventions to give the save command (such as a certain keystroke); in turn, the application would pass the command to the operating system, which would direct the hardware to record the document on the hard drive. The diagram below illustrates the interaction between the application software, the operating system, and the computer hardware.

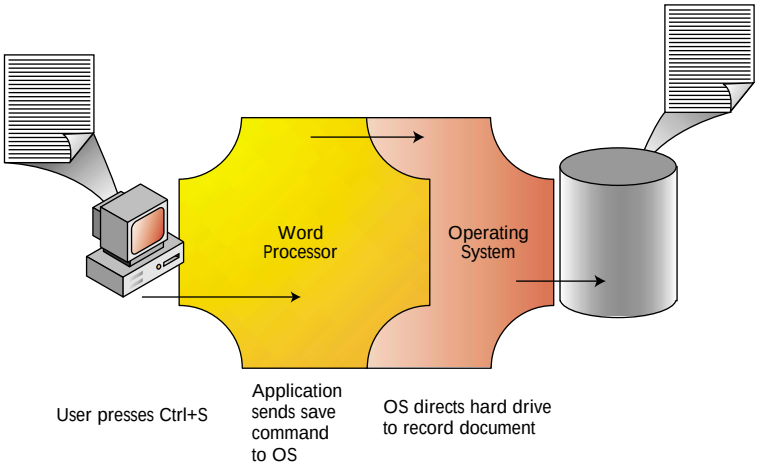


Figure 1

The function of applications and operating systems within a computer

Once you understand how applications function in this scenario, it is easy to see the importance of the operating system through which the application gains access to network resources and services. There are two types of operating systems necessary in computer networking: the desktop operating system and the network operating system. The next section discusses how the desktop operating system creates the environment necessary for application software to do its work.

Desktop Operating Systems

Each workstation on the network must have desktop operating system software that manages the interaction between the workstation's applications and its other resources. There are various commonly used desktop operating systems, including Windows 2000, Windows NT, Windows 95/98, Windows 3.x, UNIX, PC-DOS, OS/2, Linux, MS-DOS, and several versions of the Macintosh operating system. With any of these operating systems a workstation can be used to access files from local hard disks, display information on a monitor, coordinate local printing, and so on.

The desktop operating system controls access to computer resources, storage devices, and any peripheral devices. It also contains very basic networking abilities, allowing you to share information with users on other computers. Two or more computers running the same operating system can be hooked together, using appropriate hardware, to form a simple network by which the computers can share information. This sharing of information is the basis of computer networking. Although this type of network is limited in its capabilities and not often used in today's businesses, it will serve to introduce the concepts of computer networking.

Sharing information between computers, even on a simple network, is a complex process. The information from the application of origin must be converted into electronic data and then sent through the operating system to the hardware that connects the two computers. The receiving computer must then decode the electronic data it receives from the connecting hardware and reconfigure it so it will be recognized by the receiving application. This process involves a complex series of events and some very specific networking hardware. The process of converting information into electronic data and then moving it from one computer to another is explained in the following section, "Data Transmission."

Data Transmission

Although we routinely use the terms "data" and "information" interchangeably, they are not technically the same thing. Computer data is a series of electrical charges arranged in patterns to represent information. In other words, the term "data" refers to the form of the information (the electrical patterns), not the information itself.

Conversely, the term "information" refers to data that has been decoded. In other words, information is the real-world, useful form of data. For example, the data in an electronic file can be decoded and displayed on a computer screen or printed onto paper as a business letter.

Encoding and Decoding Data

To store meaningful information as data and to retrieve the information, computers use encoding schemes: series of electrical patterns that represent each of the discrete pieces of information to be stored and retrieved. For example, a particular series of electrical patterns represents the alphabetic character "A." There are many encoding schemes in use. One common data-encoding scheme is American Standard Code for Information Interchange (ASCII).

To encode information into data and later decode that data back into information, we use electronic devices, such as the computer, that generate electronic signals. Signals are simply the electric or electromagnetic encoding of data. Various components in a computer enable it to generate signals to perform encoding and decoding tasks.

To guarantee reliable transmission of this data across a network, there must be an agreed-on method that governs how data is sent, received, and decoded. That method must address questions such as: How does a sending computer indicate to which computer it is sending data? If the data will be passed through intervening devices, how are these devices to understand how to handle the data so that it will get to the intended destination? What if the sending and receiving computers use different data formats and data exchange conventions—how will data be translated to allow its exchange?

In response to these questions, a communication model known as the OSI model was developed. It is the basis for controlling data transmission on computer networks. Understanding the OSI model will allow you to understand how data can be transferred between two networked computers.

ISO and the OSI Model

The OSI model was developed by the International Organization for Standardization (ISO) as a guideline for developing standards to enable the interconnection of dissimilar computing devices. It is important to understand that the OSI model is not itself a communication standard. In other words, it is not an agreed-on method that governs how data is sent and received; it is only a guideline for developing such standards.

The Importance of the OSI Model

It would be difficult to overstate the importance of the OSI model. Virtually all networking vendors and users understand how important it is that network computing products adhere to and fully support the networking standards this model has generated.

When a vendor's products adhere to the standards the OSI model has generated, connecting those products to other vendors' products is relatively simple. Conversely, the further a vendor departs from those standards, the more difficult it becomes to connect that vendor's products to those of other vendors.

In addition, if a vendor were to depart from the communication standards the model has engendered, software development efforts would be very difficult because the vendor would have to build every part of all necessary software, rather than being able to build on the existing work of other vendors.

The first two problems give rise to a third significant problem for vendors: a vendor's products become less marketable as they become more difficult to connect with other vendors' products.

The Seven Layers of the OSI Model

Because the task of controlling communications across a computer network is too complex to be defined by one standard, the ISO divided the task into seven subtasks. Thus, the OSI model contains seven layers, each named to correspond to one of the seven defined subtasks.

Each layer of the OSI model contains a logically grouped subset of the functions required for controlling network communications. The seven layers of the OSI model and the general purpose of each are shown in Figure 2.

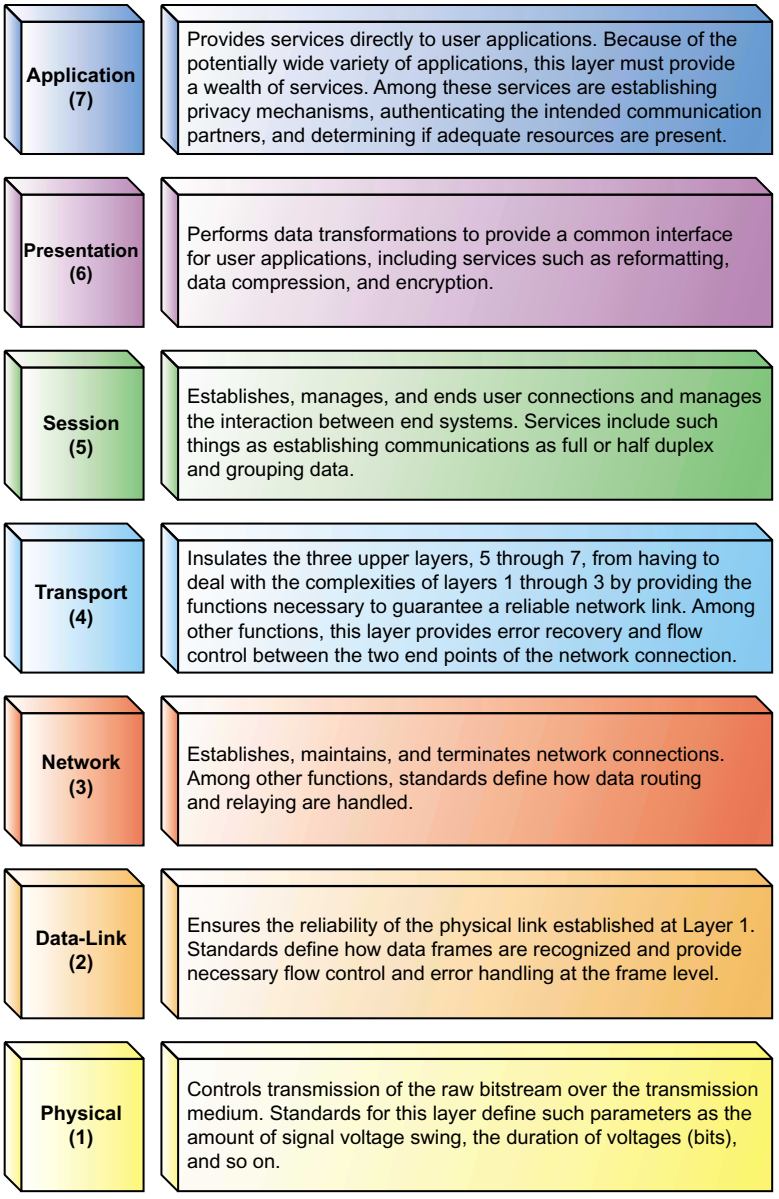


Figure 2

The OSI model

Network Communications through the OSI Model

Using the seven layers of the OSI model, we can explore more fully how data can be transferred between two networked computers. Figure 3 uses the OSI model to illustrate how such communications are accomplished.

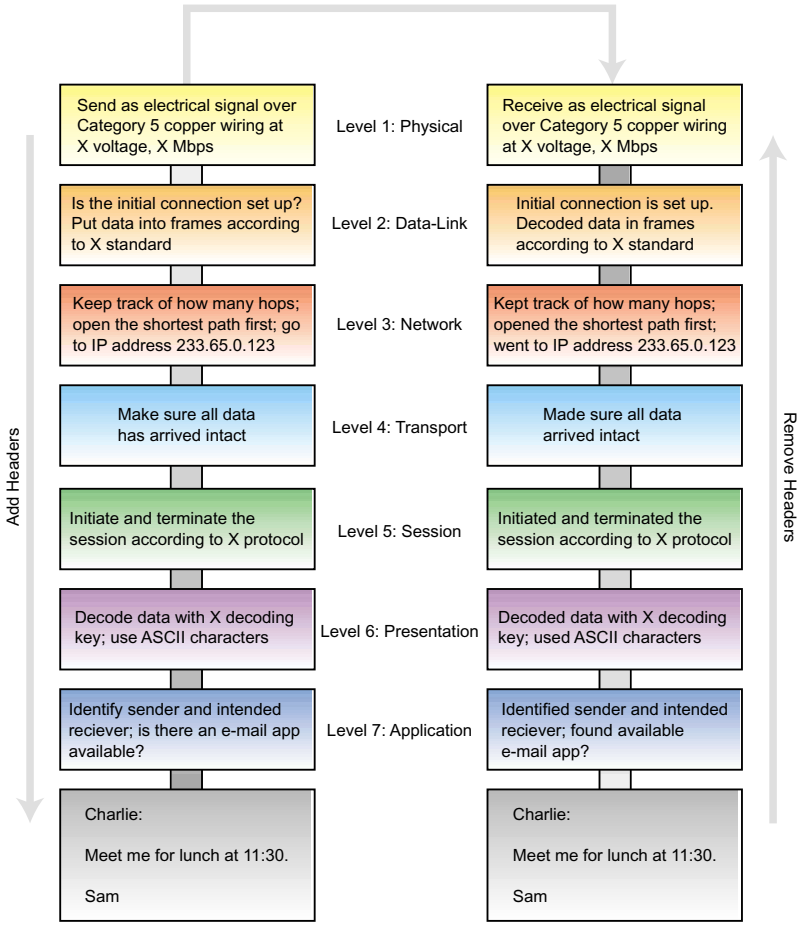


Figure 3

Networked computers communicating through the OSI model

The figure represents two networked computers. They are running identical operating systems and applications and are using identical protocols (or rules) at all OSI layers. Working in conjunction, the applications, the OS, and the hardware implement the seven functions described in the OSI model.

Each computer is also running an e-mail program that is independent of the OSI layers. The e-mail program enables the users of the two computers to exchange messages. Our figure represents the transmission of one brief message from Sam to Charlie.

The transmission starts when Sam types in a message to Charlie and presses the "send" key. Sam's operating system appends to the message (or "encapsulates") a set of application-layer instructions (OSI Layer 7) that will be read and executed by the application layer on Charlie's computer. The message with its Layer 7 header is then transferred to the part of the operating system that deals with presentation issues (OSI Layer 6) where a Layer 6 header is appended to the message. The process repeats through all the layers until each layer has appended a header. The headers function as an escort for the message so that it can successfully negotiate the software and hardware in the network and arrive intact at its destination.

When the data-link-layer header is added at Layer 2, the data unit is known as a "frame." The final header, the physical-layer header (OSI Layer 1) tells the hardware in Sam's computer the electrical specifics of how the message will be sent (which medium, at which voltage, at which speed, etc.). Although it is the final header to be added, the Layer 1 header is the first in line when the message travels through the medium to the receiving computer.

When the message with its seven headers arrives at Charlie's computer, the hardware in his computer is the first to handle the message. It reads the instructions in the Layer 1 header, executes them, and strips off the header before passing the message to the Layer 2 components. These Layer 2 components execute those instructions, strip off the header, and pass the message to Layer 3, and so on. Each layer's header is successively stripped off after its instructions have been read so that by the time the message arrives at Charlie's e-mail application, the message has been properly received, authenticated, decoded, and presented.

Commonly Used Standards and Protocols

National and international standards organizations have developed standards for each of the seven OSI layers. These standards define methods for controlling the communication functions of one or more layers of the OSI model and, if necessary, for interfacing those functions with the layers above and below.

A standard for any layer of the OSI model specifies the communication services to be provided and a protocol that will be used as a means to provide those services. A protocol is a set of rules network devices must follow (at any OSI layer) to communicate. A protocol consists of the control functions, control codes, and procedures necessary for the successful transfer of data.

More than one protocol standard exists for every layer of the OSI model. This is because a number of standards were proposed for each layer, and because the various organizations that defined those standards—specifically, the standards committees inside these organizations—decided that more than one of the proposed standards had real merit. Thus, they allowed for the use of different standards to satisfy different networking needs. As technologies develop and change, some standards win a larger share of the market than others, and some dominate to the point of becoming "de facto" standards.

To understand the capabilities of computer networking products, it will help to know the OSI layer at which particular protocols operate and why the standard for each layer is important. By converting protocols or using multiple protocols at different layers of the OSI model, it becomes possible for different computer systems to share data, even if they use different software applications, operating systems, and data-encoding techniques.

Figure 4 shows some commonly used standards and the OSI layer at which they operate.

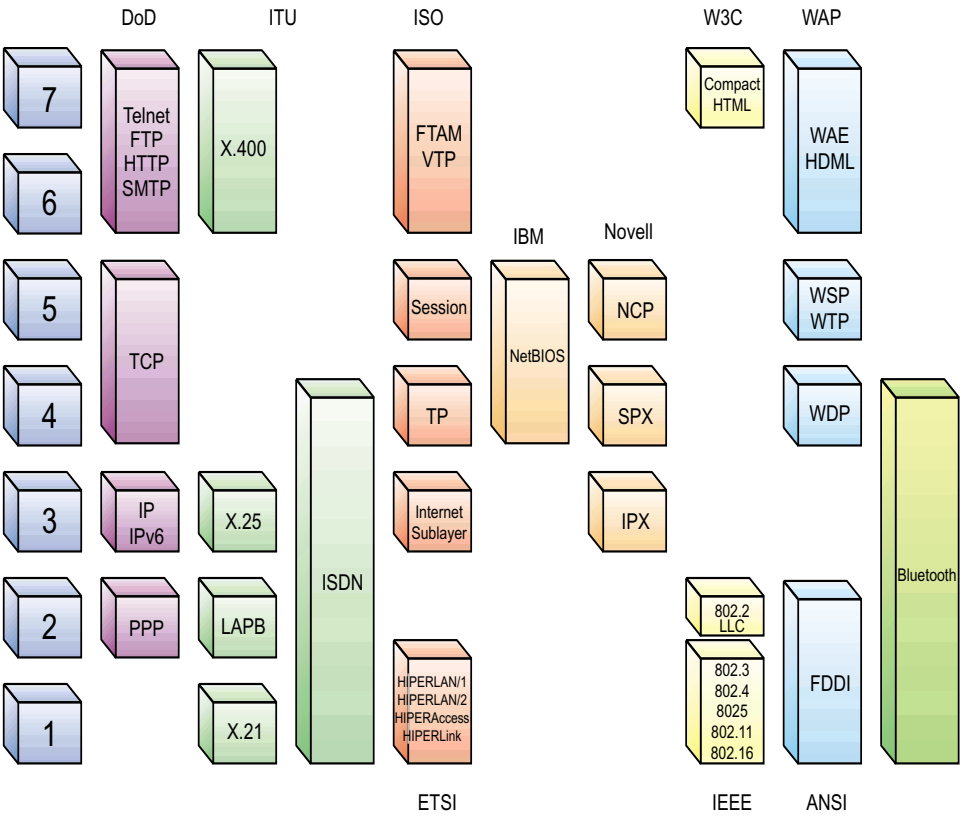


Figure 4

Important standards at various OSI layers

Layer 7 and Layer 6 Standards: Application and Presentation

The application layer performs high-level services such as making sure necessary resources are present (such as a modem on the receiving computer) and authenticating users when appropriate (to authenticate is to grant access after verifying that the you are who you say you are). The presentation layer, usually part of an operating system, converts incoming and outgoing data from one presentation format to another. Presentation-layer services include data encryption and text compression. Most standards at this level specify Layer 7 and Layer 6 functions in one standard.

The predominant standards at Layer 7 and Layer 6 were developed by the Department of Defense (DoD) as part of the Transmission Control Protocol/Internet Protocol (TCP/IP) suite. This suite consists of the following protocols, among others: File Transfer Protocol (FTP), the protocol most often used to download files from the Internet; Telnet, which enables you to connect to mainframe computers over the Internet; HyperText Transfer Protocol (HTTP), which delivers Web pages; and Simple Mail Transfer Protocol (SMTP), which is used to send e-mail messages. These are all Layer 7 protocols; the TCP/IP suite consists of more than 40 protocols at several layers of the OSI model.

X.400 is an International Telecommunication Union (ITU) standard that encompasses both the presentation and application layers. X.400 provides message handling and e-mail services. It is the basis for a number of e-mail applications (primarily in Europe and Canada) as well as for other messaging products. Another ITU standard in the presentation layer is the X.500 protocol, which provides directory access and management.

File Transfer, Access, and Management (FTAM) and Virtual Terminal Protocol (VTP) are ISO standards that encompass the application layer. FTAM provides user applications with useful file transfer and management functions. VTP is similar to Telnet; it specifies how to connect to a mainframe over the Internet via a “virtual terminal” or terminal emulation. In other words, you can see and use a mainframe’s terminal display on your own PC. These two standards have been largely eclipsed by the DoD standards.

Wireless Application Protocol (WAP) is a suite developed by the WAP Forum, whose members include many wireless device manufacturers and computer software and hardware companies, including Novell. WAP is for handheld devices such as cellular phones, pagers, and other wireless terminals that have limited bandwidth, screen size, memory, battery life, CPU, and user-interface controls. At the application and presentation layers is the Wireless Application Environment (WAE). WAE contains the Wireless Markup Language (WML), WMLScript—a scripting microlanguage similar to JavaScript—and the Wireless Telephony Application (WTA). Handheld Device Markup Language (HDML) and Handheld Device Transfer Protocol (HDTP) are also part of the WAP suite.

Compact HTML is defined by the World Wide Web Consortium (W3C) and is a subset of HTML protocols. Like WAP, it addresses small-client limitations by excluding functions such as JPEG images, tables, image maps, multiple character fonts and styles, background colors and images, and frame style sheets.

Layer 5 Standards: Session

As its name implies, the session layer establishes, manages, and terminates sessions between applications. Sessions consist of dialogue between the presentation layer (OSI Layer 6) of the sending computer and the presentation layer of the receiving computer. The session layer synchronizes dialogue between these presentation layer entities and manages their data exchange. In addition to basic regulation of conversations (sessions), the session layer offers provisions for data expedition, class of service, and exception reporting of problems in the session, presentation, and application layers.

Transmission Control Protocol (TCP)—part of the TCP/IP suite—performs important functions at this layer as does the ISO session standard, named simply "session." In a NetWare environment the NetWare Core Protocol™ (NCP™) provides most of the necessary session-layer functions. The Service Advertising Protocol (SAP) also provides functions at this layer. Both NCP and SAP are discussed in greater detail in the "Internetworking" section of this primer.

Wireless Session Protocol (WSP), part of the WAP suite, provides WAE with two session services: a connection-oriented session over Wireless Transaction Protocol (WTP) and a connectionless session over Wireless Datagram Protocol (WDP).

Wireless Transaction Protocol (WTP), also part of the WAP suite, runs on top of UDP and performs many of the same tasks as TCP but in a way optimized for wireless devices. For example, WTP does not include a provision for rearranging out-of-order packets; because there is only one route between the WAP proxy and the handset, packets will not arrive out of order as they might on a wired network.

Layer 4 Standards: Transport

Standards at this OSI layer work to ensure that all packets have arrived. This layer also isolates the upper three layers—which handle user and application requirements—from the details that are required to manage the end-to-end connection.

IBM's Network Basic Input/Output System (NetBIOS) protocol is an important protocol at this layer and at the session layer. However, designed specifically for a single network, this protocol does not support a routing mechanism to allow messages to travel from one network to another. For routing to take place, NetBIOS must be used in conjunction with another "transport mechanism" such as TCP. TCP provides all functions required for the transport layer.

WDP is the transport-layer protocol for WAP that allows WAP to be bearer-independent; that is, regardless of which protocol is used for Layer 3—USSD, SMS, FLEX, or CDMA—WDP adapts the transport-layer protocols so that WAP can operate on top of them.

Layer 3 Standards: Network

The function of the network layer is to manage communications: principally, the routing and relaying of data between nodes. (A node is a device such as a workstation or a server that is connected to a network and is capable of communicating with other network devices.) Probably the most important network-layer standard is Internet Protocol (IP), another part of the TCP/IP suite. This protocol is the basis for the Internet and for all intranet technology. IP has also become the standard for many LANs.

The ITU X.25 standard has been a common fixture in the network layer, but newer, faster standards are quickly replacing it, especially in the United States. It specifies the interface for connecting computers on different networks by means of an intermediate connection made through a packet-switched network (for example, a common carrier network such as Tymnet). The X.25 standard includes X.21, the physical-layer protocol and link access protocol balanced (LAPB), the data-link-layer protocol.

Layer 2 Standards: Data-Link (Media Access Control and Logical Link Control)

The most commonly used Layer 2 protocols are those specified in the Institute of Electrical and Electronics Engineering (IEEE): 802.2 Logical Link Control, 802.3 Ethernet, 802.4 Token Bus, and 802.5 Token Ring. Most PC networking products use one of these standards. A few Layer 2 standards under development or that have recently been proposed to IEEE are 802.1P Generic Attribute Registration Protocol (GARP) for virtual bridge LANs, 802.1Q Virtual LAN (VLAN), and 802.15 Wireless Personal Area Network (WPAN), which will define standards used to link mobile computers, mobile phones, and other portable handheld devices, and to provide connectivity to the Internet. Another Layer 2 standard is Cells In Frames (CIF), which provides a way to send Asynchronous Transfer Mode (ATM) cells over legacy LAN frames.

ATM is another important technology at Layer 2, as are 100Base-T (IEEE 802.2u), and frame relay. These technologies are treated in greater detail in the “Important WAN and High-Speed Technologies” section.

Layer 2 standards encompass two sublayers: media access control (MAC) and logical link control.

Media Access Control

The media access control protocol specifies how workstations cooperatively share the transmission medium. Within the MAC sublayer there are several standards governing how data accesses the transmission medium.

The IEEE 802.3 standard specifies a media access method known as “carrier sense multiple access with collision detection” (CSMA/CD), and the IEEE 802.4, 802.5, and fiber distributed data interface (FDDI) standards all specify some form of token passing as the MAC method. These standards are discussed in greater detail in the “Network Topologies” section.

The token-ring MAC method is not as prominent in computer networks as it once was: Ethernet, which uses CSMA/CD, has become the more popular networking protocol for linking workstations and servers. The token-ring technology of ARCnet (Attached Resource Computer network), however, has become the preferred method for embedded and real-time systems such as automobiles, factory control systems, casino games, and heating, ventilation, and cooling systems.

Logical Link Control

The function of the logical link control sublayer is to ensure the reliability of the physical connection. The IEEE 802.2 standard (also called Logical Link Control or LLC) is the most commonly used logical link control standard because it works with either the CSMA/CD or token-ring standards. The Point-to-Point Protocol (PPP) is another standard at this OSI level. This protocol is typically used to connect two computers through a serial interface, such as when connecting a personal computer to a server through a phone line or a T1 or T3 line. PPP encapsulates TCP/IP packets and forwards them to a server, which then forwards them to the Internet. The advantage to using PPP is that it is a "full-duplex" protocol, which means that it can carry a sending and a receiving signal simultaneously over the same line. It can also be used over twisted-pair wiring, fiber optic cable, and satellite transmissions.

Layer 1 Standards: Physical

Standards at the physical layer include protocols for transmitting a bitstream over media such as baseband coaxial cable, unshielded twisted-pair wiring, optical fiber cable, or through the air. The most commonly used are those specified in the IEEE 802.3, 802.4, and 802.5 standards. Use of the American National Standards Institute (ANSI) FDDI standard has declined as Ethernet has replaced token-ring technologies. Much of the FDDI market has largely been replaced by Synchronous Optical Network (SONET) and Asynchronous Transfer Mode (ATM). The different types of network cable and other network hardware will be discussed in greater detail in the "Hardware Technology" section.

Further Perspective: Standards and Open Systems

You probably noticed from looking at Figure 4 that most accepted standards do not include all (and only) those services specified for any OSI layer. In fact, most common standards encompass parts of multiple OSI layers.

Product vendors' actual implementation of OSI layers is divided less neatly. Vendors implement accepted standards—which already include mixed services from multiple layers—in different ways.

The OSI model was never intended to foster a rigid, unbreakable set of rules: it was expected that networking vendors would be free to use whichever standard for each layer they deemed most appropriate. They would also be free to implement each standard in the manner best suited to the purposes of their products.

However, it is clearly in a vendor's best interest to manufacture products that conform to the intentions behind the OSI model. To do this, a vendor must provide the services required at each OSI model layer in a manner that will enable the vendor's system to be connected to the systems of other vendors easily. Systems that conform to these standards and offer a high degree of interoperability with heterogeneous environments are called open systems. Systems that provide interoperability with components from only one vendor are called proprietary systems. These systems use standards created or modified by the vendor and are designed to operate in a homogeneous or single-vendor environment.

Hardware Technology

Now that we understand how information is converted to data and how computers send and receive data over the network, we can discuss the hardware used to transport the data from one computer to another. This hardware can generally be divided into two categories: network transmission media and transmitting and receiving devices. Network transmission media refers to the various types of media used to carry the signal between computers. Transmitting and receiving devices are the devices placed at either end of the network transmission medium to either send or receive the information on the medium.

Network Transmission Media

When data is sent across the network it is converted into electrical signals. These signals are generated as electromagnetic waves (analog signaling) or as a sequence of voltage pulses (digital signaling). To be sent from one location to another, a signal must travel along a physical path. The physical path that is used to carry a signal between a signal transmitter and a signal receiver is called the transmission medium. There are two types of transmission media: guided and unguided.

Guided Media

Guided media are manufactured so that signals will be confined to a narrow path and will behave predictably. The three most commonly used types of guided media are twisted-pair wiring, coaxial cable, and optical fiber cable.

Twisted-Pair Wiring

Twisted-pair wiring refers to a type of cable composed of two (or more) copper wires twisted around each other within a plastic sheath. The wires are twisted to reduce crosstalk (electrical interference passing from one wire to the other). There are "shielded" and "unshielded" varieties of twisted-pair cables. Shielded cables have a metal shield encasing the wires that acts as a ground for electromagnetic interference. Unshielded twisted-pair cable is the most common in business networks because it is inexpensive and extremely flexible. The RJ-45 connectors on twisted-pair cables resemble large telephone jacks.

Coaxial Cable

This type of cable is referred to as "coaxial" because it contains one copper wire (or physical data channel) that carries the signal and is surrounded by another concentric physical channel consisting of a wire mesh or foil. The outer channel serves as a ground for electrical interference. Because of this grounding feature, several coaxial cables can be placed within a single conduit or sheath without significant loss of data integrity. Coaxial cable is divided into two different types: thinnet and thicknet.

Thinnet coaxial cable is similar to the cable used by cable television companies. Thinnet is not as flexible as twisted-pair, but it is still used in LAN environments. The connectors on coaxial cable are called BNC twist-on connectors and resemble those found on television cables.

Thicknet is similar to thinnet except that it is larger in diameter. The increase in size translates into an increase in maximum effective distance. The drawback to the increase in size, however, is a loss of flexibility. Because thicknet is much more rigid than thinnet, the deployment possibilities are much more limited and the connectors are much more complex. Thicknet is used primarily as a network backbone with thinnet "branches" to the individual network components.

Optical Fiber Cable

10Base-FL and 100Base-FX optical fiber cable, better known as "fiber optic," are the same types of cable used by most telephone companies for long-distance service. As this usage would imply, optical fiber cable can transmit data over very long distances with little loss in data integrity. In addition, because data is transferred as a pulse of light rather than an electronic pulse, optical fiber is not subject to electromagnetic interference. The light pulses travel through a glass or plastic wire or fiber encased in an insulating sheath.

As with thicknet, optical fiber's increased maximum effective distance comes at a price. Optical fiber is more fragile than wire, difficult to split, and very labor-intensive to install. For these reasons, optical fiber is used primarily to transmit data over extended distances where the hardware required to relay the data signal on less expensive media would exceed the cost of optical fiber installation. It is also used where very large amounts of data need to be transmitted on a regular basis.

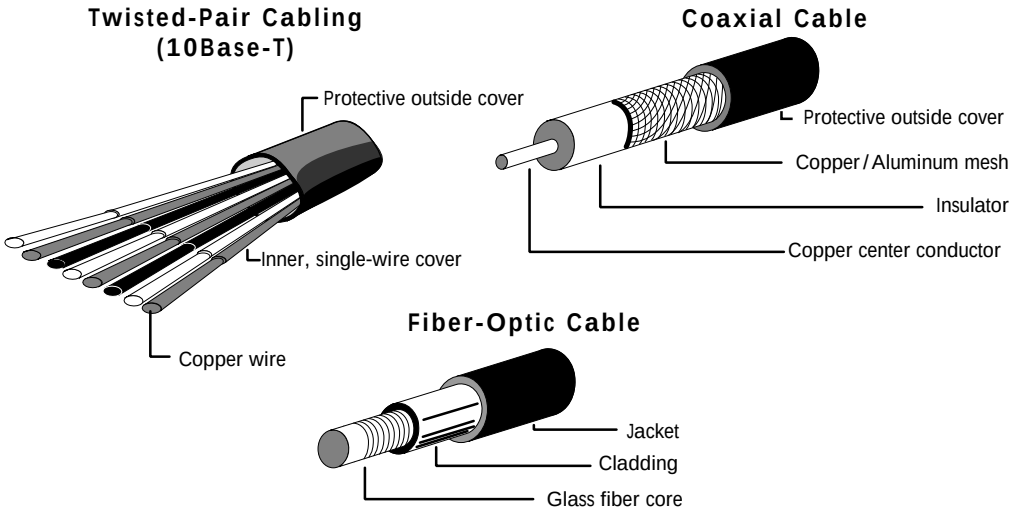


Figure 5

Common guided transmission media

Unguided Media

Unguided media are natural parts of the Earth's environment that can be used as physical paths to carry electrical signals. The atmosphere and outer space are examples of unguided media that are commonly used to carry signals. These media can carry such electromagnetic signals as microwave, infrared light waves, and radio waves.

Network signals are transmitted through all transmission media as a type of waveform. When transmitted through wire and cable, the signal is an electrical waveform. When transmitted through fiber-optic cable, the signal is a light wave: either visible or infrared light. When transmitted through Earth's atmosphere or outer space, the signal can take the form of waves in the radio spectrum, including VHF and microwaves, or it can be light waves, including infrared or visible light (for example, lasers).

Recent advances in radio hardware technology have produced significant advancements in wireless networking devices: the cellular telephone, wireless modems, and wireless LANs. These devices use technology that in some cases has been around for decades but until recently was too impractical or expensive for widespread consumer use. The next few sections explain technologies unique to unguided media that are especially of concern to networking.

Spread Spectrum Technology

Wireless transmission introduces several challenges not found in wired transmission. First is the fact that when data travels through the air, any device tuned to its frequency can intercept it, such as the way every radio in a city can pick up the same signal broadcast by a radio station. Second, if many devices transmitting on the same frequency are in the same geographical area, the signals can interfere with each other, a phenomenon known as crosstalk.

To prevent wireless transmissions from being intercepted by unauthorized devices and to reduce crosstalk, “spread spectrum” technology is used. A product of the military, spread spectrum technology has only recently become inexpensive and compact enough for use in commercial applications. As its name denotes, spread spectrum technology involves spreading a signal over a bandwidth larger than is needed, according to a special pattern. Only the devices at each end of the transmission know what the pattern is. In this way, several devices transmitting at the same frequency in the same location will not interfere with each other nor can they “listen in” on each other.

Spread spectrum technology can be performed using one of two techniques: Direct Sequence Spread Spectrum (DSSS) and Frequency Hopping Spread Spectrum (FHSS). In DSSS the sending device encodes the digital signal prior to transmission, using another digital signal as the key. The signal's power is then spread across a range of frequencies as it is transmitted. The receiving device has the same key, and upon receiving the transmission uses the key to interpret the signal. Because each connection between devices uses a unique key, the devices “hear” only those signals encoded with that key; all other signals are ignored. Also, by spreading a signal’s power over a broader-than-needed spectrum, several signals can be transmitted over the same range of frequencies without interfering with each other.

With FHSS the signal hops from one frequency to another in rapid succession and according to a pattern unique to that transmission. The Federal Communications Commission (FCC) requires that a minimum of 75 frequencies be used per transmission and that the maximum time spent on each frequency be no longer than 400 milliseconds. Because the device at the other end knows to which frequencies the signal will hop and for how long the signal will stay on each frequency, it knows where to find the signal each time it hops. Any other device using FHSS in the same geographical location would be looking for signals that hop frequencies according to a different pattern.

Each method has benefits and drawbacks. DSSS is the faster method of the two: it can achieve data transmission rates in excess of 2 Mbps whereas FHSS data transmission rates do not exceed 2 Mbps. DSSS is also more expensive and consumes more power. FHSS is therefore more cost-effective, but DSSS is best when higher data transfer rates are required.

Because of spread spectrum technology, data transmitted through the air is in many ways more secure than data transmitted over wires. With wired media the frequency at which data is sent remains constant, so a person with a good antenna and some skill could sit in the parking lot of a corporation and intercept unencrypted signals as they travelled over the wires. On the other hand, spread spectrum transmissions cannot be decoded except by the intended device.

Transmitting and Receiving Devices

Once you have selected a transmission medium, you need devices that can propagate signals across the medium and devices that can receive the signals when they reach the other end of the medium. Such devices are designed to propagate a particular type of signal across a particular type of transmission medium. Transmitting and receiving devices used in computer networks include network adapters, repeaters, wiring concentrators, hubs, switches, and infrared, microwave, and other radio-band transmitters and receivers.

Network Adapters

A network adapter is the hardware installed in computers that enables them to communicate on a network. Network adapters are manufactured in a variety of forms. The most common form is the printed circuit board, which is designed to be installed directly into a standard expansion slot inside a PC. Many manufacturers of desktop workstation motherboards include network adapters as part of the motherboard. Other network adapters are designed for mobile computing: they are small and lightweight and can be connected to portable (laptop and notebook) computers so that the computer and network adapter can be easily transported from network to network.

Network adapters are manufactured for connection to virtually any type of guided medium, including twisted-pair wire, coaxial cable, and fiber-optic cable. They are also manufactured for connection to devices that transmit and receive visible light, infrared light, and radio microwaves.

The hardware used to make connections between network adapters and different transmission media depends on the type of medium used. Figure 6 illustrates a snap-in RJ-45 connector that is ordinarily used for a 10Mbps Ethernet connection.

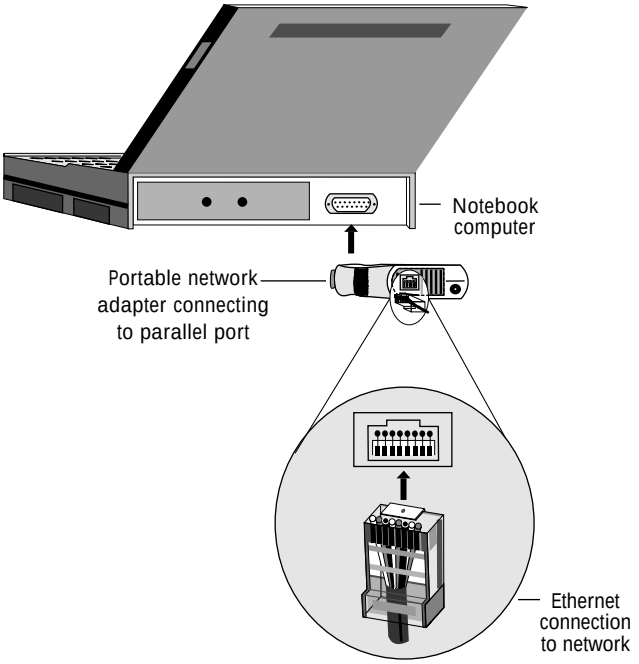


Figure 6

An RJ-45 connector links the adapter to the transmission media.

Repeaters

Repeaters are used to increase the distance over which a network signal can be propagated.

As a signal travels through a transmission medium, it encounters resistance and gradually becomes weak and distorted. The technical term for this signal weakening is "attenuation." All signals attenuate, and at some point they become too weak and distorted to be received reliably. Repeaters are used to overcome this problem.

A simple, dedicated repeater is a device that receives the network signal and retransmits it at the original transmission strength. Repeaters are placed between transmitting and receiving devices on the transmission medium at a point at which the signal is still strong enough to be retransmitted.

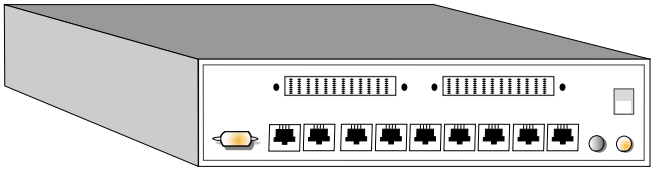
In today's networks, dedicated repeaters are seldom used. Repeaters are "dumb" devices, meaning that they do not have the capability to analyze what they're repeating. They therefore will repeat all signals, including those that should not be repeated, which increases network traffic. Repeating capabilities are now built into other, more complex networking devices that can analyze and filter signals. For example, virtually all modern network adapters, hubs, and switches incorporate repeating capabilities.

Wiring Concentrators, Hubs, and Switches

Wiring concentrators, hubs, and switches provide a common physical connection point for computing devices. (We limit this discussion to devices used for making physical connections. The term "concentrator" can mean something different in a mainframe or minicomputer environment.) Most hubs and all wiring concentrators and switches have built-in signal repeating capability to perform signal repair and retransmission. (These devices also perform other functions.)

In most cases, hubs, wiring concentrators, and switches are proprietary, standalone hardware. There are a number of companies that manufacture such equipment. Occasionally, hub technology consists of hub cards and software that work together in a standard computer.

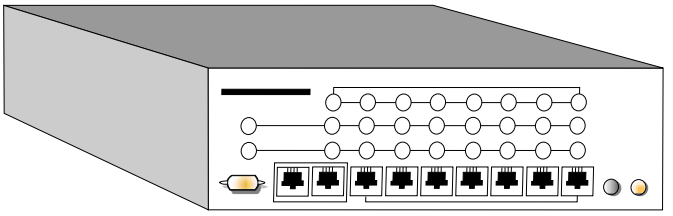
Figure 7 shows two common hardware-based connection devices: a token-ring switch and an Ethernet 10Base-T concentrator.



Token-ring switch

Figure 7

Token-ring switch and Ethernet 10Base-T concentrator



10Base-T concentrator

Modems

Modems provide the means to transmit digital computer data over analog transmission media, such as ordinary, voice-grade telephone lines. The transmitting modem converts the encoded data signal to an audible signal and transmits it. A modem connected at the other end of the line receives the audible signal and converts it back into a digital signal for the receiving computer. Modems are commonly used for inexpensive, intermittent communications between a network and geographically isolated computers.

The word “modem” is derived from “MOdulate and DEModulate”—modems convert digital (computer) signals to analog (audio) signals and vice versa by modulating and demodulating the frequency. However, analog signals consist of a sound wave with *three* states that can be altered: amplitude, frequency, and phase. Low-speed modems modulate only frequency, but faster modems modulate two or three states at the same time, usually frequency and phase. Faster modems also use full-duplex communication—they utilize both incoming and outgoing telephone lines to transmit data—which further increases their speed.

Microwave Transmitters

Microwave transmitters and receivers, especially satellite systems, are commonly used to transmit network signals over great distances. A microwave transmitter uses the atmosphere or outer space as the transmission medium to send the signal to a microwave receiver. The microwave receiver then either relays the signal to another microwave transmitter or translates the signal to some other form, such as digital impulses, and relays it on another suitable medium to its destination. Figure 8 shows a satellite microwave link.

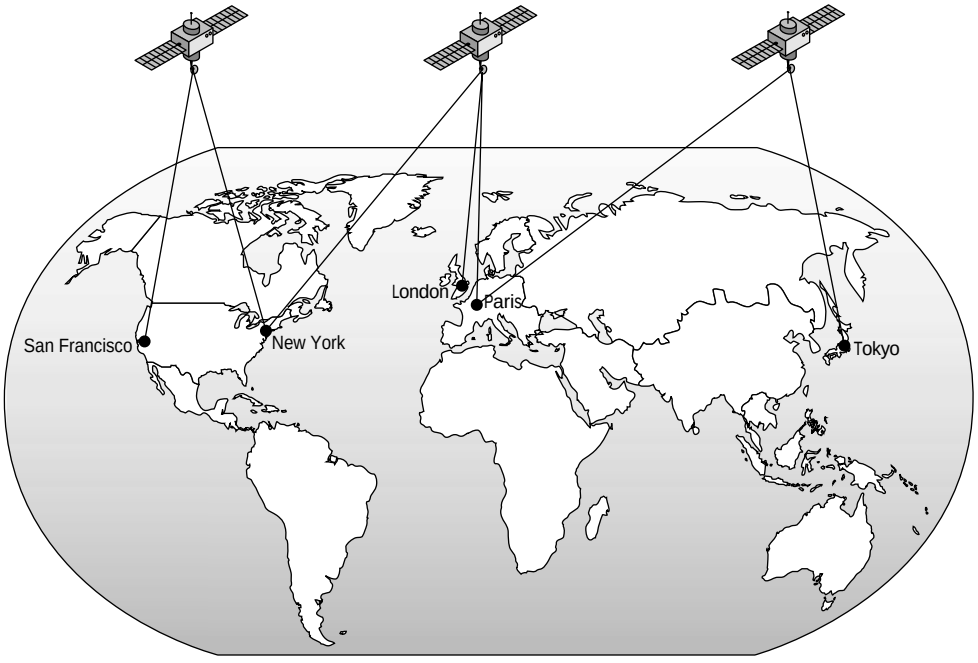


Figure 8

Satellite microwave link

Originally, this technology was used almost exclusively for satellite and long-range communication. Recently, however, there have been developments in cellular technology that allow you complete wireless access to networks, intranets, and the Internet. IEEE 802.11 defines a MAC and physical access control for wireless connection to networks.

Infrared and Laser Transmitters

Infrared and laser transmitters are similar to microwave systems: they use the atmosphere and outer space as transmission media. However, because they transmit light waves rather than radio waves, they require a line-of-sight transmission path.

Infrared and laser transmissions are useful for signaling across short distances where it is impractical to lay cable—for instance, when networks are at sites a few miles apart. Because infrared and laser signals are in the light spectrum, rain, fog, and other environmental factors can cause transmission problems.

Cellular Transmitters

Cellular transmissions are radio transmissions and therefore have the advantage of being able to penetrate solid objects. The cellular base station at the center of each cell consists of low-power transmitters, receivers, antennas, and common control computer equipment. The cell tower usually has a triangular array of antennas on top. Unlike conventional radio and television transmitters, whose primary purpose is to cover the largest area possible, cellular transmitters emit signals that do not carry much farther than a few cells. Cellular devices are likewise configured to operate at low power to avoid interfering with other cellular devices in the area.

Wireless LAN Transmitters

Wireless devices interface with LANs at wireless access points (APs). These APs function like hubs and switches in a wired environment, only they propagate signals through radio waves or infrared light instead of wires. An AP consists of a transceiver, usually positioned in a high place such as a tower or near the ceiling, that physically connects to the hard wiring of the LAN. An AP that is connected to the LAN via radio waves is called an extension point (EP). Wireless networking operates under the same principal as cellular phones: each AP or EP covers a cell, and users are handed off from one cell to the next. Therefore, a user with a handheld device can connect to the network in one room and walk to another part of the building or campus and still maintain connectivity.

Other kinds of wireless transmitters reside in wireless devices and interface directly with similar devices, creating an ad-hoc, peer-to-peer network when they are near one another. These transmitters also operate at very low power to avoid unwanted interference.

Currently, technology is being developed to use the human body as a “wet-wire” transmitter. The personal area network (PAN) takes advantage of the conductive powers of living tissue to transmit signals. The PAN device, which can be worn on a belt, in a pocket, or as a watch, transmits extremely low-power signals (less than 1 MHz) through the body. With a handshake, users could exchange business cards or other information with little fear of eavesdropping from remote users. The PAN specification encompasses all seven layers of the OSI model, meaning that it can address application and file transfer as well. (Note: The term PAN is also used to describe ad hoc, peer-to-peer networks.)

The Network Operating System

Now that you have read about data transmission, the OSI model, and the network hardware involved in network communication, you can begin to understand just how complex network communication really is. In order for a network to communicate successfully, all the separate functions of the individual components discussed in the preceding sections must be coordinated. This task is performed by the network operating system (NOS). The NOS is the "brain" of the entire network, acting as the command center and enabling the network hardware and software to function as one cohesive system.

Network operating systems are divided into two categories: peer-to-peer and client-server. Networks based on peer-to-peer NOSs, much like the example we used in the OSI model discussion, involve computers that are basically equal, all with the same networking abilities. On the other hand, networks based on client-server NOSs are comprised of client workstations that access network resources made available through the server. The advantages and disadvantages of each are discussed in the following sections.

Peer-to-Peer Networks

Peer-to-peer networks enable networked computers to function as both servers and workstations. In a wired peer-to-peer network the NOS is installed on every networked computer so that any networked computer can provide resources and services to all other networked computers. For example, each networked computer can allow other computers to access its files and use connected printers while it is in use as a workstation. In a wireless peer-to-peer network, each networked device contains a short-range transceiver that interfaces with the transceivers of nearby devices or with APs. Like their wired counterparts, wireless peer-to-peer networks offer file and resource sharing.

Peer-to-peer NOSs provide many of the same resources and services as do client-server NOSs, and in the appropriate environment can deliver acceptable performance. They are also easy to install and are usually inexpensive.

However, peer-to-peer networks provide fewer services than client-server networks. Also, the services they provide are less robust than those provided by mature, full-featured client-server networks. Moreover, the performance of peer-to-peer networks decreases significantly both with heavy use and as the network grows. Maintenance is also often more difficult. Because there is no method of centralized management, there can be many servers to manage (rather than one centralized server), and many people may have the rights to change the configuration of different server computers. In the case of wireless peer-to-peer networks, however, an AP may be one node in the network, allowing users both to share files directly from their hard drives and to access resources from the servers on the LAN.

Client-Server Networks

In a client-server network the NOS runs on a computer called the network server. The server must be a specific type of computer. For example, the most commonly used client-server version of the NetWare NOS runs on Intel-based computers.

A client-server NOS is responsible for coordinating the use of all resources and services available from the server on which it is running.

The client part of a client-server network is any other network device or process that makes requests to use server resources and services. For example, network users at workstations request the use of services and resources though client software, which runs in the workstation and communicates with the NOS in the server by means of a common protocol.

On a NetWare client-server network, you "log on" to the network server from the workstation. To log on, you provide your user name and password—also known as a login—to the server. If your user name and password are valid, the server authenticates you and allows you access to all network services and resources to which you have been granted rights. As long as you have proper network rights, the client-server NOS provides the services or resources requested by the applications running on your workstation.

"Resources" generally refers to physical devices that an application may need to access: hardware such as hard disks, random access memory (RAM), printers, and modems. The network file system is also a server resource. The NOS manages access to all these server resources.

The NOS also provides many "services," which are tasks performed or offered by a server such as coordinating file access and file sharing (including file and record locking), managing server memory, managing data security, scheduling tasks for processing, coordinating printer access, and managing internetwork communications.

Among the most important functions performed by a client-server NOS are ensuring the reliability of data stored on the server and managing server security.

There are many other functions that can and should be performed by a network operating system. Choosing the right operating system is extremely important. NetWare NOSs are robust systems that provide many capabilities not found in less mature systems. NetWare NOSs also provide a level of performance and reliability that exceeds that found in most other NOSs.

Thin Client-Server Networks

A variation on the client-server network is the server-based network or thin client-server network. This kind of network also consists of servers and clients, but the relationship between client and server is different. Thin clients are similar to terminals connected to mainframes: the bulk of the processing is performed by the server and the client presents the interface. Unlike mainframe terminals, however, thin clients are connected to a network, not directly to the server, which means the client does not have to be physically near the server.

The term "thin client" usually refers to a specialized PC that possesses little computing power and is optimized for network connections. Windows-based terminal (WBT) and network computer (NC) are two terms often used interchangeably with thin client. These machines are usually devoid of floppy drives, expansion slots, and hard disks; consequently, the "box" or central processing unit is much smaller than that of a conventional PC.

The "thin" in thin client refers both to the client's reduced processing capabilities and to the amount of traffic generated between client and server. In a typical thin-client environment, only the keystrokes, mouse movements, and screen updates travel across the connection. (The term "thin" is also used generically to describe any computing process or component that uses minimal resources.)

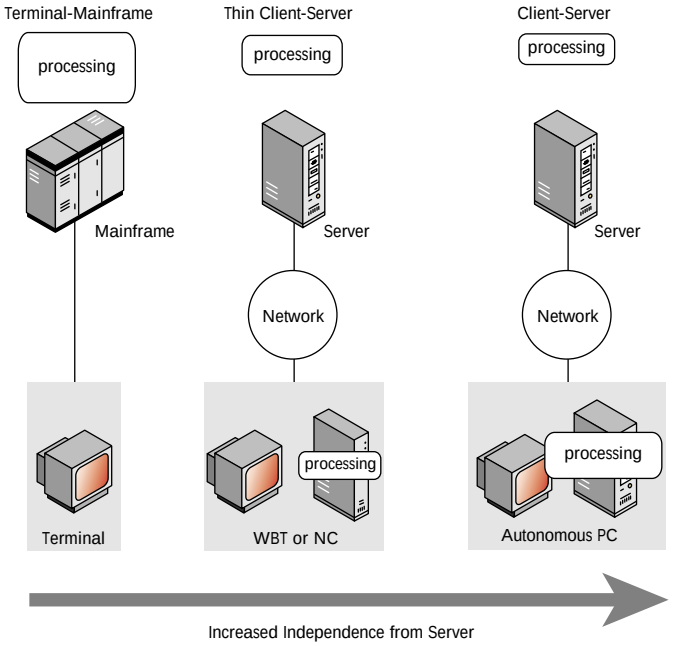


Figure 9

Thin clients range from complete dependence on the server to the autonomous PC, which can both run its own applications and act as a terminal.

Figure 9 shows where clients fall on the "thinness" continuum: mainframe terminals are the thinnest of all, followed by thin clients and conventional PCs. Thin clients are "fatter" than mainframe terminals because they run some software locally—a scaled-back operating system, a browser, and a network client—but they do not store files or run any other applications. PCs, on the other hand, can either be fully autonomous—running all applications and storing all files locally—or they can run browser or terminal-emulation software to function as thin clients.

Unlike mainframe terminals, which show text-only, platform-specific screens, thin clients display the familiar Windows desktop and icons. Furthermore, the Windows display remains consistent even when using non-Windows applications, so you do not have to learn new interfaces in heterogeneous network environments.

Server-based computing usually involves "server farms," which are groups of interconnected servers that function as one. Thin clients link to the farm instead of a particular server. If a single server fails, the other servers in the farm automatically take over the functions of the failed server so that work is not interrupted and data is not lost.

The two primary protocols for thin-client computing are remote display protocol (RDP) and independent computing architecture (ICA). RDP was developed by Microsoft for its Terminal Server and ICA is Citrix technology. Both protocols separate the application logic from the user interface; that is, they pick out the part of the application that interacts with you such as keyboard and mouse input and screens. Only the user interface is sent to the client, leaving the rest of the application to run on the server. This method drastically reduces network traffic and client hardware requirements. ICA clients, for example, can have processors as slow as an Intel 286 and connection speeds as low as 14.4 kilobits per second (Kbps).

Although RDP is the older protocol, ICA has become the de facto standard for server-based computing. ICA presents some distinct advantages over RDP, not the least of which is ICA's platform independence. ICA transmits the user interface over all standard networking protocols—TCP/IP, IPX, SPX, PPP, NetBEUI, and NetBIOS—whereas RDP supports only TCP/IP. ICA also supports all standard clients from Windows to UNIX to Macintosh, but RDP can be used only with Windows 3.11 and later. Furthermore, RDP is a streaming protocol that continuously uses bandwidth while the client is connected, whereas ICA sends packets over the network only when the mouse or keyboard is in use. As a result, most network administrators run ICA on top of RDP to obtain the best functionality.

Server-based computing is best used in environments where only a few applications are needed or when many people will be using the same machine, such as in shift work. For example, if you use only a spreadsheet, a word processor, and e-mail, a thin client may be an ideal solution. Likewise, if the applications rely on databases and directories that are already server-based, such as with airline reservations or patient charts, thin-client computing might be a good choice. Networks with many different platforms can also benefit from server-based computing: you can directly access UNIX, Macintosh, mainframe, or other non-Windows applications via ICA without the mediation of cumbersome translation applications. If, however, you need to use high-end applications such as desktop publishing, graphics, or computer-aided design, the conventional PC with its local computing power provides the only viable option.

Thin-client computing has several other advantages. Because of their simplicity, thin clients are easier for an IT staff to maintain: users cannot tamper with the settings or introduce flawed or virus-infected software into the system. The server-centric model also allows upgrades to be performed at the server level instead of the client level, which is much less time consuming and costly than updating individual PCs. Thin clients typically do not become obsolete as quickly as their fatter counterparts—the servers will, but they are fewer in number and therefore easier to upgrade or replace. Furthermore, thin clients are less likely to be stolen: because they cannot function without a server, they are useless in a home environment.

Disadvantages of thin clients include reduced computing power, which makes them practical only in limited circumstances, and absolute reliance on the network. With conventional PCs users can run applications locally, so when the network goes down, they do not necessarily experience work stoppage. On the other hand, the slightest power outage can cripple a thin-client network for long time: after power is restored, all the clients request the initial kernel from the server at the same time. Also, it is difficult if not impossible to customize a thin client. If you need to install a scanner or other peripheral device, a thin client cannot accommodate it (printers are supported). Furthermore, you cannot customize the look and feel of your desktop, which for some may be disheartening or frustrating.

Nevertheless, thin-client computing has its place, albeit an ironic one: whereas the PC represented progress beyond the terminal/mainframe paradigm, thin clients represent a return to it (though with considerably better technology). Analysts foresee thin-client computing occupying a significant niche among mobile users and application service providers (ASPs). In the near future, when applications are made available over the Internet, thin-client computing will in some cases supplant the autonomous PC.

The term “network topology” refers to the layout of a network. Due to the specific nature of computer network technology, networks must be arranged in a particular way in order to work properly. These arrangements are based on the network hardware’s capabilities and the characteristics of the various modes of data transfer. Because of these factors, network topologies are further subdivided into two categories: physical topologies and logical topologies.

Physical Topologies

The physical topology of a LAN refers to the actual physical organization of the computers on the network and the subsequent guided transmission media connections. Physical topologies vary depending on cost and functionality. We will discuss the three most common physical topologies, including their advantages and disadvantages.

Physical Bus

The simplest form of a physical bus topology consists of a trunk (main) cable with only two end points. When the trunk cable is installed, it is run from area to area and device to device—close enough to each device so that all devices can be connected to it with short drop cables and T-connectors. The principal advantage of this topology is cost: no hubs are required, and shorter lengths of cable can be used. It is also easy to expand. This simple “one wire, two ends” physical bus topology is illustrated in Figure 10.

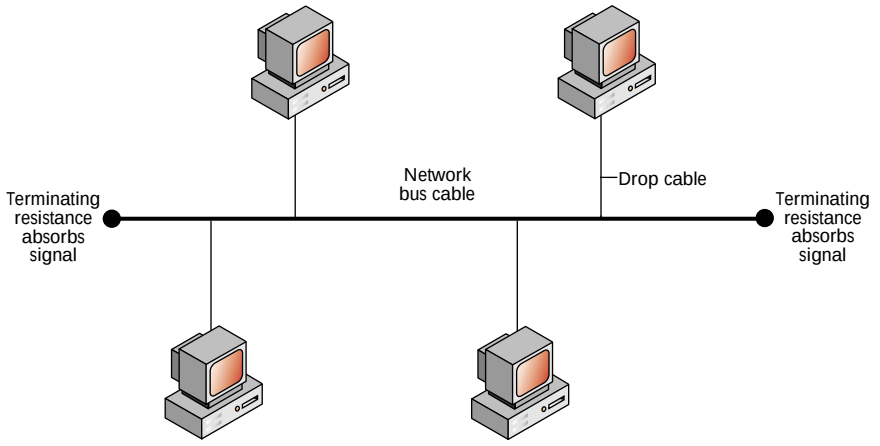


Figure 10

Physical bus topology

Distributed Bus

A more complex form of the physical bus topology is the distributed bus. In the distributed bus, the trunk cable starts at what is called a “root” or “head end,” and branches at various points along the way. Unlike the simple bus topology described above, this variation uses a trunk cable with more than two end points. Where the trunk cable branches, the division is made by means of a simple connector. This topology is susceptible to bottlenecking and single-point failure. The distributed bus topology is illustrated in Figure 11.

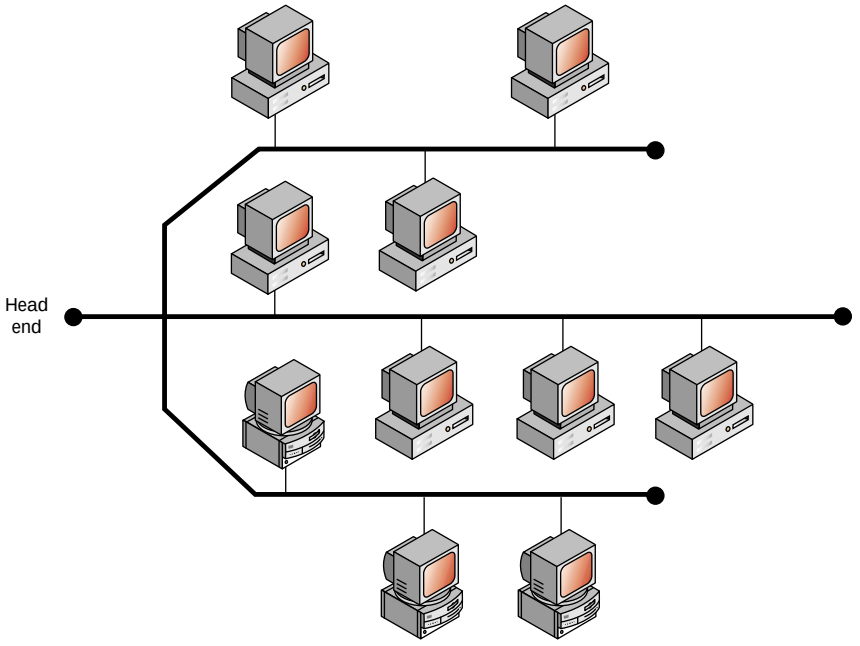


Figure 11

Distributed bus topology

Physical Star

The simplest form of the physical star topology consists of multiple cables—one for each network device—attached to a single, central connection device. 10Base-T Ethernet networks, for example, are based on a physical star topology: each network device is attached to a 10Base-T hub by means of twisted-pair cable.

In even a simple physical star topology, the actual layout of the transmission media need not form a recognizable star pattern; the only required physical characteristic is that each network device be connected by its own cable to the central connection point. Like the distributed bus topology, this topology is vulnerable to single-point failure and bottlenecking.

The simplest form of the physical star topology is illustrated in Figure 12.

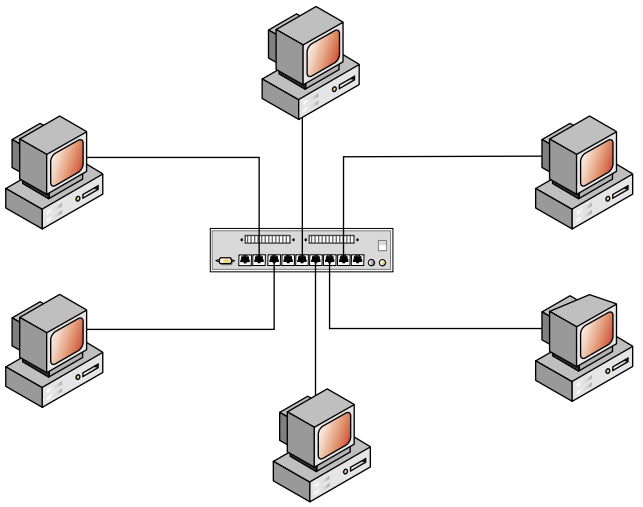


Figure 12

Physical star topology

The distributed star topology, illustrated in Figure 13, is a more complex form of the physical star topology, with multiple central connection points connected to form a string of stars.

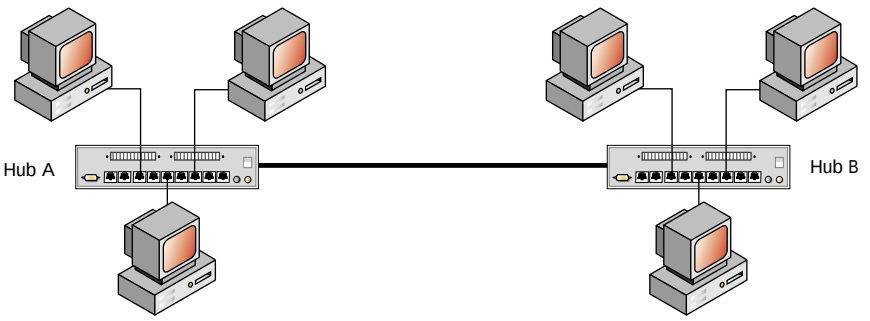


Figure 13

Distributed star topology

Physical Star-Wired Ring

In the star-wired ring physical topology, individual devices are connected to a central hub, just as they are in a star or distributed star network. However, within each hub the physical connections form a ring. Where multiple hubs are used, the ring in each hub is opened, leaving two ends. Each open end is connected to an open end of some other hub (each to a different hub), so that the entire network cable forms one physical ring. This physical topology, which is used in IBM's Token-Ring network, is illustrated in Figure 14.

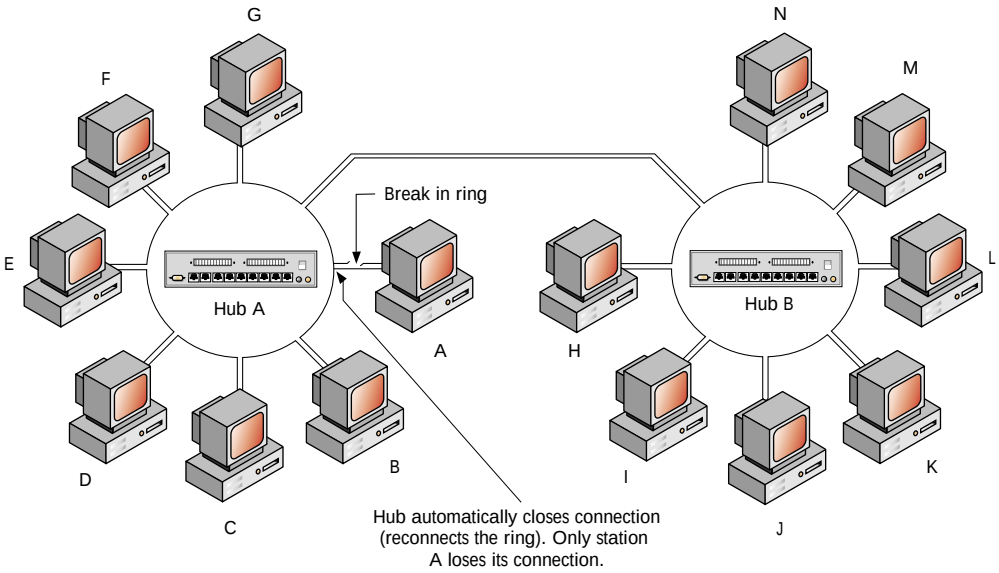


Figure 14

Physical star-wired ring topology

In the star-wired ring physical topology, the hubs are "intelligent." If the physical ring is somehow broken, each hub is able to close the physical circuit at any point in its internal ring, so that the ring is restored. Refer to details shown in Figure 14, Hub A, to see how this works.

Currently, the star topology and its derivatives are preferred by most network designers and installers because these topologies make it simple to add network devices anywhere on the network. In most cases, you can simply install one new cable between the central connection point and the desired location of the new network device without moving or adding to a trunk cable or making the network unavailable for use by other stations. However, the star topology and its derivatives are also susceptible to bottlenecking and single-point failure; the latter is often remedied by providing a redundant backup of the hub node.

Tree Topology

Also called a "hierarchical" or "star of stars" topology, tree topology is a combination of bus and star topologies. Nodes are connected in groups of star-configured workstations that branch out from a single "root," as shown in Figure 15. The root node usually controls the network and sometimes network traffic flow. This topology is easy to extend: when new users need to be added, it is simply a matter of adding a new hub. It also is easy to control because the root provides centralized management and monitoring. The principal disadvantage is obvious: when the entire network depends on one node, failure of that node will bring the whole network down. Also, the tree topology is difficult to configure, wire, and maintain, especially in extensive networks.

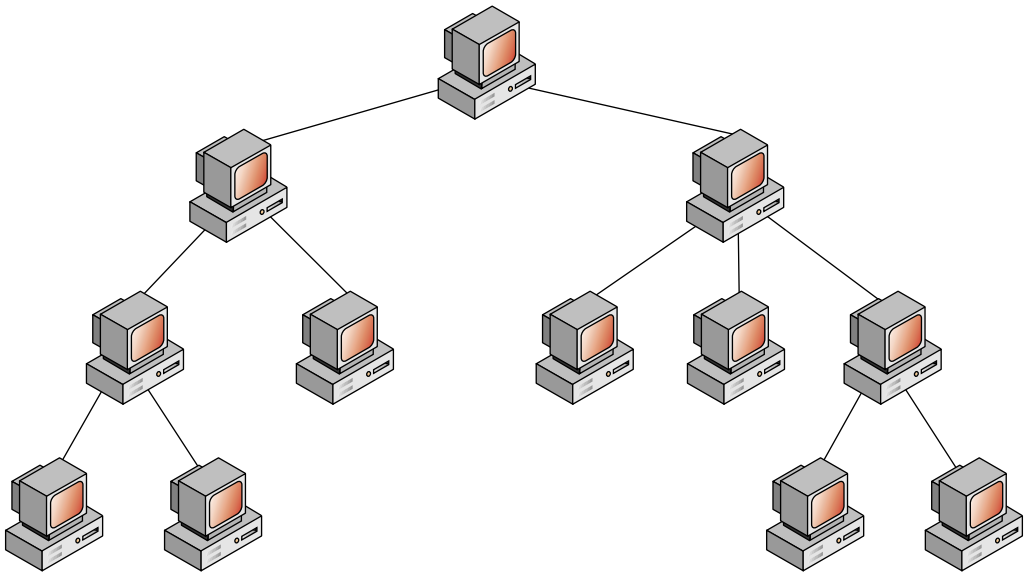


Figure 15

The tree topology is centrally controlled, making it easy to manage but highly vulnerable to single-point failure.

Mesh Topology

A topology gaining popularity in recent years is mesh topology. In a full mesh topology, each node is physically connected to every other node. Partial mesh topology uses fewer connections, and though less expensive is also less fault-tolerant. In a hybrid mesh the mesh is complete in some places but partial in others. Full mesh is generally utilized as a backbone where there are few nodes but a great need for fault tolerance, such as the backbone of a telecommunications company or ISP. Partial and hybrid meshes are usually found in peripheral networks connected to a full-mesh backbone.

The primary advantage of this topology is that it is highly fault tolerant: when one node fails, traffic can easily be diverted to other nodes. It is also not especially vulnerable to bottlenecks. On the other hand, as Figure 16 shows, full mesh topology can require inordinate amounts of cabling if there are more than just a few nodes. A full mesh is also complex and difficult to set up. In a partial or hybrid mesh there is a lack of symmetry—some nodes have more connections than others—which can cause problems with load balancing and traffic.

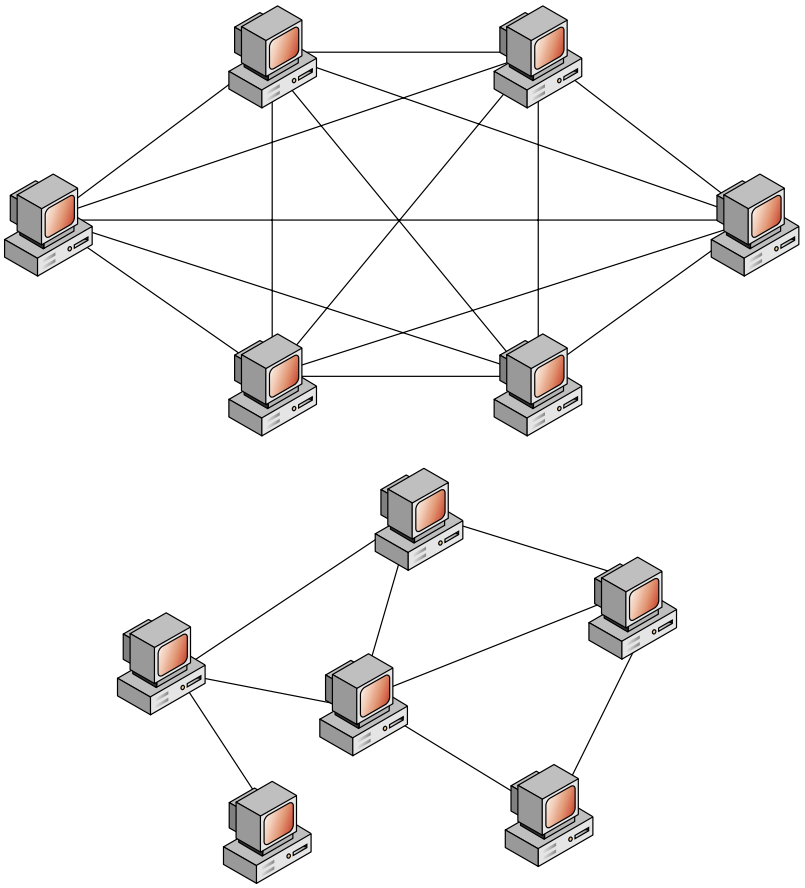


Figure 16

The full mesh, on the top, is both highly complex and highly fault tolerant. The partial mesh sacrifices some fault tolerance in favor of increased simplicity.

Wireless Topologies

Because the medium through which the signals are propagated (radio frequencies) has different properties than wires, wireless topologies differ greatly from wired topologies. The principles used in creating wireless networking solutions are based on the technology currently in use with cellular telephone systems.

Cellular technologies are often described in terms of their "generation": first, second, or third. The first generation is the analog cellular system, second-generation wireless is digital, and the third generation, which has yet to be developed, is often called UMTS: Universal Mobile Telecommunications System. This system is designed to provide digital, packet-switched, high-bandwidth, always-on service for everything from voice to video to data transfer. Once UMTS is implemented, it is hoped that it will be the only standard to which all cellular and wireless devices are built, thereby creating a universal wireless standard.

Cellular Technology

Literally at the center of any cellular technology is the cellular transceiver, an omnidirectional antenna whose range projects a circular "footprint." This footprint is the "cell" that gives cellular technology its name.

Cellular providers are allotted a set of frequencies within a specified area called a metropolitan statistical area (MSA) or a rural statistical area (RSA) (usually, two providers obtain rights to the same MSA or RSA). They divide their statistical areas into cells and place an antenna at the center of each. Ideally, the antennas are located such that their entire allotted statistical area is covered by cells. In this way, a cellular user can move throughout the provider's statistical area and always be in range of an antenna.

As the cellular user moves from one cell to another, the user's signal is transferred from one antenna to another in a process called "handing off." The handing-off process is governed by a mobile telephone switching office (MTSO)—the hub through which all cellular calls are routed. Figure 17 shows how handing off is accomplished. Cellular antennas continuously send out a control signal to whichever mobile devices are in their cells. When a cellular device initiates a transmission in Cell A, for example, the device receives the control signal and sends back another signal in reply. This signal is received by several nearby antennas, but to the antenna closest to it (in Cell A) the signal will be strongest. The MTSO will therefore route the cellular transmission through the antenna in Cell A; other antennas will ignore the signal. As the user moves from Cell A to Cell B, the MTSO will detect that the signal in Cell A is becoming weaker while in Cell B it is becoming stronger. When the user moves into Cell B, the MTSO assigns the device a new frequency and routes the transmission solely through the antenna in Cell B.

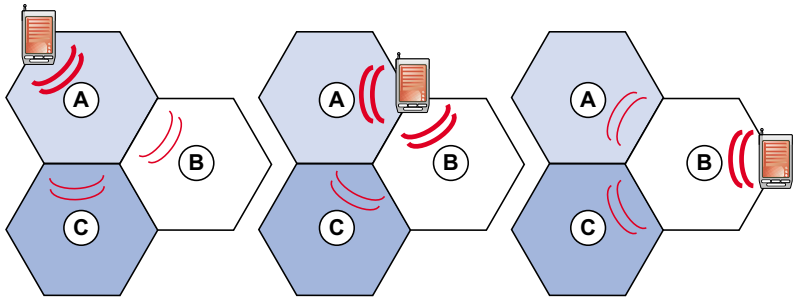


Figure 17

Hand-off from one cell to the next

Frequency Reuse

In most wireless technologies, only one party can transmit a signal at a given frequency within a defined geographical location. With regard to cellular service, however, a person on a mobile call monopolizes their allocated frequency only within their current cell; someone in a cell across town can use that frequency at the same time. The concept of multiple users operating at the same frequency in the same geographic area is known as "frequency reuse."

For frequency reuse to be effective, every cell phone in the area must put out only enough power to reach the antenna of the cell that they are in. Too much power, and the signal would be picked up by unintended antennas in other cells; this could interfere with conversations in those cells being transmitted at the same frequency.

Cells using the same frequency, however, cannot be adjacent to each other. This is because a cellular caller on the border between two cells would put out just enough power to reach both the intended antenna in one cell and the antenna in the other. This would interfere with any conversations taking place at the same frequency in the unintended cell. Of course, in order for the cellular provider to allow more callers to use the same frequency to carry on separate conversations, the provider will want to place cells using that frequency as close together as possible. The number of cells of separation usually ranges from four to 21.

Air Interfaces

The term "air interface" describes the way the signal is modulated between the wireless device and the base station. Air interfaces generally use modulation schemes designed to increase the information throughput of the wireless system. Three principal air interfaces are Frequency Division Multiple Access (FDMA), Time Division Multiple Access (TDMA), and Code Division Multiple Access (CDMA).

FDMA is the oldest of these schemes and the least efficient. With FDMA, only one transmission is propagated over each channel at a time. The channel is dedicated to that one transmission regardless of whether data is being transmitted, and the channel is not available for another user until the device using it terminates the transmission. Figure 18 shows how FDMA allocates one channel for each user. This scheme is used by Advanced Mobile Phone System (AMPS), the analog cellular system in the United States, and Total Access Communication System (TACS), the analog cellular system in Europe. FDMA is for analog signals only and therefore can handle only voice transmissions, not data or fax.

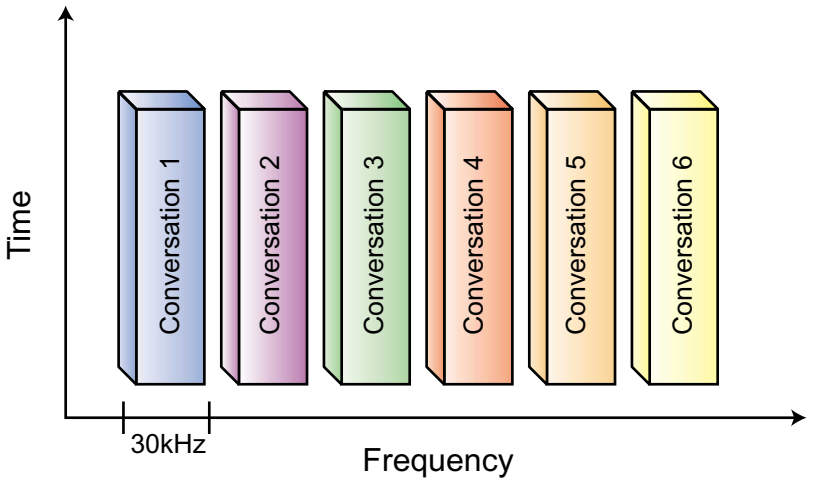


Figure 18

FDMA channel allocation

TDMA is often called digital FDMA. It increases the number of transmissions per channel by taking advantage of the properties of digital technology. With digital transmissions, data is sent in discrete packets rather than continuous analog waves. TDMA assigns these packets a time slot on a frequency and alternates between transmissions so that more transmissions can be sent per channel. To send three telephone conversations on one channel, for example, TDMA sends first a packet from the first conversation, then a packet from the second, then from the third, then from the first, etc., as shown in Figure 19. By separating the conversations by time, TDMA can "simultaneously" transmit all three conversations over the same channel. The IS-54 and IS-136 implementations of TDMA allow three users to occupy the same channel at once, and there are implementations that allow six. In the future, TDMA will be able to accommodate up to 40 signals per channel. TDMA is used by GSM, the European digital standard, and PDC, the Japanese digital standard.

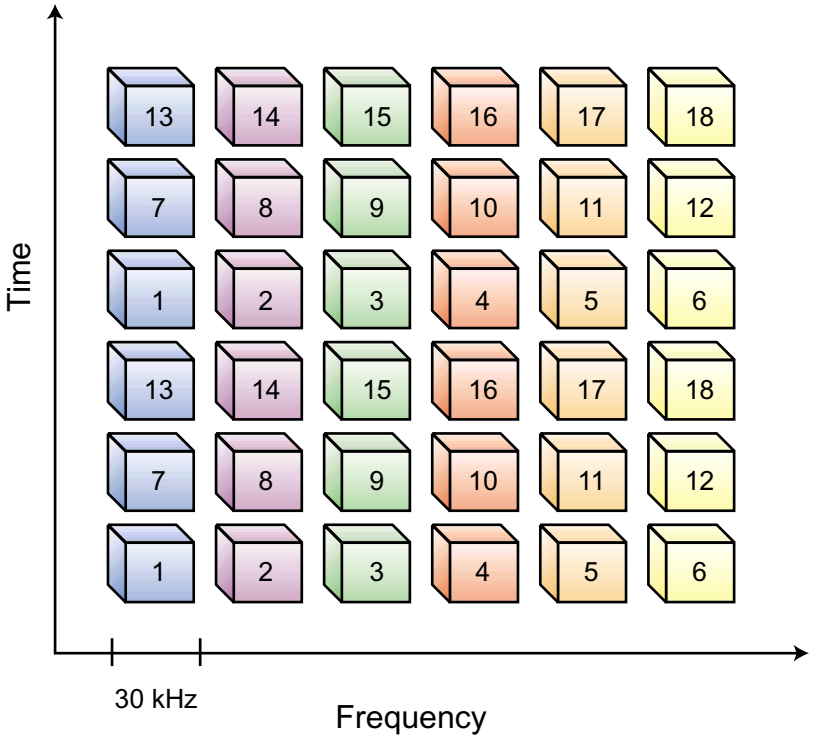


Figure 19

TDMA divides each channel into multiple time slots.

However, even though TDMA packs more transmissions into each channel, the time slots may go empty when data is not being sent, such as during the pauses in a conversation. An advanced implementation of TDMA, called Extended TDMA (ETDMA), assigns data to the time slots dynamically so that slots are always filled.

CDMA is DSSS technology. It attaches "pseudo-random code sequences" to each packet sent. The code is known only to the sender and the receiver. The signal is also spread across a range of frequencies, which allows many users to transmit across the same range at once, as shown in Figure 20. This spread-spectrum technology allows up to 20 times more transmissions per cell than with FDMA. CDMA is the predominant technology used by PCS networks in the United States.

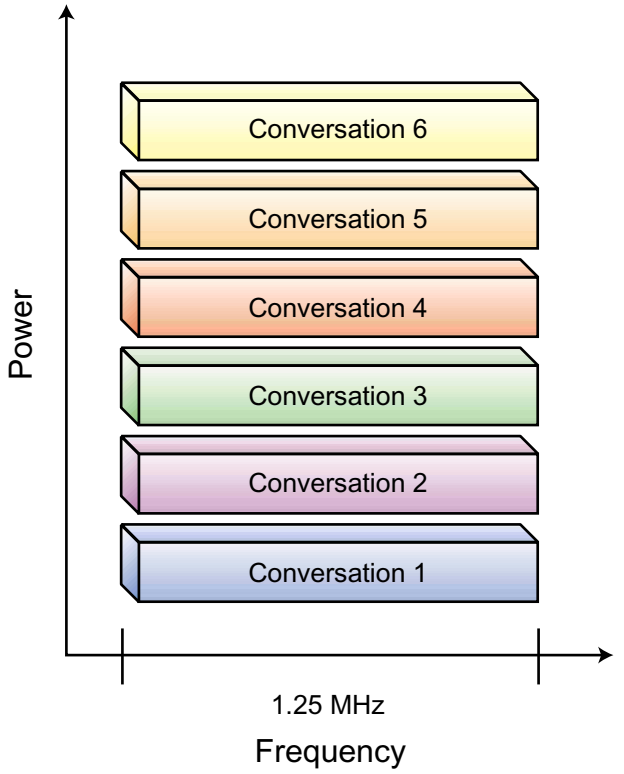


Figure 20

CDMA spreads each transmission over a range of frequencies.

Wireless LANs

Wireless LANs (WLANs) are variations on wired LANs rather than new topologies, but the hardware used to create them is different. The most obvious hardware difference is the absence of wiring, which provides several advantages, not the least of which are lower cost and greater flexibility and mobility.

A wireless peer-to-peer network consists of two or more wireless-enabled devices such as PCs with WLAN cards or handheld devices that are in close proximity to one another. Figure 21 shows a wireless peer-to-peer network configuration.

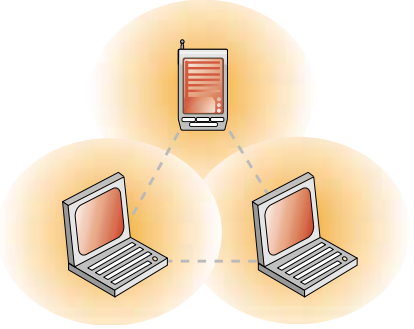


Figure 21

The wireless peer-to-peer network consists of several wireless devices

A wireless topology that more closely mimics a wired LAN involves one or more APs or EPs. Figure 22 shows how an AP, wired to the LAN backbone, functions much like a hub or a switch.

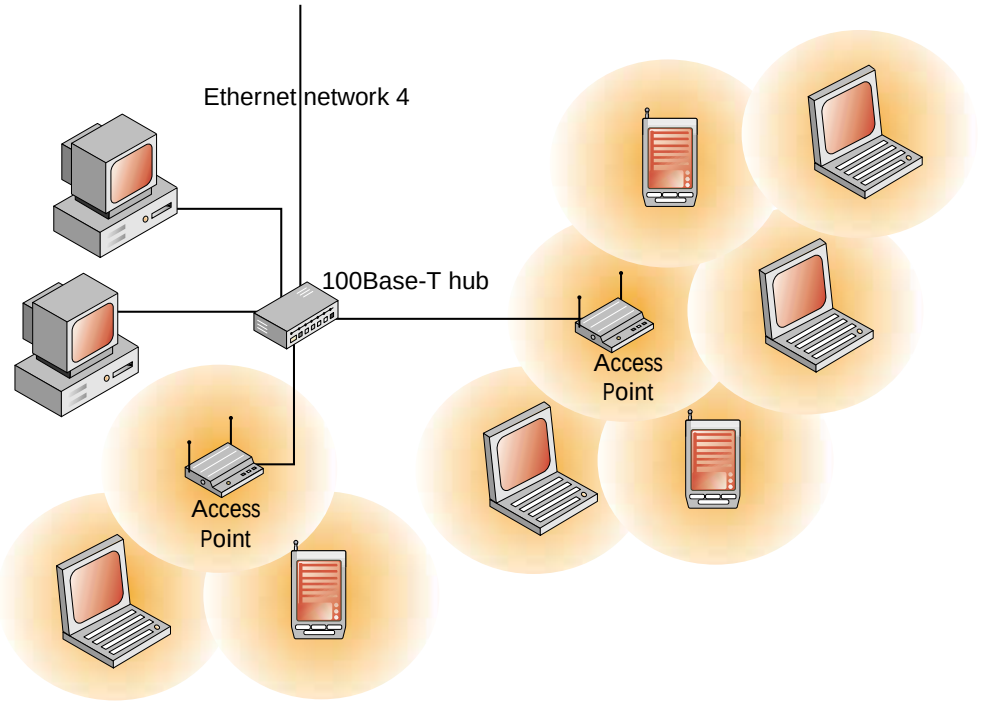


Figure 22

Client and access point

Logical Topologies

The logical topology is the schema of the actual path the data follows within the physical topology. It differs from the physical topology in that not only does it show the location of network components, it also shows the path the data follows through these components as well as the direction of travel. It is used to enable network devices to transmit and receive data across the transmission media without interfering with each other.

Because the logical topology is associated with the path and direction of data, it is closely linked with the MAC methods in the media access layer of the OSI model. Specific MAC methods are required for each of the logical topologies in order to monitor and control data flow. These methods will be discussed in conjunction with the appropriate logical topology.

There are three basic logical topologies: logical bus, logical ring, and logical star (switching). Each of these topologies offers distinct advantages in specific situations. As you study the figures representing these topologies, remember that they illustrate a logical (electronic) rather than a physical connection scheme.

Logical Bus

In the logical bus topology, transmissions (called frames) are broadcast simultaneously in every direction to every point on the transmission media. Every network station checks each frame to determine for whom it is intended. When the signal reaches any end point on the transmission media, it is absorbed (removed from the media) by an appropriate device. Removing the signal prevents it from being reflected back along the transmission media and interfering with subsequent transmissions.

On a logical bus network the transmission media are shared. To prevent transmission interference, only one station may transmit at a time. Thus, there must be a method for determining when each station is allowed to use the media. The methods used to control how data is sent on the network are the MAC methods that we discussed briefly in the "Data Transmission" section.

The MAC method most commonly used for a logical bus network is CSMA/CD. This MAC method is similar to the access scheme used on a telephone party line. When any station wants to send a transmission, it "listens" (carrier sense) to determine if another station is currently transmitting on the media. If another station is transmitting, the station that wants to transmit waits. When the media become free, the waiting station transmits. If two or more stations determine that the media are free and transmit simultaneously, there is a data "collision." All transmitting stations detect the collision and transmit a brief signal to inform all other stations that a collision has occurred. All stations then wait a random amount of time before attempting to retransmit.

A logical bus network may also use token passing for media access control. In this MAC method each network station is assigned a logical position in an ordered sequence with the last number of the sequence pointing back to the first (the logical order that the stations are assigned need not correspond to any physical order). A control frame called a "token" is used to control which station can use the media: a station can transmit only when in possession of the token. Furthermore, a station can have the token for a limited time only; it then must pass the token to the next station. The token starts at the first station in the predefined logical order. While the first station has the token, it transmits, polls stations, and receives responses until the allotted time expires; or, it passes the token when it no longer needs control of the media. The first station passes the token to the second station in the logical sequence. This sequential token passing continues nonstop while the network is running so that every station gets equal access to the transmission media.

The logical bus transmission scheme is used in combination with both the physical bus and physical star topology. The MAC method can vary from case to case. For example, while thin Ethernet and 10Base-T Ethernet use the logical bus transmission scheme, cable on thin Ethernet networks is laid out as a physical bus, and on 10Base-T networks as a physical star. Thin Ethernet (physical bus) and 10Base-T Ethernet (physical star), however, both use the CSMA/CD MAC method.

Figure 23 shows a thin Ethernet network (physical bus, logical bus), and Figure 24 shows a 10Base-T Ethernet network (physical star, logical bus). In both figures, notice that the network signal (shown by the arrows) emanates from the sending station and travels in all directions to all parts of the transmission media.

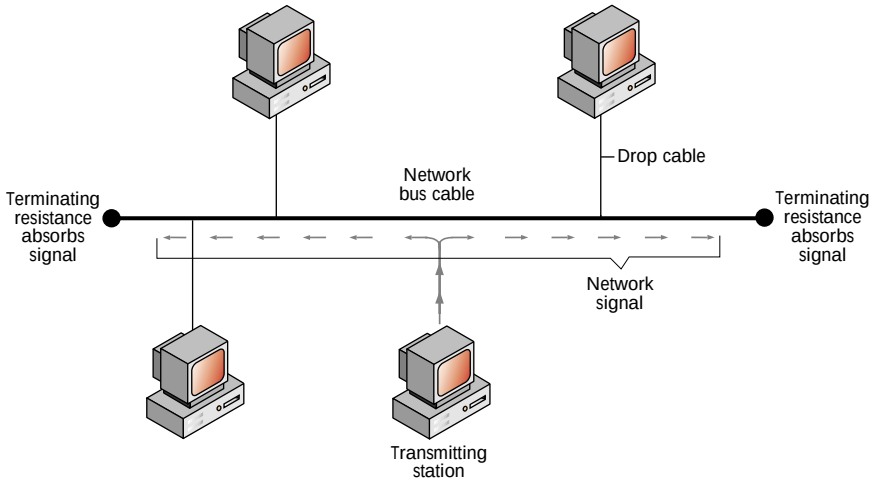


Figure 23

Thin Ethernet network (physical bus, logical bus)

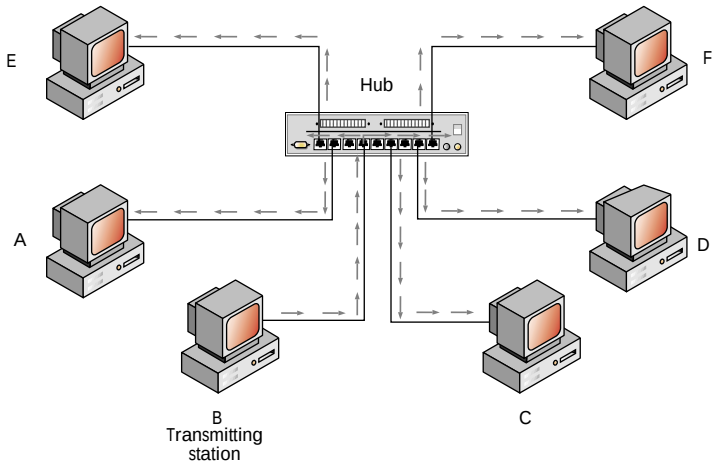


Figure 24

10Base-T Ethernet network (physical star, logical bus)

Logical Ring

In the logical ring topology, frames are transmitted in one direction around a physical ring until they have passed every point on the transmission media. The logical ring must be used in combination with a physical ring topology such as the star-wired ring explained earlier. Each station on the physical ring receives the signal from the previous station and repeats the signal for the next station. When a station transmits data, it gives the data the address of another station on the ring. The data is circulated

around the ring through each station's repeater until it reaches the station to which it is addressed and is copied. The receiving station adds an acknowledgment of receipt to the frame. The frame continues on around the ring until it returns to the station from which it was originally transmitted. This station reads the acknowledgment and removes the signal from the ring. Figure 25 shows how data would flow on a logical ring network with a star-wired ring physical topology.

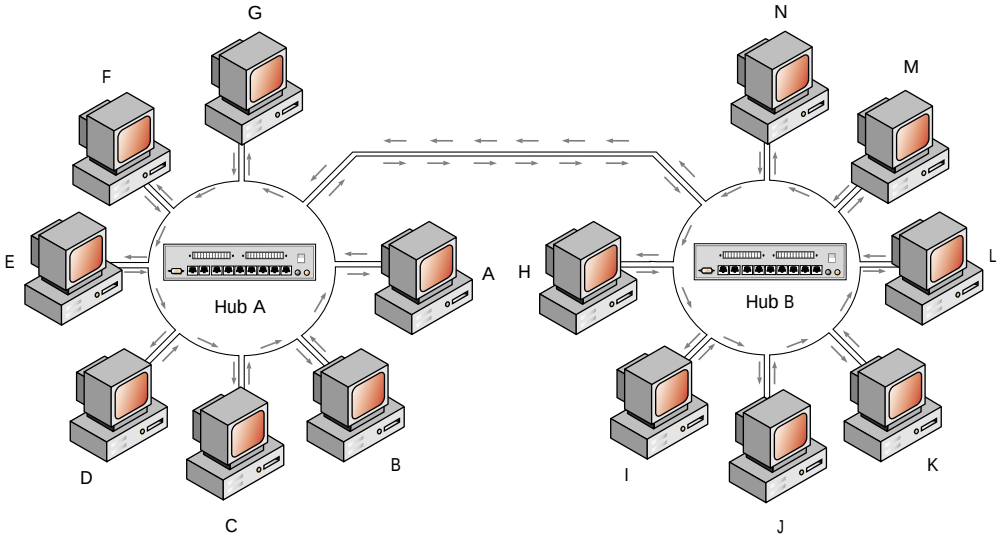


Figure 25

Logical ring topology

Media access control for the logical ring topology is almost always based on a form of token passing. However, stations are not necessarily granted media access in the same order in which they receive frames on the physical ring. IBM's Token-Ring network is a logical ring network based on the star-wired ring physical topology.

Logical Star (Switching)

In the logical star topology, network switches are used to restrict transmissions to a specific part of the transmission medium. Transmission path restriction is the identifying characteristic of a logical star.

In its pure form, switching provides a dedicated line for each end station. When one station transmits a signal to another station on the same switch, the switch transmits the signal only on the two paths connecting the sending and receiving station. Figure 26 shows how data would be transmitted from one station to another if two stations were directly connected to the same switch.

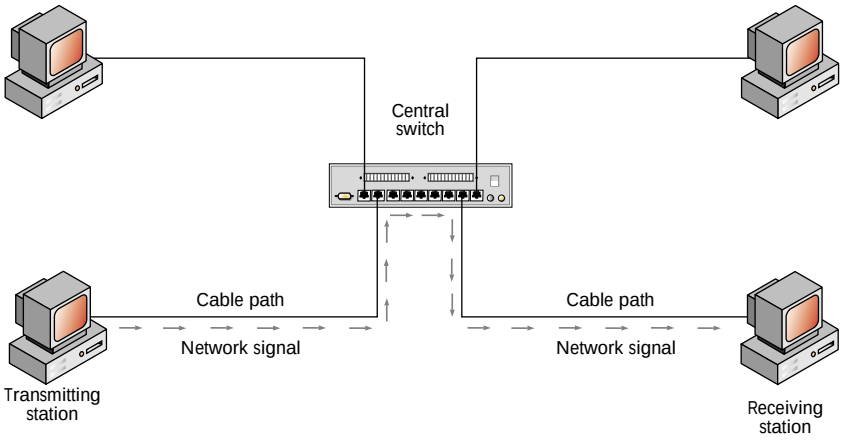


Figure 26

Switching

Most switching technology adds switching capability to existing connection standards, incorporating the logical connection schemes of the existing standards, such as the MAC methods. For example, a 10Base-T Ethernet switch supports the Ethernet CSMA/CD MAC method.

Switches have built-in connection logic and a significant quantity of fast memory. They can simultaneously service all connected stations at full access speed. Thus, when you connect a station directly to a switch, you can increase the total throughput of your network—a significant performance advantage.

Switching illustrates that a logical topology consists of the total of the various aspects of the electronic connection scheme, not just the MAC method. By combining new switching capabilities with existing logical connection schemes, engineers create a new logical topology.

Multiple switches can be connected using one or more physical topologies. Switches can be used not only to connect individual stations, but to connect groups of network stations, known as segments. Thus, in many circumstances switching can be used to improve the performance of your network.

Connecting a Simple Network

Now that we have discussed the hardware pieces that make up a network and considered the difference between physical and logical topology, we will illustrate how to connect some hardware devices to form a simple network. Figure 27 shows some of the hardware items we have discussed, connected to form a basic computer network.

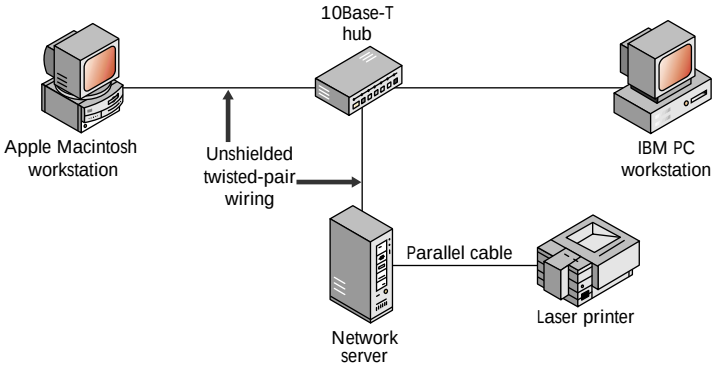


Figure 27

Various networking hardware connected to form a simple network

The network in this figure includes the following components: three computers connected with a 10Base-T hub by means of unshielded twisted-pair wiring, an Ethernet 10Base-T network adapter installed inside each of the nodes, and a laser printer that is connected to one of the nodes.

The node at the bottom center of the illustration is a network server, and it controls the network. The other two nodes are workstations. The workstations use the network under the control of the network server. One workstation is an IBM PC and the other is an Apple Macintosh computer.

The 10Base-T hub serves as a common connection point for the three computers. It also repeats network signals.

The lines between the different components of the network represent the transmission medium: twisted-pair wiring. This 10Base-T network is connected in a physical star, but it is based on a logical bus that uses a contention scheme for the workstations to gain access to the transmission medium.

The printer in this network is connected directly to the server by means of a parallel interface cable, which is a standard connection method. The server accepts print jobs from either workstation and sends the jobs through the cable to the printer. While this is the simplest way to enable both workstations to use the printer, there are other ways to connect printers to a network. You can, for example, attach them to a computer set up as a dedicated print server, or to a computer running special software by which it functions as both workstation and print server. Many high-end printers are now manufactured with an internal network adapter so that they can be attached directly to the transmission medium at any physical point in the network.

Internetworking

As your business grows, it may become necessary to divide a network or connect two separate networks. When a network is split or when two networks with different addresses are connected, an internetwork is created. An internetwork has subnetworks, or network segments, that have different network addresses. Even a modest-sized business often has several subnetworks operating, each serving a specific portion of the organization.

The most common reason for segmenting a network is to enhance network performance. If a network has too many users or devices that need access to resources or services, the transmission media can become so busy that devices have to wait for an unacceptable period of time to transmit. When this happens, you begin to notice delays when you try to save or open files or perform other operations.

When you segment a network, you give each subnetwork its own network address. This results in two separate transmission media segments, which can be used simultaneously. Both segments will have only half the users of the original network. Thus, you double network performance. Moreover, on some networks performance can be more than doubled, because on an overloaded network the resources required to manage transmission collisions use more network "bandwidth" (the amount of data that can be transmitted in a fixed amount of time) than those on a modestly busy network. Networks are also segmented to enhance data security and to minimize the effect of equipment failure on any part of the network.

There are several devices and protocols used to internetwork subnetworks. The following sections discuss these items and how they apply to an internetworked environment.

Internetworking Devices

The devices used to interconnect network segments are divided into three classifications: bridges, routers, and gateways. Each of these devices plays a very specific role in internetworking. Bridges and routers are generally used to connect networks that use similar protocols, while gateways are used to connect networks that use dissimilar protocols.

Bridges and routers are usually separate hardware components that are connected directly to the transmission media at the intersection point of the two separate networks. There are also bridges and routers that are software-based and function as part of a server's NOS or run in conjunction with the NOS. Software-based bridges and routers can also be installed on standard computers to create dedicated, standalone devices.

Gateways, on the other hand, are usually a combination of both hardware and software, and they perform much more advanced functions than either bridges or routers. The following sections explain the differences between these internetworking devices.

Bridges

A bridge operates at the data-link layer (Layer 2) of the OSI model. A bridge acts as an address filter: based on information contained at the MAC level, it relays data between subnetworks.

Simple bridges are used to connect networks that use the same physical-layer protocol and the same MAC and logical link protocols (OSI Layers 1 and 2). Simple bridges are not capable of translating between different protocols.

Other types of bridges, such as translational bridges, can connect networks that use different Layer 1 and MAC-level protocols; they are capable of translating, then relaying frames.

After a physical connection is made (at OSI Layer 1), a bridge receives all frames from each connected subnetwork and checks the network address of each received frame. The network address is contained in the MAC header. When a bridge receives a frame from one subnetwork that is addressed to a workstation on another subnetwork, it passes the frame to the intended subnetwork. Figure 28 provides a simple illustration of how a bridge relays frames between subnetworks.

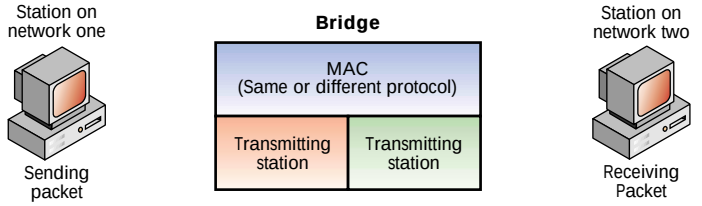


Figure 28

Internetworking through a bridge

A bridge assumes that all communication protocols used above the data-link layer at which it operates (OSI Layers 3 through 7) are the same on both sides of the communication link. If not, translation between unlike protocols at Layers 3 through 7 will need to be performed by something other than the bridge.

Spanning Trees and Source-Route Bridging

There are two terms connected with bridging that are useful to understand: Spanning Tree Protocol and source-route bridging.

Spanning Tree Protocol prevents problems resulting from the interconnection of multiple networks by means of parallel transmission paths. In various bridging circumstances, it is possible to have multiple transmission routes between workstations on different networks. If multiple transmission routes exist, it is also possible to have an endless duplication and expansion of routing errors that will saturate the network with useless transmissions, quickly disabling it. Spanning Tree Protocol is used to specify one, and only one, transmission route. When bridges use this protocol, they send out special frames to each other so that they can be “aware” of the network’s topology; they then disable all redundant pathways.

Source-route bridging is a means of determining the path used to transfer data from one workstation to another. Workstations that use source routing participate in route discovery and specify the route to be used for each transmitted packet. Source-route bridges carry out the routing instructions that are placed into each data packet when the packet is assembled by the sending workstation—hence the name "source routing." Source-route bridging is used on IBM Token-Ring networks.

Although it includes the term "routing," source-routing is a part of bridging technology. Bridging technologies and routing methods can be combined in various ways. For example, there is an IEEE specification for a source-route transparent bridge, a bridging scheme that merges source-route bridging and transparent bridging in one device. When choosing internetworking products, it is important to select those that support multiple bridging methods.

Routers

Routers function at the network layer of the OSI model (one layer above bridges). To communicate with each other, routers must use the same network-layer protocol. The sending and receiving workstations on different networks must either share identical protocols at all OSI layers above Layer 3 or something must perform the protocol translation at these layers.

Like some bridges, routers can allow the transfer of data between networks that use different protocols at OSI Layers 1 and 2 (the physical layer and the data-link layer). Routers can receive, reformat, and retransmit data packets assembled by different Layer 1 and Layer 2 protocols. Different routers are built to manage different protocol sets. Figure 29 illustrates how a router transfers data packets.

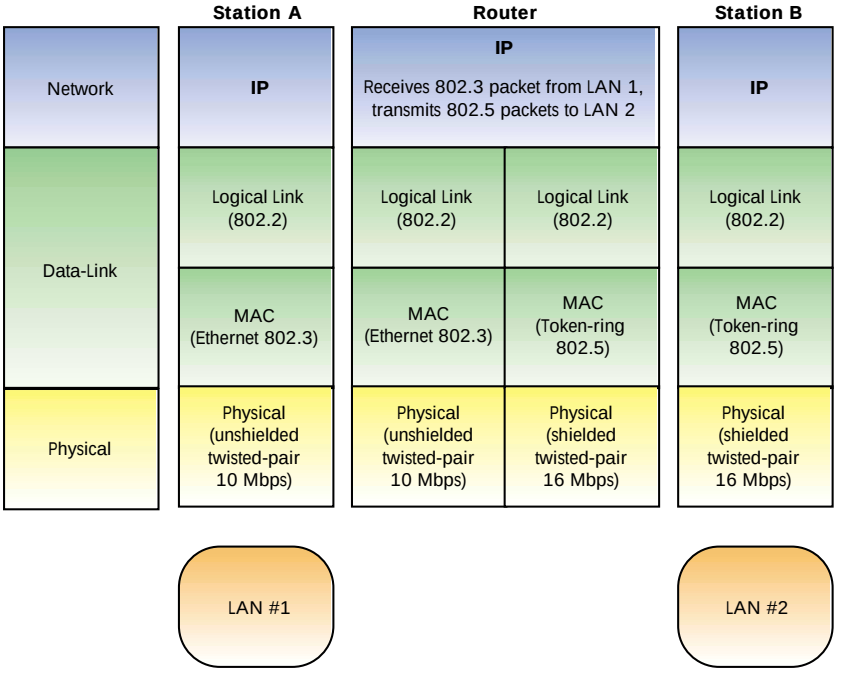


Figure 29

Internetworking through a router

Wireless routers can be used to join two remote LANs or to connect a LAN with the Internet instead of using expensive WAN technology such as a leased line. Wireless routers can have a transmission range of up to 30 miles and a transmission rate of up to 11 Mbps.

Gateways

In contrast to bridges and routers, which function at only one layer of the OSI model, a gateway translates protocols at more than one OSI layer. Therefore, a gateway is used to interconnect computer systems that have different architectures and that therefore use different communication protocols at several OSI layers.

A gateway can connect entirely dissimilar networks or it can connect dissimilar systems on the same network (thus, using a gateway does not necessarily involve internetworking). For example, a gateway might translate protocols at several different OSI layers to allow transparent communications between Internetwork Packet Exchange™ (IPX™)-based systems and systems based on TCP/IP, Systems Network Architecture (SNA), or AppleTalk. Figure 30 illustrates how a gateway is used to translate protocols to enable communications between two heterogeneous systems.

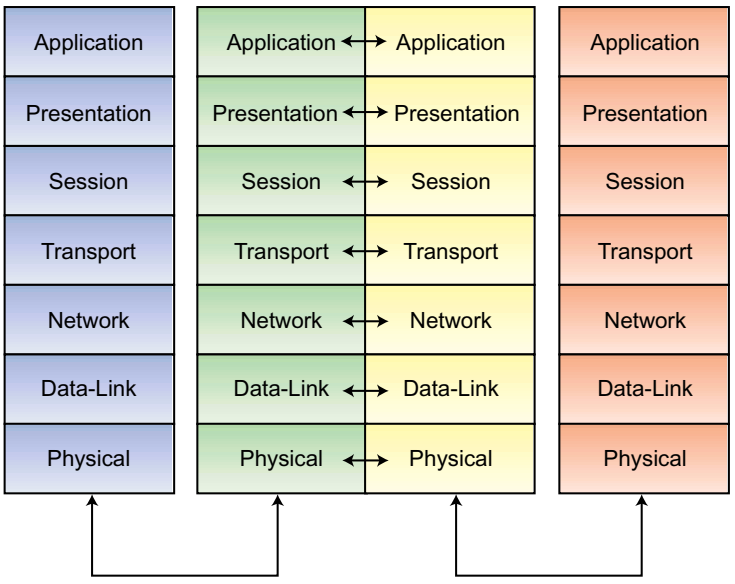


Figure 30

Gateways provide protocol translation between dissimilar systems at more than one OSI layer.

A gateway may consist of hardware or software but is usually a combination of the two. It also may provide translation at all or at only some of the different OSI layers, depending on the types of systems it connects.

Differing systems use different protocols for network communication. The following sections discuss several internetworking protocols that Novell has developed, adopted, or adapted, as well as how they fit into the OSI model.

NetWare Internetworking Protocols

Each of the protocols shown in Figure 31 plays a role, directly or indirectly, in NetWare internetworking. These protocols were developed to facilitate the transfer of data in networked and internetworked environments.

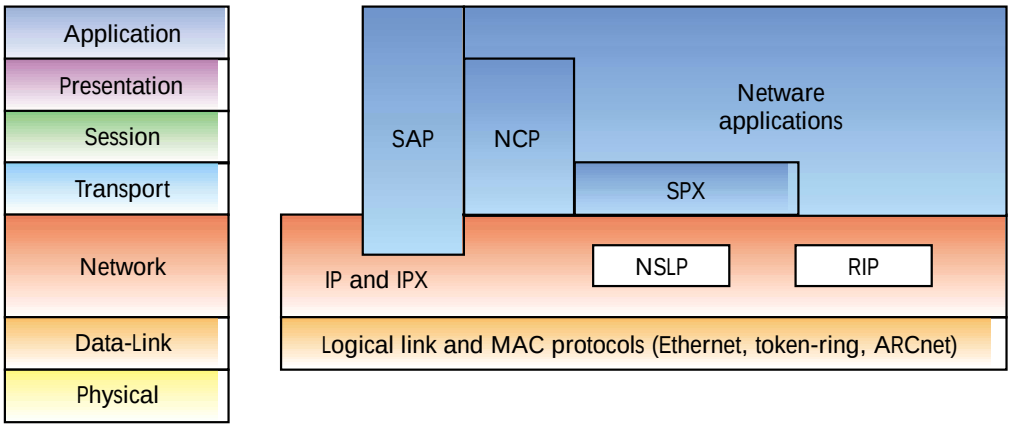


Figure 31

Where NetWare protocols fit in the OSI model

IP and IPX: Network-Layer Protocols

TCP/IP is the Internet's fundamental communications protocol suite. The popularity of the Internet—and of corporate intranets—has therefore made IP the world's dominant networking protocol. As a result, Novell has also adopted IP as its primary network-layer (OSI Layer 3) protocol. However, Novell continues to support their proprietary network-layer protocol, IPX.

In a NetWare environment, internetwork packet routing is accomplished at the network layer. In conjunction with industry-standard MAC protocols, NetWare IP and NetWare IPX provide the NetWare addressing mechanism that delivers communication packets to their destination and routes packets between internetworked computers.

IP and IPX base routing decisions on address fields in packet headers (provided by the MAC protocol) and on information received from other internetworking protocols. For example, IP and IPX use information supplied by Routing Information Protocol (RIP) to forward packets to the destination computer or to the next router. Similarly, NetWare Link Services Protocol™ (NLSP™) is a companion protocol to IPX for the exchange of routing information in a Novell network.

Both IP and IPX also employ SAP, a protocol that enables networked devices such as network servers and routers to exchange information about available network services.

RIP and NLSP: Routing Protocols

NetWare routers use distance-vector and link-state routing protocols to exchange routing information with neighboring routers. In an internetwork using distance-vector routing—the traditional method of router communication—routers periodically receive information about the internetwork's topology from neighboring routers, consolidate this information within their own routing tables, and broadcast packets to other neighboring routers that summarize the information they now have.

IP RIP and IPX RIP are well-known distance-vector routing protocols. Examples of other such protocols include Cisco Internet Gateway Routing Protocol (IGRP), which is part of the IP protocol suite, and Routing Table Maintenance Protocol (RTMP), which is part of the AppleTalk protocol suite. Enhanced IGRP can handle AppleTalk and IPX—in addition to IP—routing information.

Link-state protocols adapt more quickly to network topology changes than do distance-vector protocols. Unlike distance-vector routers, each link-state router builds its own routing map: it does not rely upon secondhand summaries from other routers. Moreover, routing transmissions are made only when the internetwork changes, not at predefined intervals, which reduces network traffic. Thus, link-state protocols are better than distance-vector protocols at managing internetworking on large, complex internetworks.

NLSP is a link-state routing protocol. Examples of other link-state protocols include Open Shortest Path First (OSPF), which is part of the TCP/IP suite, and Intermediate System-to-Intermediate System (IS-IS), a router-to-router protocol that is part of the OSI suite.

Various distance-vector and link-state routing protocols can coexist on the same NetWare internetwork. Furthermore, individual routers can be configured to accept or reject individual protocols.

NCP: NetWare Core Protocol

NCP is a set of service protocols that a server's operating system follows to accept and respond to service requests.NCP does not play a direct role in routing. However, it does provide session control and packet-level error checking between NetWare workstations and routers.

TCP and SPX: Transport-Layer Protocols

TCP and Sequenced Packet Exchange™ (SPX™) are transport-layer (OSI Layer 4) protocols. Standards at this OSI layer demand reliability from the end-to-end communication link. Accordingly, TCP and SPX provide guaranteed packet delivery and packet sequencing for IP and IPX, respectively.

Like NCP, TCP and SPX do not play a direct role in routing. These protocols are connected with internetworking only in that they guarantee the delivery of all routed packets.

Real-World Networking

In the real world, computer networks can combine a variety of physical and logical topologies. For the sake of simplicity and clarity, all of the subnetworks in the following discussion will be based on the NetWare client-server networking model.

So far we have discussed basic internetworks existing at one local site, with computers and other devices directly connected by some type of cabling. These types of networks are commonly referred to as LANs. In real-world applications, however, LANs are much more complex than the simple models we have discussed. In addition, there are networks composed of complex multiserver internetworks that exist at separate sites—which might be any number of miles apart—connected by long-range transmission media. These types of networks are known as WANs. The following sections discuss internetwork configurations that are more likely to be encountered in real-world situations.

One Server, Multiple Networks of the Same Type

The simplest form of an internetwork is two cabling (media) segments of the same MAC type sharing one network server.

For example, one server could contain two Ethernet network adapters, each supporting a different cable segment. There could be several computers connected to each cable segment in a star physical layout, with each cable segment using contention (CSMA/CD) for the MAC. Each of the cable segments would have a different network address; thus, each would be an independent subnetwork.

Together, the two separate networks would form an internetwork connected by internal routing capabilities built into the server.

Figure 32 illustrates the one-server internetwork described above.

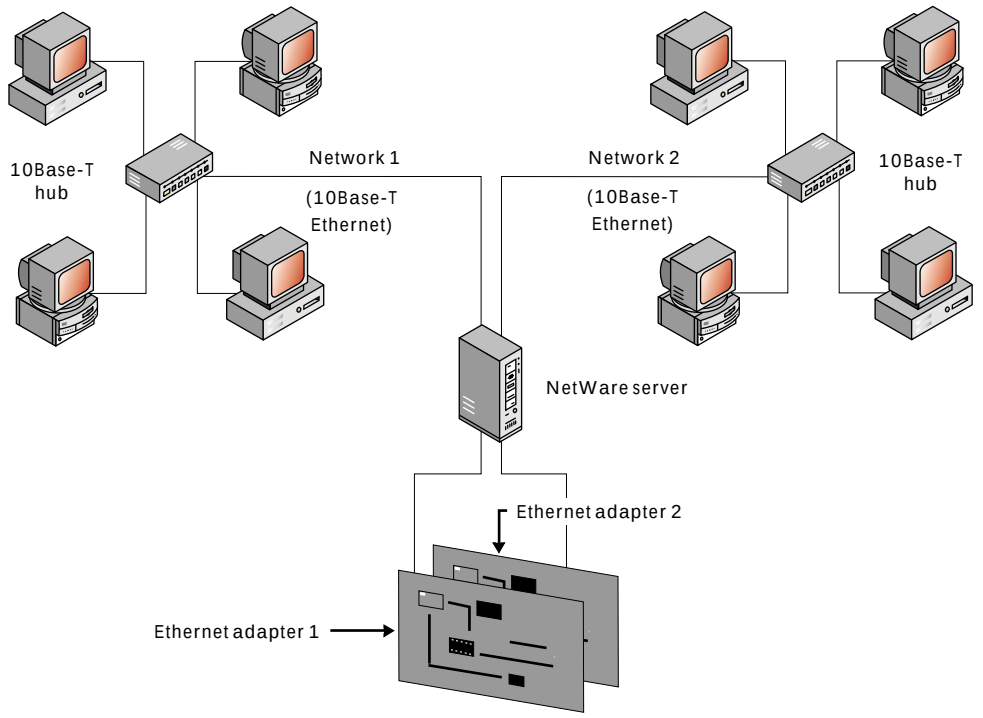


Figure 32

Internetworking two networks using the same type of network adapter in one NetWare server, by means of the server's internal routers

In the network illustrated in Figure 32, routing would be performed using NetWare IP or NetWare IPX, with support from the other NetWare routing protocols, as previously described.

Every NetWare server is capable of using internal routers to accomplish local network routing by means of the NetWare routing protocol set and AppleTalk. For larger or more complicated internetworks, or for departments with heavy server-processing requirements, NetWare or dedicated routers from other vendors provide the necessary routing power and capabilities.

One Server, Multiple Networks of Differing Types

In a slightly more complex internetwork, a NetWare server could support multiple cable segments using the same physical layouts but different MACs.

For example, a server could contain one Ethernet network adapter and one token-ring network adapter, with a cable segment attached to each. The Ethernet network might be connected in a physical star and use CSMA/CD for the MAC. The token-ring network might also be connected in a physical star, but it would use token passing for the MAC. Like the simpler configuration explained in the previous section, each cable segment would have a different network address. Figure 33 illustrates this more complex one-server internetwork.

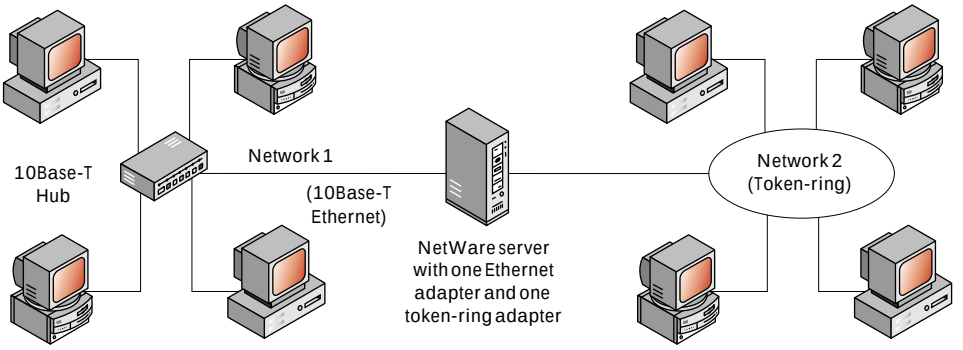


Figure 33

Internetworking two networks using different types of network adapters in one NetWare server, by means of the server's internal routers

In the case of the internetwork shown above, routing would again be performed using NetWare IP or NetWare IPX, with support from the other NetWare routing protocols.

Both one-server networks illustrated above support only two separate subnetworks. However, all NetWare servers are capable of supporting as many as 32 different network adapters (32 separate subnetworks) in any combination of same or different types.

Even though the token-ring network above was described as a physical star, it is drawn as a ring to signify that it is a token-ring network. In nearly all illustrations it is more important to make the logical topology clear than to be concerned with the physical topology.

Multiple Servers with Multiple Networks of Different Types

In an even more complex internetwork there may be multiple servers. For example, a complex internetwork might consist of two one-server subnetworks connected by a standalone router. Each server might contain multiple network interface adapters.

One server might contain two Ethernet network adapters and one token-ring network adapter, with a cable segment attached to each. One of the Ethernet adapters might support a PC network, and the other might connect to both PCs and Macintosh computers.

The other server might contain two Ethernet adapters that support PCs and Macintoshes. On one Ethernet, some nodes might be APs supporting wireless PCs and handheld devices.

Each server would have a unique internal number (server address), and each cable segment in each server would have a unique physical network (cable segment) address.

In this case, there would be five subnetworks on the internetwork: three attached to one server and two attached to the other. The internal server routers would perform the routing between any two workstations on subnetworks attached to the same server. Both the internal server routers and the intermediate standalone router would be involved in the routing between any two workstations on subnetworks attached to different servers.

Figure 34 illustrates the two-server internetwork described above.

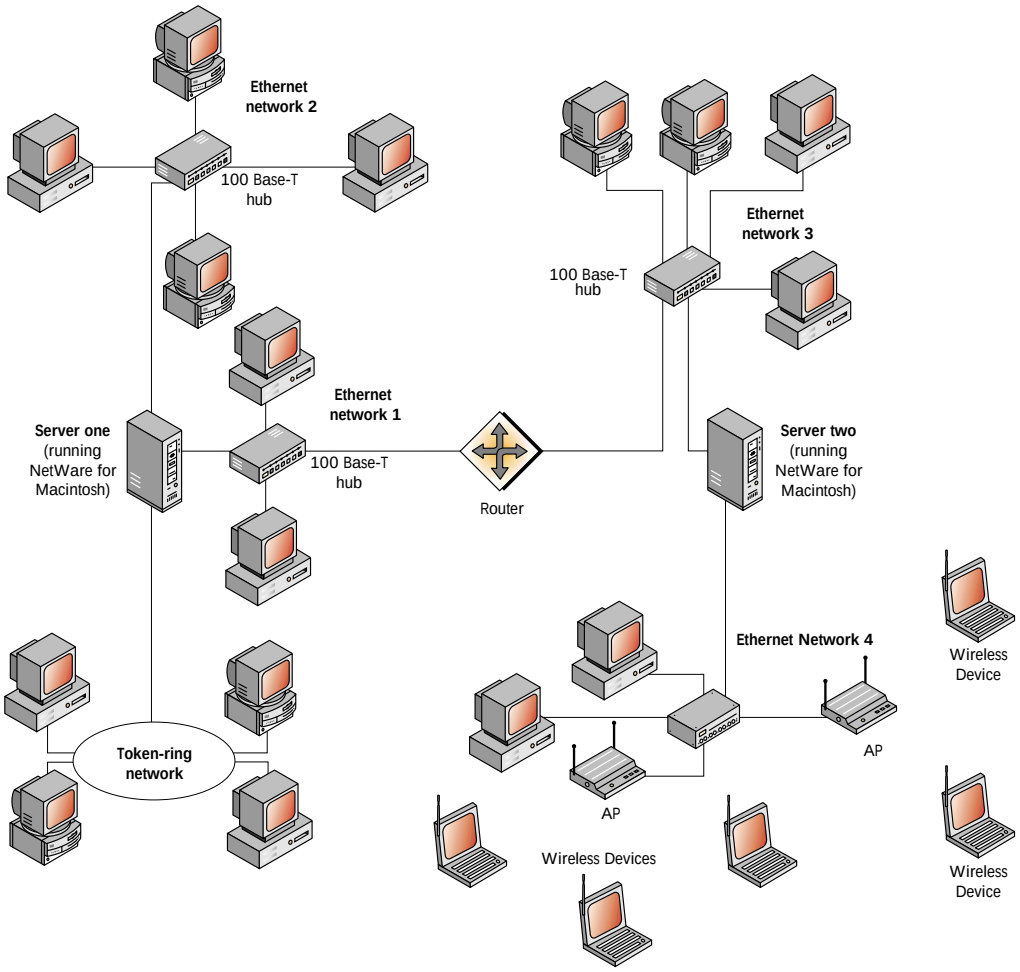


Figure 34

Internetworking multiple networks using different types of network adapters in two NetWare servers, by means of internal and standalone routers

Host Access

An already complex multiserver internetwork becomes even more complex with the addition of connections to host computer systems: mainframe computers such as IBM mainframes, minicomputers such as IBM's AS/400 or a VAX system, or other hosts such as UNIX workstations.

Host systems can provide access to other application software, additional resources such as data storage devices and printers, and increased processing power. For example, you might want to log on to an IBM AS/400 minicomputer to run an application available only on that computer, or to use its processing power for one task while you were using the processing power of your own workstation for another task. Or, you might want to print a large report on a high-speed printer connected to an AS/400 minicomputer.

The illustration in Figure 35 shows a multiserver NetWare network with an IBM mainframe, an IBM AS/400 minicomputer, and several UNIX workstations connected as host computers.

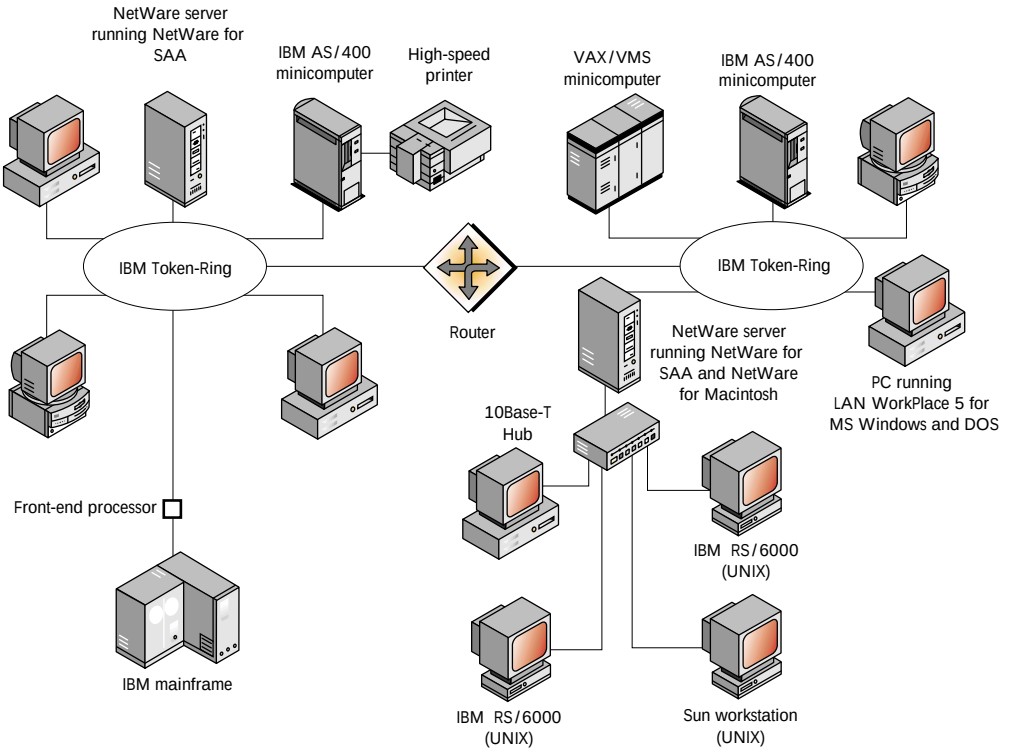


Figure 35

Host systems connected to a complex multiserver NetWare network

Clustering Services

As a recent addition to the world of computer networking, clustering services take networking to a new level. With the services, you can combine several servers on your network into a single unit with increased fault tolerance. The servers are "clustered" using specific software and programmed to keep tabs on one another so that the failure of any individual server will not bring the network down. When one server goes down, the others can immediately step in and provide the services it offered. This allows the administrators to repair downed servers or perform routine maintenance without interrupting network productivity.

Wide Area Networking

The traditional definition of wide area networking has been "connecting two or more networks existing at widely separate geographic sites." Some traditionalists also specify that the separate networks must be connected by means of common carrier telecommunication facilities (private companies that rent resources such as T1 lines and microwave transmission equipment).

Like any general term used in connection with rapidly changing technology, not everyone will agree on an exact definition of wide area networking. What is “widely separate”? Does the connection really have to be through a common carrier? Many major companies now own their own equipment linking networks that are many miles apart.

Suppose you connect two networks in two different buildings 100 yards apart by means of asynchronous modems and common telephone lines. Is that wide area networking? Most networking experts would say no: they would describe it as “one-site” or “campus” networking. What if the networks were two miles apart and separated by a major interstate highway? Or, what if they were 15 miles apart, on opposite sides of a major city? Many experts would still not call this wide area networking; they might use the term “metropolitan area networking.” Others consider metropolitan area networking a part of wide area networking.

Figure 36 shows two separate branch office internetworks connected to a third internetwork at a main corporate office. Each internetwork has multiple servers and existing host connections. One of the branch office networks is connected to the corporate network with a wireless router. The other branch office network is connected with a common carrier-provided intermediate link: in this case, a frame relay packet-switching network. However, either network could be connected by other means such as ATM or a dedicated leased-line link, perhaps using PPP.

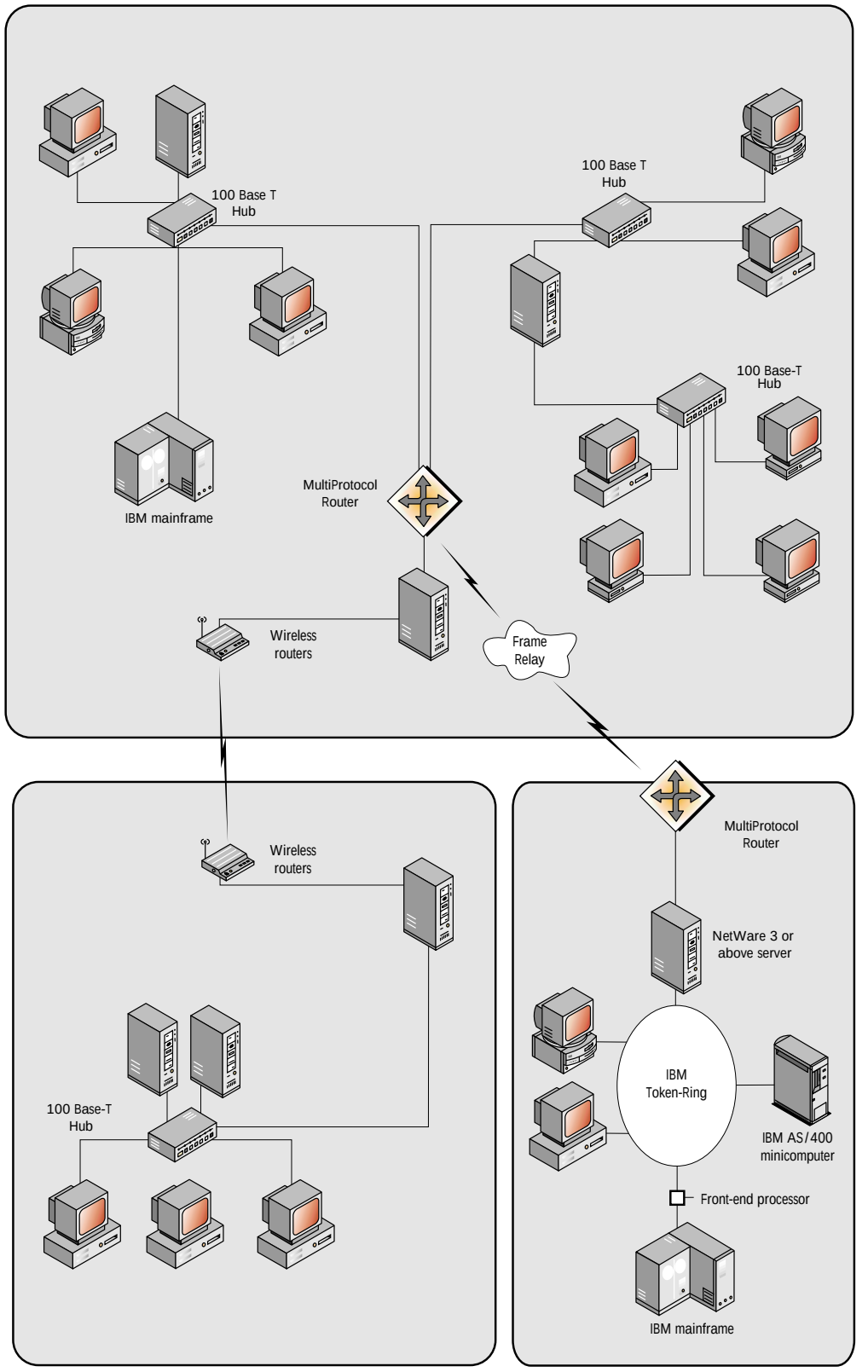


Figure 36

Wide area networking: three networks at widely separated sites connected through wireless routers and a frame relay connection

Important LAN and WAN High-Speed Technologies

In today's business environment, among the most widely discussed networking topics are the technologies that make networks faster, such as Fast Ethernet and Gigabit Ethernet, as well as the technologies that connect geographically distant networks, such as frame relay, ATM, and SONET. As networks grow, so does the amount of information sent across these networks. For this reason, transmission speed is of utmost importance. This section will provide brief descriptions of the following technologies that are dramatically improving the transmission speed on computer networks.

- Fast Ethernet
- Gigabit Ethernet and 10 Gigabit Ethernet
- Firewire and USB
- Fiber Channel
- Bluetooth
- 802.11
- HIPERLAN/1
- FDDI
- X.25
- Frame Relay
- ATM
- ISDN
- xDSL and Cable
- SONET
- Wireless Packet-Switched Networks
- CSC
- GPRS

The first seven items—Fast Ethernet, Gigabit Ethernet and 10 Gigabit Ethernet, Firewire and USB, Fibre Channel, Bluetooth, 802.11, and HIPERLAN/1—apply primarily to LANs. The other technologies in this list are reserved almost exclusively for WANs.

Fast Ethernet

100Base-T—also known as Fast Ethernet—is a high-speed LAN technology. It has been designated as the IEEE 802.3u standard and functions at the data-link (OSI Layer 2) layer's MAC sublayer, providing data transfer rates as high as 100 megabits per second (Mbps). Three kinds of wiring carry Fast Ethernet: 100Base-T4, which is four pairs of twisted-pair wires; 100Base-TX, which is two pairs of data-grade twisted pair wires; and 100Base-FX, which is two strands of fiber optic cable.

Distinguishing Characteristics

Like 10Base-T Ethernet, 100Base-T uses CSMA/CD as the MAC method. 100Base-T is based on the scalability of CSMA/CD. Scalability means that you can easily enlarge or downsize your network without degrading network performance, reliability, and manageability.

CSMA/CD was known to be scalable before the 100Base-T standard was created: a scaled-down version of Ethernet (1Base-5) uses CSMA/CD, provides data transfer rates of 1 Mbps, and enables longer transmission distances between repeaters. Because CSMA/CD could be scaled down, people reasoned that it could be scaled up. Specifying changes such as decreased transmission distances between repeaters produced a reliable data transfer rate of 100 Mbps, 10 times faster than traditional 10Base-T Ethernet.

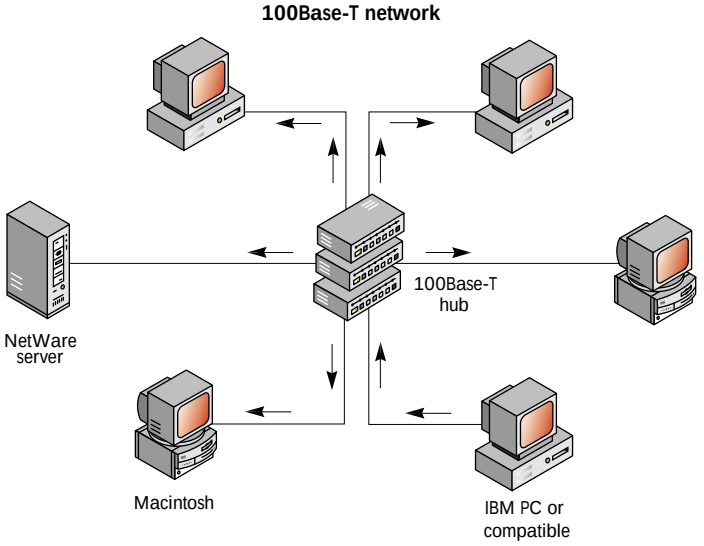


Figure 37

On 100Base-T networks, the physical topology is a star and the logical topology is a bus. A broadcast signal travels to all parts of the cable.

Advantages

100Base-T adapter cards and compatible cable are currently available from various vendors.

In addition, it is easy to upgrade from 10Base-T Ethernet to 100Base-T Ethernet. Both use CSMA/CD, and most network cards now support both 10 Mbps and 100 Mbps Ethernet. The adapter cards automatically sense whether it is a 10 Mbps or 100 Mbps environment and adjust their speed accordingly. Because 10Base-T and 100Base-T Ethernet can coexist, network supervisors can upgrade network stations from 10Base-T to 100Base-T one at a time, as needed. Moreover, most network supervisors are already familiar with CSMA/CD, so there is no need for expensive retraining.

100Base-T can be an inexpensive way to make your network faster. Adapter cards are not significantly more expensive than 10Base-T cards. In addition, Category 5 UTP cable (also called “CAT 5”) is relatively inexpensive.

Disadvantages

100Base-T reduces the maximum network size compared to 10Base-T because the standard specifies shorter transmission distances between repeaters. Compared to 10Base-T, 100Base-T reduces the maximum network diameter from 500 to 205 meters. For existing networks that exceed 205 meters, routers must be installed between 100Base-T network segments.

Gigabit Ethernet and 10 Gigabit Ethernet

Faster standards are always being developed and established and those for Ethernet are no exception. One standard is Gigabit Ethernet, also known as 1000Base-T or 802.3z. Gigabit Ethernet increases transmission speed on a standard Ethernet network to 1000 Mbps, or ten times that of 100Base-T. It was designed to function on the same cabling as 100Base-T so that upgrades would be inexpensive and straightforward. Right now, the primary focus for Gigabit Ethernet is as a backbone service for 100Base-T networks. As the hardware becomes more prevalent, however, 1000Base-T subnetworks and workstations should become more common.

In addition to Gigabit Ethernet, there is also an emerging standard known as 10 Gigabit Ethernet or 802.3ae. 10 Gigabit Ethernet will support data transmission speeds of 10,000 Mbps. Analysts predict that in some cases it could replace high-speed WAN technologies such as ATM and SONET.

Although Gigabit Ethernet is enjoying widespread adoption, 10 Gigabit Ethernet has not yet reached wide-scale acceptance in the world of computer networking. Part of the problem is that details of the standard, such as the medium over which it will be propagated, are not yet defined. Also, because Ethernet has not been a long-range standard in the past, it does not include management-control capabilities to alert network administrators when something goes wrong. However, Gigabit and 10 Gigabit Ethernet are gaining popularity as a solution for the thin-client application-on-demand model.

IEEE 1394 (Firewire) and USB

The IEEE 1394 standard (also known as Firewire) and the Universal Serial Bus (USB) standard are two standards that apply to data transmission between computers and peripheral hardware. Although these two standards are different, their application is fundamentally related and therefore they are covered in the same section here.

Distinguishing Characteristics

The IEEE 1394 standard is a high-speed standard developed for processing-intensive peripherals such as scanners, digital cameras, and removable storage devices. As a complementary standard, USB is more suitable for peripherals that do not require as much speed, such as mice and keyboards. Both standards use a simple cable with jacks similar to telephone jacks or Ethernet RJ-45 jacks. Most newer computers include USB ports; IEEE 1394 ports are integrated into higher-end computers.

Advantages

The most obvious advantage of both IEEE 1394 and USB is the ease of use. Both standards support "hot swapping" of peripheral components. This means that one device can be unplugged from the computer and another plugged in (and recognized by the computer) without having to reboot the system. This is not possible with standard parallel port or serial port connections. In addition, both 1394 and USB standards allow you to connect peripherals according to a tree topology or as a "daisy chain" (linked in a straight line) and then attach them to a single port on the computer—each peripheral does not require its own port. The IEEE 1394 standard allows for 63 devices per port, whereas USB will support more than one hundred. In addition, both standards support data transfer speeds higher than conventional ports. USB supports 12 Mbps data transfer, USB 2.0 supports 36–48 Mbps, and IEEE 1394 supports up to 400 Mbps (with several faster versions currently under development). USB is expected to replace serial and parallel ports for most simple peripherals such as mice and keyboards, and 1394 will meet the high-throughput requirements of video and real-time equipment.

Disadvantages

Although USB has gained widespread acceptance—USB-compatible peripherals are readily available—IEEE 1394-compatible hardware is harder to find. Part of the reason is that 1394 is the intellectual property of Apple Computer, which tightly controls the dissemination of the standard. In addition, 1394-compatible peripherals tend to be on the high end of the technological spectrum and therefore are much more expensive than their USB counterparts. Firewire and USB are also not compatible, so hardware manufacturers either have to install both ports or choose between the two: given the popularity of USB, Firewire often loses.

Fiber Channel

Fiber channel refers to a relatively new application for optical fiber components. The most common usage of fiber channel is in storage area networks (SANs), where it is used to connect clustered servers to storage systems. This technology is also being considered as an internal drive interface (between the hard drive and the processor within a computer) and as a high-speed switching service to connect several server clusters and SANs into a large interconnected network.

Distinguishing Characteristics

Fiber channel technology consists of optical fiber cables, specialized hubs, and Gigabit interface converters (GBICs). The GBICs are used to convert electrical signals into optical signals and vice versa. The cabling is divided into two categories: multimode and single-mode fiber. Multimode fiber has a larger diameter core and allows multiple transmissions to travel simultaneously. Single-mode fiber allows only one transmission path.

Advantages

Fiber channel technology has several advantages over other transmission media, but the most important is the speed of data transmission. Fiber channel supports data-transmission speeds of 100 Mbps. In addition, since the data is transmitted as a pulse of light rather than an electronic signal, it can travel much greater distances (up to 10 kilometers) before suffering any signal degradation. Likewise, the data is immune to electromagnetic interference and radiates no energy (no heat-shielding required).

Disadvantages

The main disadvantage of fiber channel is the cost: optical fiber is much more costly than conventional copper cable and more expensive to install. The advantages of fiber channel, however, greatly outweigh the cost in many applications.

Bluetooth

Bluetooth, a wireless standard developed by Ericsson, IBM, Intel, Nokia and Toshiba, is named after a Danish king who united Denmark and Norway in the 10th century. It is designed for short-range radio transmissions between devices no more than 10 meters apart. Bluetooth operates at a frequency of 2.4GHz and transmits at speeds up to 1 Mbps. (A next generation of Bluetooth will transmit at 2 Mbps.) Bluetooth is primarily for use in mobile devices to provide connectivity and synchronization; for example, two Bluetooth-enabled handheld devices a few meters apart can synchronize phone lists or schedules. The devices can connect on a one-to-one or one-to-many basis so that when near each other, Bluetooth devices create a “piconet”: an ad-hoc, peer-to-peer network of up to 10 nodes. For example, if all the participants at a meeting have Bluetooth-enabled laptops, they can create a piconet for sharing documents and messages. A Bluetooth-enabled printer in the room could be used by all without the need for cabling.

Advantages

Bluetooth is already a de facto standard that has garnered widespread interest and support among vendors. It supports both data and voice transmissions and does not require a line of sight.

Disadvantages

Bluetooth’s limited range makes it useful only for instances when the device is near another Bluetooth transmitter. Its data transmission rates are not nearly as fast as those of 802.11. It has so far been prohibitively expensive to implement, but it is predicted that the price will go down as vendor interest increases.

802.11

IEEE 802.11 is an extension of the Ethernet standard, adapted for wireless LANs. It consists of one MAC-layer standard and three physical-layer standards: two for radio transmissions (DSSS and FHSS) and one for infrared. 802.11 operates at 2.4GHz and can transmit at speeds up to 2 Mbps at a range of 30–100 meters. IEEE 802.11b, ratified in 1999, boasts transmission rates of up to 11 Mbps, but only over the DSSS physical layer.

Distinguishing Characteristics

Similar to its wired Ethernet counterpart, the 802.11 MAC layer uses a variation of CSMA/CD called carrier sense multiple access with collision avoidance (CSMA/CA). Unlike wired devices, a wireless device is unable to "listen" and transmit at the same time because the noise from transmission drowns out incoming signals; therefore, if there is a collision during transmission the device will not detect it. To solve the problem CSMA/CA provides for explicit packet acknowledgement. After each packet is sent, the receiver sends an acknowledge (ACK) packet to confirm that the packet arrived intact. If the ACK packet does not arrive, the sender assumes the original packet did not arrive, probably because of a collision, and sends the packet again.

Wireless environments are also susceptible to the "hidden node" issue in which a wireless node can hear traffic from the AP but not from another wireless node, due to distance or an obstruction. To prevent a collision between hidden nodes, 802.11 specifies an optional Request to Send/Clear to Send (RTS/CTS) protocol, also at the MAC layer. When this protocol is engaged, a sending device sends an RTS packet to the AP and waits for the CTS packet before transmitting. Because all other devices interfacing with that AP receive the same CTS packet, they delay any intended transmissions until the medium is free.

Advantages

Because 802.11 is a true Ethernet specification, 802.11 devices can be integrated seamlessly into conventional Ethernet LANs. With a laptop and an 802.11 network adapter card, an employee can roam throughout a building, go from building to building, or even go to a remote office and always be connected to the network.

Disadvantages

Unlike Bluetooth, 802.11 does not support voice transmissions, and the additional overhead from the ACK and RTS/CTS packets means 802.11 transmissions will always be slower than wired Ethernet. It is also feared that 802.11 transmissions can inadvertently disrupt critical Bluetooth transmissions, such as wireless medical or manufacturing monitoring devices.

HIPERLAN/1

High-Performance Radio LAN, Type 1 (HIPERLAN/1) is a standard developed by the European Telecommunications Standards Institute (ETSI) to improve on the data throughput rates of 802.11. It is the first in a suite of HIPERLAN standards that operate in the 5GHz range: HIPERLAN/2 is Wireless ATM, HIPERLAN/3 (renamed HIPERAccess) is for wireless local loop (the last segment between a home and the telephone system), and HIPERLAN/4 (renamed HIPERLink) is for wireless point-to-point connections. Of the four standards, only HIPERLAN/1 has been approved; the others are still in development.

Distinguishing Characteristics

The HIPERLAN/1 transmission scheme is the same as that for GSM, which means it uses TDMA as its air interface and Gaussian Minimum Shift Keying (GMSK) as its modulation scheme. HIPERLAN/1 can achieve data transfer rates up to 23.5 Mbps.

With HIPERLAN the MAC layer is subdivided into the Channel Access Control (CAC) layer, and the MAC layer. The CAC layer defines how a given channel access attempt will be made, depending on whether the channel is busy or idle and at what priority level the attempt will be made, if contention is necessary. Packets receive higher priority as they age.

Multi-hop routing support (not included in 802.11) is part of the HIPERLAN/1 specification. HIPERLAN-enabled devices choose a nearby "controller" and forward all outgoing traffic to that controller. The controller will then route the packet toward its destination. HIPERLAN-enabled devices also employ "hello" packets to announce their presence to other devices. In this sense, HIPERLAN-enabled devices behave similarly to conventional routers and are therefore able to structure their own network.

Advantages

HIPERLAN/1 is compatible with wired and wireless Ethernet. It is very stable and flexible.

Disadvantages

HIPERLAN is an emerging technology, so only a few HIPERLAN/1-compatible devices exist.

Fiber Distributed Data Interface

FDDI is also a high-speed LAN technology. It is not generally used for direct connection to desktop computers, but rather as a network backbone connecting two or more LAN segments to provide a path for data transmission between them. A simple backbone might connect two servers through a high-speed link consisting of network adapter cards and cable. An example of such a backbone is illustrated in Figure 38.

FDDI has been designated ANSI X3T9.5 and operates at the physical and data-link layers (Layers 1 and 2) of the OSI model. Like 100Base-T, FDDI provides data transfer rates as high as 100 Mbps.

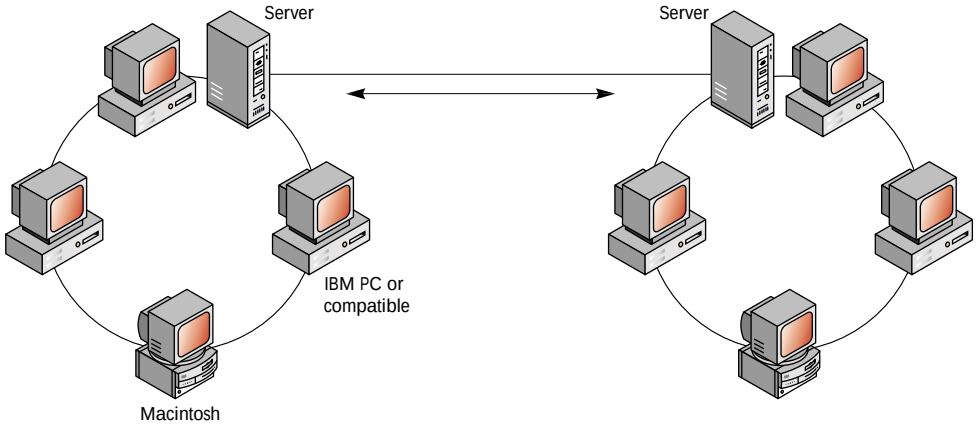


Figure 38

A simple server-based backbone connecting two LAN segments

Distinguishing Characteristics

FDDI networks have a dual, counter-rotating ring topology. This topology consists of two logical closed signal paths called “rings.” Signals on the rings travel in opposite directions from each other. Although both rings can carry data, the primary ring usually carries the data while the secondary ring serves as a backup.

On FDDI networks every node acts as a repeater. FDDI supports four kinds of nodes: dual-attached stations (DASs), single-attached stations (SASs), single-attached concentrators (SACs), and dual-attached concentrators (DACs). DASs and DACs attach to both rings, while SASs and SACs attach only to the primary ring. Several SASs often attach to the primary ring through a concentrator, so that an SAS failure will not bring down the entire network. If the cable is cut or a link between nodes fails, DASs or DACs on either side of the failure route signals around the failed segment, using the secondary ring to keep the network functioning.

FDDI uses token passing for its MAC method and is implemented using fiber-optic cable.

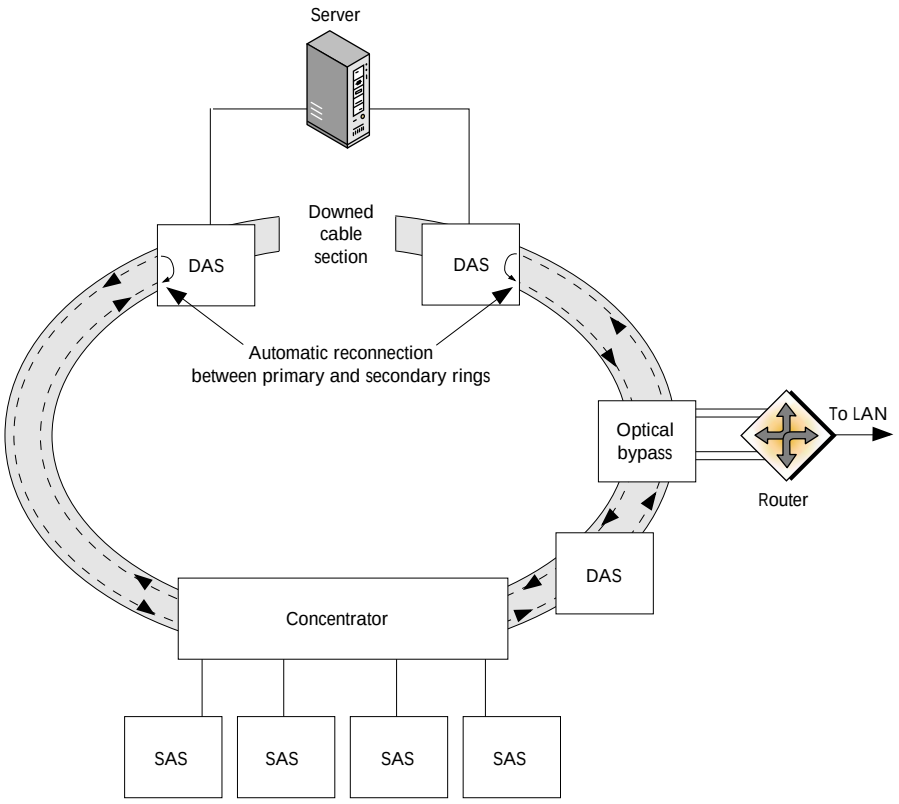


Figure 39

If a cable section on an FDDI network goes down, DASs on either side of the failed section automatically reconnect the primary and secondary rings. Also note that the server has a redundant connection to improve reliability.

Advantages

FDDI is a fast, reliable standard. The dual, counter-rotating ring topology increases the network's reliability by keeping it functioning even if a cable is damaged. FDDI also offers network management support, which was designed directly into the standard. In addition, the standard includes the Copper Distributed Data Interface (CDDI) specification for building a network using UTP cable (which is less expensive than fiber-optic cable).

Disadvantages

FDDI's main disadvantages are availability and price. Because FDDI is not useful for transmitting large graphic and sound files (such as video), it is being replaced by Gigabit Ethernet. And because most fiber backbones are now running SONET, FDDI hardware is difficult to find. Furthermore, FDDI adapter cards and fiber-optic cable are both expensive compared to other technologies offering the same speed. Fiber-optic cable installation also requires technicians trained specifically for this purpose. Even CDDI adapters (for copper wire), which are less expensive than FDDI adapters, are more costly than 100Base-T adapters. It is expected that FDDI will soon be replaced by other technologies.

X.25

X.25 is an ITU standard and includes data-link and physical-layer protocols (Link Access Procedure Balanced [LAPB] and X.21). X.25 provides data transfer rates of 9.6 Kbps to 256 Kbps, depending on the connection method.

Distinguishing Characteristics

X.25 specifies the interface for connecting computers on different networks with an intermediate connection through a packet-switched network. X.25 was defined when the quality of transmission media was relatively poor. As a result, the standard specifies that each node in the packet-switched network must receive each packet completely and check it for errors before forwarding it.

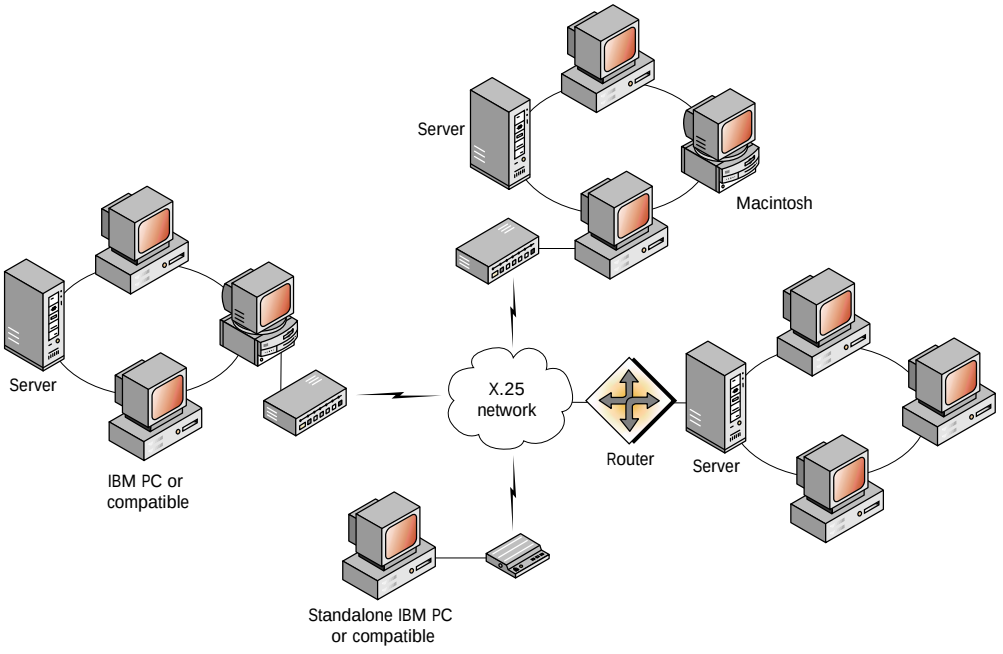


Figure 40

X.25 networks are often provided by telecommunication carriers.

Advantages

X.25 is well understood and reliable. Connections to X.25 networks can be made through the existing telephone system, Integrated Services Digital Network (ISDN), and leased lines. Because access is simple, it is comparatively inexpensive. X.25 is available worldwide, although its market share in the United States is rapidly decreasing. In countries with little digital telecommunications infrastructure, X.25 is the best WAN technology available.

Disadvantages

X.25 is slow compared to newer technologies. The process of checking each packet for errors at each node limits data transfer rates. X.25 also uses variable-size packets, which can cause transmission delays at intermediate nodes. In addition, many people connect to X.25 networks through modems, which limit data transfer rates to between 9.6 Kbps and 56 Kbps. Although X.25 is still in use in some areas, newer, faster standards such as ATM and frame relay have largely replaced it.

Frame Relay

Frame relay is a WAN technology. Approved by ANSI and the ITU, frame relay works at the data-link layer (OSI Layer 2) of the OSI model and provides data transfer rates from 56 Kbps to 1.544 Mbps.

Distinguishing Characteristics

Frame relay is an interface specification for connecting LANs over public packet-switched networks. This standard can be thought of as a simplified version of X.25 designed to take advantage of digital transmission media.

Frame relay services are typically provided by telecommunications carriers. Customers install a router and lease a line (often a T1 or fractional T1 line) to provide a permanent connection from the customer's site to the telecommunications carrier's network. This connection enables frame relay to use "permanent virtual circuits" (PVCs), which are predefined network paths between two locations.

With frame relay the router encapsulates (or frames) network-layer packets, such as IP and IPX packets, directly into a data-link-level protocol and sends them on to the packet-switched network. Like X.25, frame relay uses variable-size frames but it eliminates the error checking required on X.25 networks. A frame relay switch simply reads the header and forwards the packet, sometimes without first receiving the frame completely. Intelligent end stations must identify missing or corrupted frames and request retransmission.

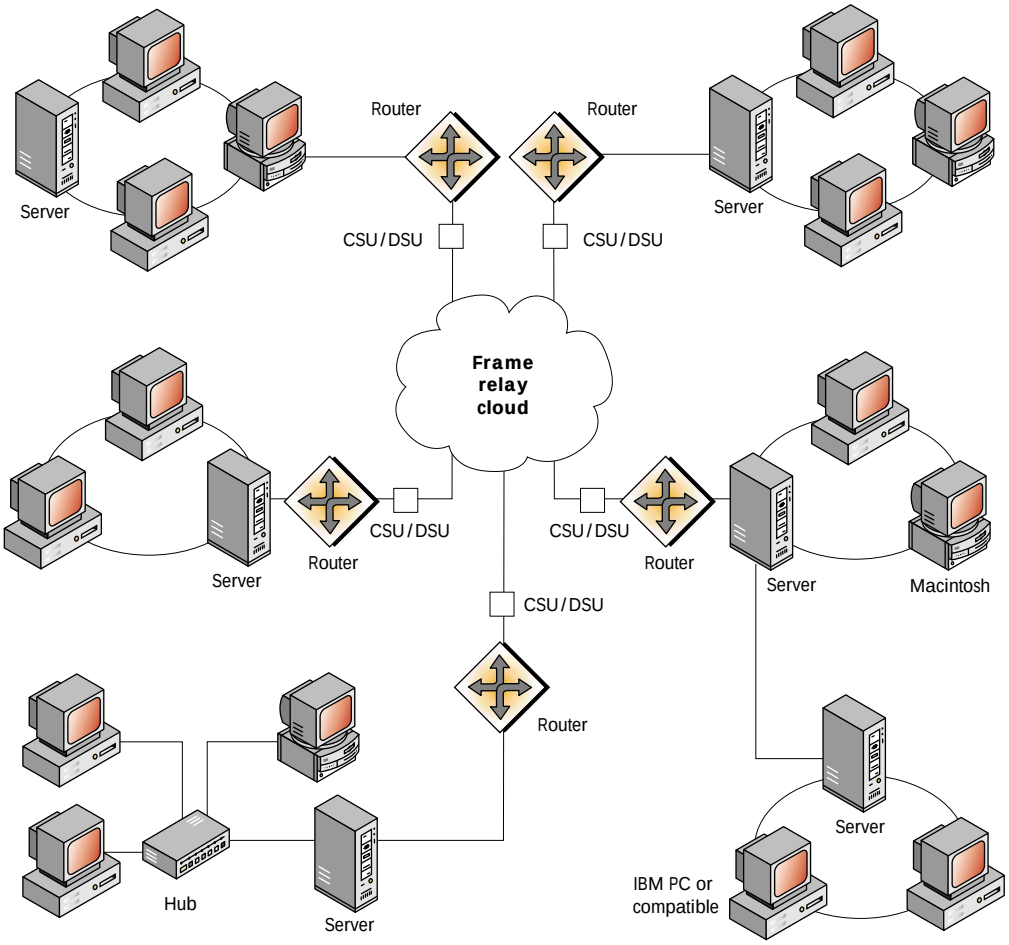


Figure 41

Frame relay is a WAN technology that enables companies to connect LANs through a telecommunications carrier's network.

Advantages

Frame relay uses PVCs over leased lines rather than a modem connection. PVCs transmit and receive data immediately, eliminating the call setup and handshaking that modems must perform. In addition, frame relay does not require error checking and flow control at the switches, thereby reducing overhead and leaving more bandwidth for data transmission. Frame relay is also a common standard in many countries. Finally, frame relay is less expensive than other WAN technologies because it provides bandwidth on demand rather than dedicating bandwidth regardless of whether data is being transmitted.

Although frame relay is fairly complex to implement, value-added resellers and most telephone companies will assist customers in determining their needs and will help install the technology.

Disadvantages

Frame relay's speed is limited because it uses variable-size frames, which can cause delays at switches along the frame's path. As a result, frame relay cannot support applications that require low latency (delayed response time) such as real-time video.

Asynchronous Transfer Mode

ATM is a WAN technology that is generally implemented as a backbone technology. The exact relationship of the ATM layers to the OSI model is currently undefined, although ATM LAN emulation works at the data-link layer (OSI Layer 2).

ATM provides data transfer rates of 100 Mbps and 155 Mbps. At the high end, WAN implementations using ATM and SONET together have achieved data transfer rates of 2.4 Gbps.

Distinguishing Characteristics

ATM is a cell-relay technology, meaning that it uses standard-sized packets called "cells." The size of an ATM cell is 53 bytes.

In a LAN implementation ATM functions at the data-link layer's MAC sublayer. It further divides the MAC sublayer into three layers: LAN Emulation, ATM Adaptation Layer (AAL), and ATM. LAN Emulation enables you to integrate ATM with Ethernet and token-ring networks without modifying existing Ethernet or token-ring protocols.

On a mixed network, LAN Emulation hardware sits between the Ethernet or token-ring segment and the ATM part of the network. It uses the three layers mentioned above to convert packets moving toward the ATM segment into cells, and to assemble cells moving toward the Ethernet or token-ring segment into packets. AAL and ATM put data into standard-sized cells. In most network computing situations, an ATM adaptation layer breaks packets into 48-byte blocks that are then passed to the ATM layer, where the five-byte header is attached to form a complete 53-byte cell.

Advantages

ATM offers high data-transfer rates, which have climbed into the gigabit range and are still increasing. One reason that ATM is so fast is its use of cells: because they are a standard size, ATM networks handle data in a predictable, efficient manner at the switches. Standard-sized cells and high-bandwidth media like fiber-optic cable also enable ATM to support real-time voice, video, and data traffic.

ATM also offers flexibility in its transmission media. As many as 27 ATM specifications exist for media like UTP, shielded twisted-pair (STP), and fiber-optic cable. (ATM is generally implemented with fiber-optic cable.)

Ethernet and token-ring networks can be integrated with ATM through use of LAN Emulation. The ATM network can emulate (or imitate) enough of the MAC layer of Ethernet and token-ring technologies so that higher-layer protocols can be used without modification. This allows existing network applications and network protocols to run over ATM networks, resulting in great cost savings.

Disadvantages

ATM is more expensive than the other high-speed LAN technologies and is extremely complex to set up and maintain; the expense is preventing many companies from implementing ATM.

ISDN

ISDN is a set of protocols defined by the ITU to integrate data, voice, and video signals into digital telephone lines. It functions at the physical, data-link, network, and transport layers (Layers 1 through 4) of the OSI model. ISDN offers data transfer rates between 56 Kbps and either 1.544 Mbps or 2.048 Mbps, depending on the country where it is implemented.

Distinguishing Characteristics

ISDN makes end-to-end digital connections over telephone lines. Although many telephone networks are almost completely digital, the local loop that connects a home or office to the telephone company's network usually is not. Most local loops send analog rather than digital signals. ISDN replaces local analog signaling with digital signaling, enabling end-to-end digital communications.

ISDN offers Basic Rate Interface (BRI) for individuals or small branch offices and Primary Rate Interface (PRI) for larger companies.

BRI uses two bearer, or B, channels (providing 64 Kbps each) to transmit and receive data, and one delta, or D, channel for call setup and management.

PRI is a T1 line. A T1 line in the United States consists of 23 B channels and one D channel, providing a total data transfer rate of 1.544 Mbps. A T1 line in Europe, known as an E1 line, consists of 30 B channels and one D channel, providing a total data transfer rate of 2.048 Mbps. A fractional T1 uses only some of the B channels in a T1 line (and thus offers some fraction of the total T1 data transfer rate).

ISDN requires special equipment at the customer's site, including a digital phone line and a network termination unit (NT-1). An NT-1 converts the bandwidth coming over the line into the B and D channels, and aids the phone company in diagnostic testing. The NT-1 also provides a connection for terminal equipment, such as ISDN telephones and computers that have an ISDN interface. In addition, the NT-1 provides terminal adapter (TA) equipment to connect equipment that is not compatible with ISDN. TA equipment provides an intermediary connection point: such equipment has an ISDN interface for connection to the NT-1, and a non-ISDN interface for connection to non-ISDN equipment.

Advantages

ISDN increases speed and broadens data transmission capabilities, especially for those currently using analog modems to remotely connect to an office or to access the Internet. It offers faster call setup and data transfer rates. The transfer rates are acceptable for transmitting voice, data, limited video, fax, and images. ISDN can also be used for limited LAN-to-LAN communications.

With ISDN you can transmit voice and data traffic simultaneously: over the same telephone line you can concurrently talk on the phone and download a data file to your computer. For example, one BRI ISDN configuration enables you to use the two B channels (128 Kbps) for data and part of the D channel for a telephone conversation.

Disadvantages

To understand ISDN well enough to simply order services requires considerable effort. Furthermore, configuration can be difficult. ISDN speeds are faster than those of a conventional modem (56 Kbps), but they are not as fast as ADSL or other emerging technologies.

xDSL

xDSL refers to the various types of Digital Subscriber Line (DSL), a relatively new, high-speed Internet access technology. With the explosive growth in computer networking, business networks, and the use of the Internet, the demand for fast and cost-effective access has also been growing steadily. Although there are many technologies available to provide high-speed Internet access—such as leased lines, wireless connections, and ISDN—most are expensive to install and costly to maintain. xDSL was developed as a low-cost alternative to these technologies. xDSL is a technology that uses the existing standard telephone cable (twisted pair) to provide data transmission speeds rivaling and often exceeding those of much more expensive solutions. Cable modems, which connect to the Internet via cable TV hookups, provide speeds and service options similar to xDSL. Many modems available on the market support both xDSL and cable.

Distinguishing Characteristics

xDSL lines are always “on,” which means that you do not have to establish a modem connection to use the Internet and then disconnect when you are finished—you are always connected. And unlike most dial-up Internet connections, xDSL can transmit both voice (or fax) and data signals over the same line at the same time. Conventional phone lines are capable of transmitting signals up to 1MHz, but voice transmissions only use the range between 1kHz and 4kHz. xDSL, on the other hand, transmits digital signals between 26kHz and 1.1MHz. A “splitter,” installed at the user site, divides the signal into digital and analog signals for data and voice, respectively. For this reason, xDSL is especially attractive to home and small-business users who do not want to install separate lines for voice and data transmissions.

DSL comes in a variety of “flavors,” the most common of which are listed below:

ADSL

The most popular kind of consumer xDSL is Asymmetric DSL (ADSL), also called G.dmt and Full-rate ADSL (ITU-T G.992.1). By asymmetrically dividing the available bandwidth, ADSL allows you to receive data much faster than you send it. With respect to an Internet connection, this is the optimal configuration: you send out much less data (about five percent of total transmission) than you download. Standard ADSL provides for downstream (data received) speeds of 8 Mbps and upstream (data sent) speeds of 1.5 Mbps.

ADSL Lite

A variant of ADSL is G.Lite, also ADSL Lite, splitterless DSL, and Universal DSL (ITU-T G.992.2). ADSL Lite provides a data transmission rate of 1.5 Mbps downstream and 512 Kbps upstream. ADSL Lite does not require a visit from the phone company because the splitting is managed at a central location, an arrangement which sacrifices some speed to save cost. It is expected to become the most widely installed form of DSL in private residences even though it does not handle voice and entertainment applications well.

RADSL

Rate Adaptive Asymmetric Digital Subscriber Line (RADSL) is similar to ADSL but it "listens" to see how much traffic is on the wire and adjusts its speed accordingly. Depending on the distance from the telephone company's central office, RADSL can transmit at speeds up to 7 Mbps downstream and 1.5 Mbps upstream (10,000 feet for the highest speed, 17,000 feet for the lowest).

HDSL

High bit-rate DSL (HDSL) is symmetric, meaning that it provides an equal amount of bandwidth upstream and downstream. Generally used for wideband connections within a corporation and between telephone companies and their customers, HDSL is often used in lieu of a T-1 connection. The data transmission rate varies depending on how many twisted-pair wires are used. Two wires provide 1.5 Mbps; three carry 2 Mbps. HDSL II, a variant of HDSL, provides the same speeds as HDSL but over a single wire.

SDSL

Symmetric DSL (SDSL) also provides the same speeds as HDSL but differs in two important ways: it is limited to 10,000 feet and it uses only one wire. It also uses the same modulation technique as ISDN. SDSL is a forerunner to HDSL II.

IDSL

ISDN DSL (IDSL) is a hybrid between ISDN and DSL. Using the same modulation technology as IDSN, IDSL bypasses voice lines and uses the less-busy data network instead. Transmission speeds are slightly higher than those of ISDN.

VDSL

Very high bit-rate DSL (VDSL) is a developing technology that will provide higher transmission rates over shorter distances. At 1,500 feet the speeds may be as high as 13 Mbps and at 1,000 feet an amazing 52 Mbps. High-definition television signals are one proposed use for VDSL.

	Type	Downstream	Upstream	Distance	Medium
ADSL	Asymmetric	8.0 Mbps	1.5 Mbps	18,000 ft	single twisted pair
ADSL Lite	Asymmetric	1.5 Mbps	512 Kbps	18,000 ft	single twisted pair
RADSL	Asymmetric	7.0 Mbps	1.5 Mbps	17,000 ft	single twisted pair
HDSL	Symmetric	1.5 Mbps	1.5 Mbps	15,000 ft	2 twisted-pair wires
		2.0 Mbps	2.0 Mbps	15,000 ft	3 twisted-pair wires
SDSL	Symmetric	1.5 Mbps	1.5 Mbps	10,000 ft	single twisted pair
IDSL	Symmetric	144 Kbps	144 Kbps	18,000 ft	single twisted pair
VDSL	Asymmetric	13–52 Mbps	1.5–3.2 Mbps	1,000 ft – 4,500 ft	fiber-optic cable

Advantages

The most obvious advantage of xDSL is the low cost. Although the modems are fairly expensive, the price is still less than that of leased lines and ISDN. Also, many phone companies and ISPs are offering G.Lite service for a monthly fee with low or no installation charges. And because xDSL uses existing telephone wiring, you do not need to have new lines installed. The next most important advantage is the data transmission speed: ASDL offers 8 Mbps downstream as opposed to 128 Kbps with ISDN.

Disadvantages

The primary disadvantage of xDSL is security. Because the connection is always on, it is easier for "crackers" (unscrupulous users who break in to others' computers) to find your computer on the Internet. Those who subscribe to xDSL will need to ensure they have sufficient security in place to prevent identity theft or other mischief. Some solutions are security software, firewalls, switches, and modifying OS settings to reduce the chance of a security breach.

Another disadvantage may be availability. Because of the nature of the technology there is a limit placed on how far a subscriber can be from a major telephone line switching hub. For that reason, xDSL may not be available in your area; however, most phone companies and many ISPs are rapidly increasing the scope of xDSL availability.

Synchronous Optical Network

SONET, also known in some countries as Synchronous Digital Hierarchy (SDH), is a WAN technology that functions at the physical layer (OSI Layer 1) of the OSI model. SONET has been accepted by ANSI and recommended by the ITU. It specifies a number of data transfer rates from 51.8 Mbps to 13.21 Gbps.

Distinguishing Characteristics

SONET defines a fiber-optic standard for high-speed digital traffic. This standard provides the flexibility to transport many digital signals with different capacities.

Data communications sometimes prove difficult because digital signaling rates can vary. For example, as stated in the above section on ISDN, in the United States a T1 line provides 1.544 Mbps, while in Europe an E1 line provides 2.048 Mbps. SONET resolves such problems by defining how switches and multiplexers coordinate communications over lines with different speeds, including defining data transfer rates and frame format.

SONET defines a number of Optical Carrier (OC) levels. Each level defines an optical signal and a corresponding electrical signal called Synchronous Transport Signal (STS). The base level is OC-1/STS-1 or 51.84 Mbps. Each level's rate is a multiple of 51.84 Mbps. The table below shows the OC levels and the corresponding data transfer rates that SONET defines.

OC Level	Data Rate
OC-1	51.8 Mbps
OC-3	155.5 Mbps
OC-9	466.5 Mbps
OC-12	622.0 Mbps
OC-18	933.1 Mbps
OC-24	1.24 Gbps
OC-36	1.86 Gbps
OC-48	2.48 Gbps
OC-96	4.976 Gbps
OC-192	10 Gbps
OC-255	13.21 Gbps

SONET also provides easy access for low-speed signals such as DS-0 (64 Kbps) and DS-1 (1.544 Mbps) by assigning them to sub-STS-1 signals called "Virtual Tributaries."

Advantages

The SONET standard defines data transfer rates and a frame format that all vendors and telephone companies throughout the world can use, creating a framework for global networking. SONET also includes management capabilities for telephone company equipment. Cell relay technologies such as Switched Multimegabit Data Services (SMDS) and ATM operate above SONET, making SONET the foundation for many broadband services. And because SONET uses a ring topology, line breaks and equipment failures barely affect service.

Disadvantages

SONET is primarily a technology for voice transfer. Developments in optic fiber and data-transfer technologies such as in ultra long-haul dense wavelength division multiplexing (DWDM) may make SONET obsolete.

Wireless Packet-Switched Networks

To accommodate the demand for data transmission capabilities in mobile devices, many service providers offer access to wireless packet-switched networks. A packet-switched network does not require a continuous connection the way circuit-switched networks do (the telephone system is circuit switched). In the United States, the principal wireless packet-switched networks use one of the following technologies: DataTAC, Mobitex, Cellular Digital Packet Data (CDPD) and microcellular data network (MCDN). DataTAC was created by Motorola and IBM, Mobitex was developed in Sweden by Eritel, CDPD by the Wireless Data Forum, and MCDN by Mobitel.

Distinguishing Characteristics

DataTAC networks use one of two protocols: MDC4800 and Radio Data Link Access Procedure (RD-LAP), the former providing raw throughput rates up to 4.8 Kbps and the latter up to 19.2 Kbps (actual data rates are about half that after factoring out overhead). DataTAC networks are hierarchical networks wherein base stations connect to area communications controllers that connect to message switches. Customers' fixed-end systems connect to message switches. Depending on the area, DataTAC networks use between one and ten 25 kHz channels in the 800 MHz range. Connection to wired networks is through X.25 or TCP/IP protocols.

Mobitex, on the other hand, operates on the 400MHz or 900MHz frequency with an 8 Kbps data rate on 12.5 kHz channels. Each Mobitex packet, called an MPAK, can be no more than 512 bytes long. The Mobitex system is also hierarchical, consisting of network management centers, switches, and base stations that cover up to 30 km per cell. The standard Mobitex configuration connects via the X.25 protocol to wired networks, but IP connectivity is available as well.

CDPD differs from the other two in that it transmits over the existing analog cellular infrastructure rather than radio transmission towers. CDPD is non-proprietary, which means that you can use the same software to access any CDPD network. A faster alternative than DataTAC or Mobitex, CDPD also supports more operating systems than the other two. CDPD also uses RD-LAP and is based on TCP/IP.

MCDN, called Ricochet by Metricom, is a high-speed, "desktop quality" wireless packet-switched network designed for laptop computers. It operates at speeds of 28.8 Kbps, and a 128 Kbps implementation is now under way. MCDN uses FHSS and operates at the 900MHz, 2.3GHz, and 2.4GHz frequencies. MCDN employs microcell radio transceivers and a wired access point. The wired access point cell covers about 20 square miles, and within that cell the microcell radios are positioned in a checkerboard pattern (approximately 100 per cell), connected according to a wireless mesh topology. These radios are about as big as a shoe box and are installed on streetlight and utility poles. An MCDN user has a special modem connected to the laptop that connects with the closest microcell radio. The radio then sends the signal to the next radio and the next until it reaches the wired access point. From there, the signal is carried to the Internet or a corporate LAN via a T1 or fiber optic line. MCDN is based on IP protocols.

Advantages

Devices for all four technologies operate in an "always on" mode; that is, they do not require dedicated connections that need to be established every time they are used. Because the networks are packet switched (as opposed to circuit switched, like the conventional telephone system) the devices send data in small bursts, allowing several devices to use the same frequency at the same time. The technologies also employ store-and-forward techniques to ensure that packets are delivered even when the remote device is temporarily unavailable.

Disadvantages

Except for MCDN, the technologies are not capable of transmitting large files: video conferencing, Web browsing, file transfer, and high-speed multimedia cannot be accommodated. And although MCDN can provide these services, the network is available only in a handful of cities.

Circuit-Switched Cellular

Circuit-Switched Cellular (CSC) technology, like CDPD, runs on the existing analog cellular infrastructure; however, it is capable of handling larger file transfers, albeit at a slower rate than CDPD.

Distinguishing Characteristics

CSC uses the established analog frequencies and infrastructure the same way a conventional modem uses land-based telephone lines: with a cellular phone and a CSC-enabled modem, you call another modem to establish a connection with it. Because many wired modems do not accept wireless protocols, many providers have set up modem pools that receive the cellular modem signals and translate them into the standard voice frequencies of the conventional modem before sending the signals to the intended modem. CSC operates at air frequencies between 824MHz and 894MHz and offers transmission rates between 14.4 Kbps and 20 Kbps.

Advantages

CSC technology has the same coverage area as standard analog cellular; in other words, it is nearly ubiquitous.

Disadvantages

Because it uses a continuous connection on the same frequencies as analog voice cellular, providers often charge the same rate for CSC as for voice. For infrequent data transfers this may not be expensive, but for lengthy Internet connections it can become so.

GPRS

General Packet Radio Service (GPRS) is a connection-oriented technology that runs on top of the existing GSM and TDMA infrastructure to provide a wireless packet-switched network similar to the Internet.

Distinguishing Characteristics

GPRS has been developed as an intermediary step between second- and third-generation cellular systems. Unlike CSC, GPRS traffic does not actually go through the GSM network but only uses the GSM network to look up user profile data. It employs between one and eight channels that can be used by several devices at once and offers transmission rates between 14.4 Kbps and 115 Kbps. GPRS supports Quality of Service (QoS), and because connection setup is fast, it appears to users that their connection is always on.

Advantages

GPRS can provide most of the functionality of the current wired Internet, with services such as chat, textual and visual information, still and moving images, document sharing and collaborative work, audio, e-mail, remote LAN access, and file transfer.

Disadvantages

Voice and GPRS calls occupy the same resources, which means that if there is heavy voice traffic on the GSM or TDMA network, fewer resources will be available for GPRS links and vice versa. GPRS also does not have store-and-forward capabilities.

Internet Technology

The Internet

The Internet's history can be traced to 1957, when the Department of Defense (DoD) formed the Advanced Research Projects Agency (ARPA) in response to Russia's launch of Sputnik, the first artificial earth satellite. ARPA (later renamed DARPA) sponsored a number of studies to research how to make a few university supercomputers available to many research scientists across the country. In 1969, the first computer network was created. Called ARPANET, it interconnected UCLA, Stanford Research Institute, and UC Santa Barbara in California with the University of Utah. As time passed, more and more organizations joined this growing computer network.

Now the Internet is the world's largest computer network, linking thousands of networks and millions of individual computers around the world. The interlinked networks and individual computers belong to a myriad of private individuals, government agencies, universities, elementary and high schools, hospitals, private businesses of all kinds, and other organizations in almost every country in the world. On a daily basis, millions of users send and receive e-mail, download and upload files, do research, and conduct business on the Internet. E-commerce alone is expected to account for billions of dollars in sales over the next few years.

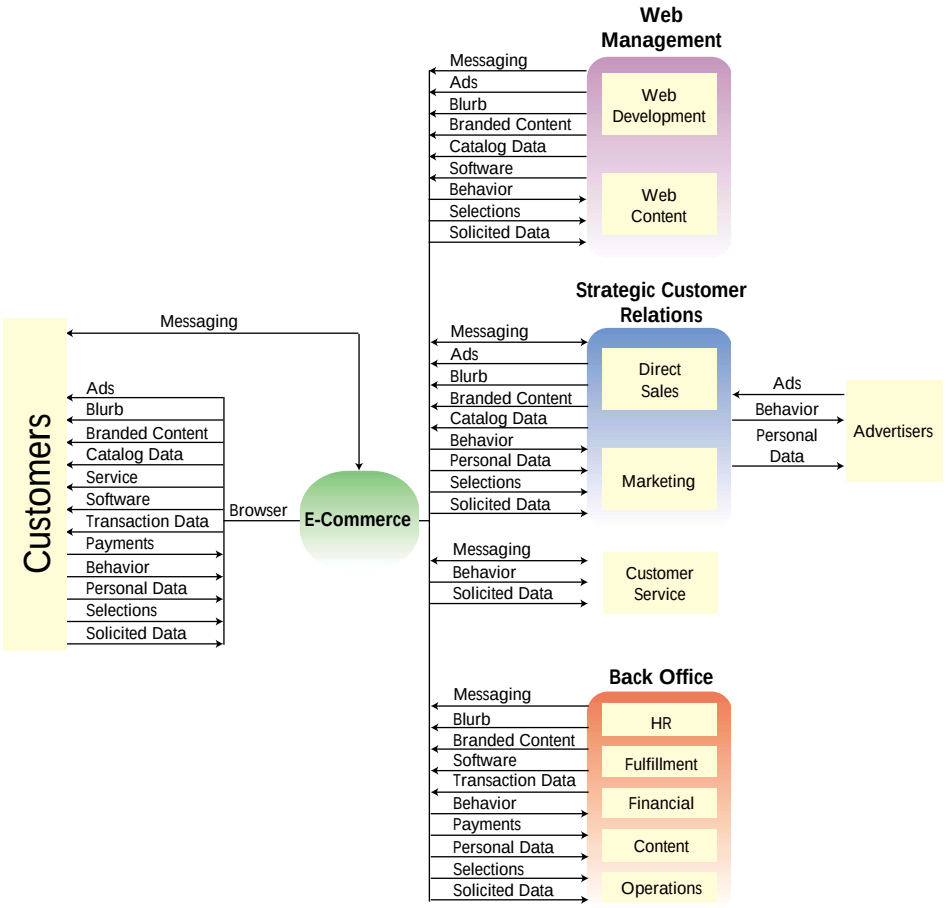


Figure 42

Performed over the Internet, e-commerce provides a direct connection between businesses and their customers. Immediate access to customer feedback and purchasing patterns can give businesses a decisive edge over their competitors.

Because the Internet is so popular and widely used, many of its technologies have spilled over into the private computer networking market. Two of the most influential technologies are the TCP/IP suite and the World Wide Web (WWW), which is based on the HTTP and HTML protocols. These technologies—and the addressing scheme that supports them—have become integral to computer networking and are shaping the future of the industry. The following sections define these technologies and explain how they are affecting the world of computer networking.

TCP / IP

The TCP/IP suite was originally developed by the DoD to connect a system of computer networks that became known as the Internet. TCP/IP is actually a group, or suite, of networking protocols used to connect computers on the Internet. TCP and IP are the two main protocols in the suite.

TCP provides transport (OSI Layer 4) functions, ensuring, among other things, that the amount of data received is the same as the amount transmitted. The IP part of TCP/IP provides the addressing and routing mechanism (OSI Layer 3).

TCP/IP uses a special transmission method that maximizes data transfer and automatically adjusts to slower circuits and other delays encountered on the network.

The TCP/IP suite includes a file transfer capability called FTP, which allows files containing text, programs, graphics, numerical data, and so on to be downloaded off of or uploaded onto a network. Simple Mail Transfer Protocol (SMTP) is TCP/IP's own messaging system for e-mail. In addition, the Telnet protocol provides terminal emulation, allowing a personal computer or workstation to act as a terminal, or access device, for a larger mainframe computer. TCP/IP also includes Telnet for remote login capabilities and User Datagram Protocol (UDP), which is used to deliver non-essential data (data which requires no confirmation of receipt) over the network.

Because of the increasing importance of the Internet as well as TCP/IP's versatility, more and more companies are using TCP/IP as the primary protocol in their LANs. A LAN that uses Internet technology such as TCP/IP and Web browsers is called an "intranet." Novell has supported TCP/IP and intranets in its NetWare network software for some time. With the release of NetWare 5 network software in 1998, TCP/IP became the default protocol of Novell networks. Novell's NetWare 5 and later NOSs enable pure TCP/IP to be used in a LAN, without the aid of any other networking protocols.

World Wide Web (WWW)

The World Wide Web has become the dominant Internet service, taking only a few short years to catch on after it was introduced to Internet users in 1991 by Tim Berners-Lee and CERN (a European consortium for nuclear research).

The World Wide Web is a client-server environment. Information is managed through Web sites on computers called Web servers. You access these sites using the client software on your individual computer and the Internet's HTTP. The client software is called a Web browser. Netscape Navigator and Microsoft's Internet Explorer are examples of popular Web browsers.

The computers and Web sites on the Internet are linked through documents called Web pages. The basic format of a Web page is a text document written in HTML, which is made up of codes that tell how the page will be displayed. The HTML document also includes the text that will be displayed as well as the addresses, or Uniform Resource Locators (URLs), of other Web pages that have links in the document. These links appear as underlined or highlighted text that includes hidden cross-references, or hyperlinks, to additional information. Clicking on this highlighted text allows you to jump to the Web page referenced by the link. Web pages may also display icons and images as links to other pages.

In order for these links to work, however, the addressing scheme must be very specific and the links must reference an appropriate URL. The URL is then used to determine the location of the site referenced by the link. This method of Internet addressing is discussed in the following section.

Internet Addresses

For Web site identification, e-mail routing, and many other purposes, every machine on the Internet is identified by a unique number known as an IP address. The IP address is a binary number 32 bits long, specified in IP, also called IPv4. It is usually written in "dotted decimal" format by dividing it into four eight-bit numbers and converting each number to its decimal equivalent, which is a number from zero to 255. The four numbers are then separated by dots, like this: 199.104.124.97.

This address system provides 4.3 billion possible addresses. When the Internet was first set up under ARPANET, this number appeared more than adequate. But current estimates indicate that all available 32-bit IP addresses will soon be exhausted. The initial framers of the Internet did not anticipate that nearly every business, organization, government institution, and human being in the world would eventually want an Internet address.

In 1994, an Internet oversight committee called the Internet Engineering Task Force (IETF) set up specifications for a new IP version, IPv6, that will solve this problem. IPv6 employs a 128-bit addressing scheme instead of the 32-bit scheme used before. As a result, IPv6 addresses are much more complex than their predecessors: instead of four eight-bit numbers, IPv6 addresses consist of eight four-digit, hexidecimal numbers. For example, an IPv6 address would look something like this: 2EG3.0000.1323.0000.6HE2.CDDE.2546.AB76. This new addressing scheme supports well over forty-undecillion (40×10^{36}) possible addresses.

For computers, keeping track of these numbers is no problem, but for their human counterparts, numbers can be difficult if not impossible to remember. Humans generally prefer names. To remedy this problem, the Domain Name System (DNS) was introduced in 1984.

DNS assigns each IP address a corresponding domain name made up of letters or words organized in a hierarchy or inverted tree. At the top of the hierarchy are the "top-level domains." The following is a list of top-level domains:

- .com: commercial organizations, as in novell.com
- .edu: educational organizations, as in ucla.edu
- .gov: governmental agencies, as in whitehouse.gov
- .mil: United States military organizations, as in army.mil
- .org: nonprofit organizations, as in redcross.org
- .net: networking entities, as in compuserve.net
- .int: international organizations, as in nato.int

Novell has created a subdomain under the .com top-level domain, and it is identified on the Internet as novell.com. Within Novell this domain is further divided into subdomains such as provo.novell.com, sjf.novell.com, and ukb.novell.com.

When you are on the Internet and you type in novell.com on the HTTP command line, your computer contacts another computer that has a list of .com domains paired with their assigned IP addresses. That computer translates novell.com into the IP address that identifies the main Novell Web server, sends your data on to routers that recognize the number, and finally connects your computer with the Novell World Wide™ home page.

Computers that keep and maintain such DNS translation lists are called name servers. They also keep lists of other name servers. If e-mail or some other message comes their way that does not apply to any of their subordinate domains, they send it off to a name server that may have the needed address. Eventually, a name server with the correct domain is found and the message is sent on to its destination. The IP address of a computer may change, but the name servers keep track of such changes and maintain the same domain name, so we do not have to worry about it.

This addressing scheme, along with the TCP/IP suite and the World Wide Web, has drastically changed the future of computer networking. Businesses and individuals who have become familiar with the Internet and its workings have recognized the advantages of this technology and applied it to their business networks in the form of intranets, virtual private networks (VPNs), and extranets. The following sections detail the various technologies used and the advantages obtained through the use of these new networking ideas.

Intranet

An intranet is a privately-owned, secure, business network based on Internet technology, although not necessarily connected to the Internet. The term “intranet” appeared when companies discovered that they could use Internet technologies to make company-internal information available to all employees, no matter where the employees were located or what kind of hardware they were using; that they could still secure the information from unwanted access by outsiders; and that, along with these advantages, they could make the information available at the lowest possible cost.

The main reason for a company to implement an intranet is that it enables a business to collect, manage, and disseminate information more quickly and easily than ever before—even much more quickly and inexpensively than with other current means of electronic communications, including e-mail and other types of cross-platform publishing (in computing parlance, “publishing” refers to the act of making a document available for others to access electronically). In fact, intranet publishing is the ultimate in cross-platform publishing because it is based on the Internet technologies that were developed specifically for the purpose of allowing information-sharing among dissimilar computing systems.

Although even a small company with only one office and a small network can benefit from an intranet, the value of an intranet increases with the number of employees, the size of the network, and the number of geographically separate sites. As a company grows, if it continues to use conventional means of information dissemination such as printed memoranda and newsletters, the cost of disseminating information to all employees increases exponentially. Other methods of sharing information, such as e-mail and file sharing, also fall short of the cost savings and immediacy that can be obtained through intranet publishing.

On an intranet any employee with a properly configured workstation and a Web browser can read documents as soon as the files are completed and copied to any Web server, regardless of where the employee is located. If a company were to instead disseminate documents as files in a public directory or by e-mail, the documents would have to be provided in multiple formats to accommodate the various computing platforms and applications used within the company. The company would need to pay employees to prepare the differently formatted documents and distribute them to the locations where they could be accessed. In even a small company this type of effort takes more time and costs far more than does publishing the same information once in HTML format on a single Web server. In a large company the time and cost differences can be enormous.

Intranet publishing has other benefits. One important advantage is that the network can update your intranet documents automatically in real time. For example, if you published a document that contained the stock price for your company or news about the market in which your company competes, you could create a Web server script that would automatically update the document every 15 minutes with the most current stock price and market news. With immediate access to up-to-date information, you can respond more quickly to changes in the marketplace. Moreover, after the script is created, the network continues to update the information at no further cost; the work of updating is not forgotten or lost because you get too busy.

In addition, you get immediate feedback about the documents published on your intranet. With paper-based documents or publicly available files stored on a server, you cannot determine whether or not people are reading the documents. If you published the documents on an intranet server, however, the network could track how many people read the documents and which documents were used the most.

Businesses are continually finding more ways to use intranets to decrease costs, especially since the specification for World Wide Web documents has been extended to include graphics, audio clips, and movies. For example, many companies have installed applications that allow employees to access company databases directly from a Web browser, thereby avoiding the cost of specialized database access programs. Recent products such as Novell’s GroupWise allow employees to read their e-mail messages and view and modify their appointment schedules directly from a Web browser.

Another factor that makes an intranet valuable is that it can be configured to enable you to access it over the Internet. Traveling employees, suppliers, and customers can access any information published on the intranet over the Internet, but you can still control access to information. For example, you might allow the general public to view some documents and restrict access to other documents to authorized users. Also, you can allow employees using your intranet to connect to the Internet and to access the vast information resources available there.

Extranet

An "extranet" is two or more intranets connected in such a way that they enable collaboration among the companies that own the separate intranets. On an extranet each connected company usually makes some selected part of its intranet accessible to the employees of one or more other companies. For example, several companies might create an extranet to consolidate data gathering and share data, to jointly develop and share training programs and other material, or to coordinate project management for a common work project. On an extranet each company uses the security inherent in its own intranet to the keep employees of other companies from accessing information they do not need to see.

The collaborative business application is a powerful extranet tool. Such applications—possibly developed jointly by participating companies—enable the employees of these companies to work together effectively without leaving their offices (which might be located in different places across the world).

For example, a consumer company might work with a supply company to connect their intranets and create a supply ordering system so that all employees of the consumer company could order whatever supplies they needed, whenever they needed them, directly from the supply company. The consumer company employees might order supplies by using their Web browsers to look through one or more electronic catalogs that the supply company published on the extranet. The employees might check a box next to each of the items they needed. Different employees might be given different rights to different catalogs so that they could see and order only from selected parts of a catalog. Underlying parts of the collaborative business application could sort all ordered items by company division, group, and employee, and fill out one daily purchase requisition containing all items ordered by all employees. Each purchase requisition could be immediately delivered over the extranet. For the supply company the application could automatically generate a shipping ticket that contained the items to be shipped, broken down by division, group, and the person to whom each item was to be delivered.

For the consumer company the end result might be the elimination of stocked supplies and a considerable reduction in purchasing costs. The consumer-company employees might get the supplies they needed in less time than ever before. And the supply company might sell more supplies and deliver them faster with fewer staff members.

Because almost all intranets and extranets will eventually be connected to the Internet, intranet technology should be designed to deal as effectively as possible with the security problems and other problems inherent in the Internet. Therefore, Novell is constantly working on new technologies such as the BorderManager family of services and iChain.

Of course, intranets and extranets need not be connected to the Internet: an intranet may be purely local or, if it is a WAN intranet, the various locations might be connected by means other than the Internet. In addition, an extranet may connect several of these types of intranets without any Internet connection. However, an important technology called virtual private networking enables you to create intranets and extranets using the Internet as a ready-made, low-cost WAN backbone.

Virtual Private Network (VPN)

A VPN is a private WAN that uses the public Internet as a low-cost WAN backbone to transport data between two or more geographically separate sites. By contrast, traditional WANs connections are made by means of dedicated communications equipment and dedicated leased lines (such as T1 lines). Although VPNs may not provide the same data-transfer performance as a dedicated-line WAN, there are some advantages that a VPN has over a dedicated-line WAN that have made VPNs increasingly popular.

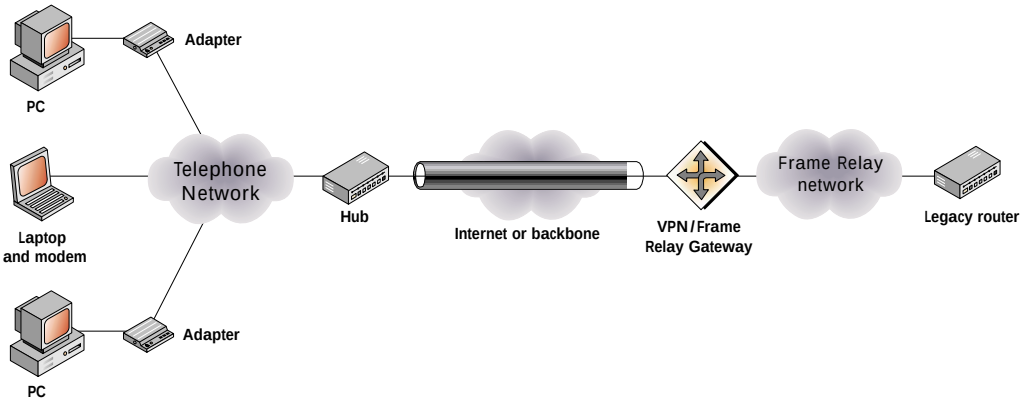


Figure 43

A VPN securely connects remote sites by using the Internet as a WAN backbone.

The most obvious advantage is the cost of implementation. Because you are using the Internet as the backbone, there is no need to lay cable or lease dedicated lines between the remote sites you wish to connect. This eliminates an incredible amount of overhead. Second, because the VPN uses the Internet, the Internet becomes part of your network. In today's computer-based market the Internet is one of the most powerful tools available for large-scale commerce. With a VPN you are preparing for future expansion onto the Internet as the need arises. You can use your Internet connection to make more information available to Internet consumers. With a conventional dedicated-line network, an additional Internet connection would be required to reach this market. Using a VPN, you can network your remote offices into one large WAN and provide access to the Internet.

The technology used in VPNs to connect remote sites is known as "tunneling." Using this technology, you can transfer data across the Internet by encapsulating it into TCP/IP packets and transmitting it across a secure Internet connection referred to as a tunnel.

Tunneling is accomplished through use of a tunneling (or VPN) protocol. Using the tunneling protocol, two sites set up and maintain a trusted session (or tunnel) between them. After the trusted session is established, each site secures data packets by encrypting the contents, encapsulating them in a TCP/IP packet, and sending them through the tunnel. The receiving site extracts the encapsulated packet, decrypts the contents, and routes the information to the appropriate local destination. The most commonly used tunneling protocols are Internet Protocol Security (IPSec) and Point-to-Point Tunneling Protocol (PPTP).

The IPsec protocol operates at the network layer (OSI Layer 3). A core part of IPsec is the subprotocol Internet Security Association Key Management Protocol (ISAKMP)/Okaley, which is the protocol used to establish a secure session. The secure session is based on a shared public key. ISAKMP/Okaley allows the receiving site to obtain a public key and authenticate the sending site using digital certificates. Many vendors currently use IPSec as the basis of their VPN products, and the IETF may adopt IPSec as the standard for a network-layer VPN security protocol.

Like IPsec, PPTP operates by transferring encapsulated data through a secure tunnel. PPTP uses a Generic Routing Encapsulation (GRE) mechanism to encapsulate PPP packets. There are two parallel components of PPTP: (1) a control connection operating over TCP; and (2) an IP tunnel that is used to transport GRE-encapsulated PPP packets. The existing PPTP specification does not detail security other than that addressed in PPP. PPP security includes Password Authentication Protocol (PAP) and Challenge Handshake Authentication Protocol (CHAP). One useful feature of PPTP is the ability to connect IP-based networks with private network addressing schemes: you can connect two networks even if their private addresses conflict with globally unique addresses already registered on the public Internet. PPTP can also authenticate remote clients accessing an intranet site across the Internet.

Network Management

Businesses rely heavily on their computer networks to provide high productivity and to perform many services essential to making a profit. When their networks go down, it usually costs them in loss of productivity and performance, and the larger the company, the greater the loss. For some global companies, network downtime is so critical that it can mean millions of dollars of profit loss for each downtime hour and even, in some cases, for each downtime minute. Hence, it is something of an understatement to say that good network management is important to modern businesses, no matter how large or small their networks.

To appreciate what it means to manage a network, you need to understand that networks, regardless of size, are almost always in a state of growth and change. You must accept that eventually all network components will fail. You must also accept that users can and will make mistakes that damage workstation configurations and server files, and that workstations, printers, and servers will need to be added or removed. In addition, new applications will be added, existing applications will be upgraded, bandwidth requirements will increase, and networking technology itself will change, requiring you to phase-out chunks, segments, and sometimes even the whole of the old networking technology.

Most literature on network management will list five key areas for network managers to focus on, as recommended by ISO:

- Fault management
- Configuration management
- Performance management
- Accounting management
- Security management

Fault Management

Fault management involves quickly identifying network problems and taking steps to isolate them. This includes the proper notification of responsible parties when network components fail. For instance, with the right equipment you can set up your network to send you a pager notice when a network problem occurs. You can even have the system display specific codes on your pager that identify the server or other network component involved.

Configuration Management

Configuration management involves changing network and user configurations in order to optimize network performance and productivity. This is closely tied to fault management, which often employs the technique of changing configurations to isolate network faults.

Performance Management

Performance management involves tracking important network occurrences such as processor and RAM usage levels, disk access requests, usage of specific programs, and data packets being sent and delivered across the network. This data is then used to project future network upgrade requirements as well as to troubleshoot network performance problems.

Accounting Management

Accounting management means tracking and billing network users for the software and other services that they use. This has become a serious issue, with some companies being taken to court and paying fines for not having purchased enough software licenses for the number of users actually using the software.

Security Management

Security management involves protecting your network from unauthorized access to critical business data or system resources. Safeguards must be installed to ensure the integrity of your network. This issue becomes even more complex when you connect your network to the Internet. Without some sort of security your network and all of its data and resources are vulnerable to unauthorized access. Because of the widespread use of computer networks and the increased dependence of most businesses on their networks, network security issues have become paramount. Network management solutions are incorporating a broad range of security features to further ensure the security of computer networks.

Network security technology can be roughly divided into two categories: those that protect against unauthorized access from within (employees or other recognized network users) and those that protect against access from without (hackers, unauthorized Internet intrusions, viruses). Authentication is the most common security measure used to protect networks against access to sensitive information from within. Authentication involves the use of user passwords and rights. When on a network you are given a password that allows you to log on to the network. In addition, you are assigned rights to specific network resources. If you are not given specific rights to a restricted resource, you are not allowed to access the resource.

As the need for higher levels of network security increase, new developments are being made that augment this authentication process. Now there are small, handheld authentication devices, such as ActivCards and Tokens, that act as “password generators” for restricted-access networks. When you wish to access the network, you must enter a personal identification number into the authentication device (which you keep with you at all times). The device will then generate a random, single-usage password that will allow access to the network. This method prevents someone from learning your password and then using it to access your terminal while you are absent.

If your network is connected to the Internet or an extranet, additional measures are required to secure it against virus infections and highly-skilled computer hackers. Virus protection software is crucial to network security. These software programs scan all data entering your network from any outside source for known viruses and warn you of any viruses encountered, so you may avoid corrupting your network software. Updates for virus software are made available through the vendor, usually on a subscription basis. These updates ensure that your virus software will be able to identify new viruses as they are discovered.

Protection against unauthorized access from outside your network is usually provided through some sort of firewall service. Firewalls are either computers or routers that are set up to provide a secured “doorway” through which you can access the Internet and Internet users can access your Web data.

Firewall services can be configured to meet specific security needs. They can be set up to screen Internet users trying to access your network, and to allow only certain authorized employees to access the Internet from within your network. In addition, many firewalls now feature remote authorization for employees using a remote (off-site) Internet connection to access restricted network resources. Other non-Internet applications for firewall services include protecting mainframes or subnetworks from general access within an organization and ensuring confidentiality of data transmitted across networks.

All of these aspects of network management are crucial to the continued success of your network. Managing these aspects, however, is a daunting task that becomes more difficult as your network grows and evolves. Luckily, the computer industry is aware of the importance of network management and is constantly developing new products to assist in management tasks. The most important of these recent developments is directory services. The following section explains directory services and how they are revolutionizing the complex task of computer network management.

Directory Services

As networks increase in size and diversity and become more complex, the administration of these networks becomes increasingly difficult. Our modern network environments often include a variety of hardware and software. Users often require multiple passwords and varying levels of access and authority—all of which must be entered in several locations across the network. And if any change occurs in the user's status, this information must again be modified at each of these locations. Heterogeneous environments, Internet access, and the security issues involved with each further compound the problem.

A solution for this increasingly difficult task of network and information management can be found in directory services. Directory services provide you with the capabilities to manage your entire network—regardless of size, operating system, or complexity—from a single location. With directory services, user information is entered once and then automatically applied across the entire network. E-mail addresses, group memberships, access rights, and heterogeneous operating system accounts are created automatically. Likewise, any changes to user or resource information are automatically updated throughout the network. Administrators no longer need worry about the security issues involved with the termination of employees. Once the user's profile is removed, all related access and authority is immediately revoked.

Currently, most directory services are based on the X.500 directory standard and, more specifically, Lightweight Directory Access Protocol (LDAP), the protocol used to access directory information. Due to its extensive use in TCP/IP-based networks, LDAP is rapidly becoming the standard for directory service access and directory-enabled applications.

Directory Basics

Directories are similar to databases in that they organize information into records and fields. Below is a table from a sample database:

UserID	Last Name	First Name	Password	Telephone	Cubicle	Title
beauliet	Beaulieu	Trace	******	7-6047	C43	Graphic Artist
boehma	Boehm	Amber	******	5-0972	A21	Manager
canninga	Canning	Amy	******	7-5436	D32	Programmer
dunbarp	Dunbar	Paul	******	7-5486	A45	Data Entry
lopezi	Lopez	Ignacio	******	5-7824	C45	Editor
mayb	May	Bonnie	******	5-4554	C23	Designer
shankarn	Shankar	Naren	******	7-6584	D45	Programmer
whitehm	Whitehead	Marcia	******	5-4962	C67	Manager

In this table each row constitutes a record and each column is a field. A relational database would consist of two or more such tables in which the field of one table would correspond to the field of another. For example, the "cubicle" field in the above table might correspond to the cubicle field in another table with fields such as "size," "floor," "workstation type," and "printer type." (This second table would keep track of cubicle location, dimensions, the equipment each contains, etc.) Relational databases work well for organizing complex data relationships, but the directory can go one step further: it can organize information into a hierarchy.

Consider the fictional company, Networking, Inc. It has offices in four locations: Tokyo, New York City, Albany, N.Y., and London. Each location houses different departments; for example, Sales and Marketing are in the Albany office and Testing and Production are in the London office. Networking, Inc. has organized its network directory according to the company hierarchy shown in Figure 44.

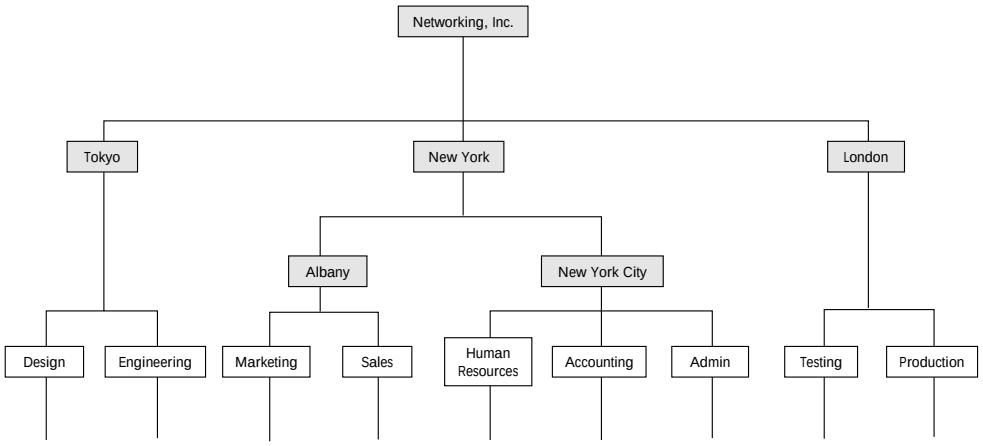


Figure 44

Hierarchical organization of a fictional company

In the directory the network is depicted as a series of "objects," which are virtual representations of network components and organizational elements.

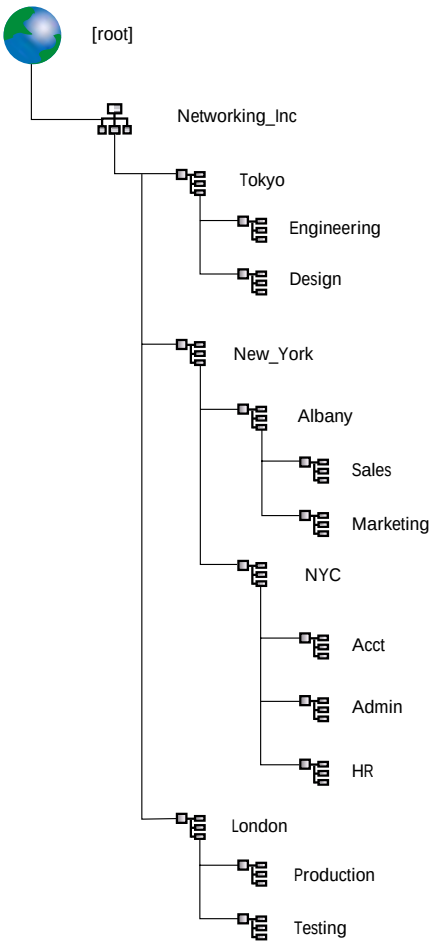


Figure 45

The directory tree

The above diagram shows a directory "tree," so called because the hierarchical organization resembles an upside-down tree with the "root" object at the top and the branches extending downward. The object labelled "Networking_Inc" is an organization object, and the rest of the icons represent "organizational units" (OUs). The root, organization, and OU objects are all "container objects," meaning that they can hold other objects. In the above example, the root object holds all the other objects, and the New York object holds the Albany and NYC objects and all their departmental OU objects. Each of these container objects may contain other objects that represent other containers, servers, volumes, applications, users, printers, or other network components. Objects that cannot contain other objects, such as printers or users, are called "leaf" objects.

The advantage of using container objects is that the changes you make to the container will affect all the objects in the container. For example, if you wanted to establish a server policy for all the servers in New York, you could assign the policy to the New_York OU and the policy would automatically "flow down" to all the servers contained in that OU. This eliminates having to assign the policy to each server individually. If you did not want a particular server to be affected by those changes you could select that server's object and designate it an exception.

Objects in the directory are comparable to records in a database, and the fields are called, collectively, the "schema." Because the nature of each object is different, the schema for each object type is different. For example, a printer object could contain fields in its schema for its make, model, speed, resolution, and IP address. A user object's schema could contain the fields in the example database table and many others. When fields can be added to a schema as needed, it is called an "extensible schema."

In a hierarchically organized directory, the name of each object shows where it fits in the hierarchy. For example, if the employee named Ignacio Lopez works in Sales, his directory "name space" would be lopezi.Sales.Albany.New_York.Networking_Inc. Likewise, the name space for Trace Beaulieu in Design would be beauliet.Design.Tokyo.Networking_Inc. Every object in the directory has one of these hierarchical name spaces.

The opposite of a hierarchical name space is the "flat" name space, which would be the user name only: lopezi or beauliet. In such a case, there could not be another lopezi or beauliet in the entire directory; each user name would have to be unique. This presents a problem if, for example, there are several HP LaserJet5si printers on the network. But with a hierarchical name space there is no problem. By including the full "context" in each name space, the directory can easily recognize the difference between hplj5si.Engineering.Tokyo.Networking_Inc and hplj5si.HR.NYC.New_York.Networking_Inc.

Security and Authentication

A primary function of the directory is to manage authentication and network security. A typical network consists of components that need to be available to all users as well as those that should be available only to a few. For example, everyone may need access to a particular printer, but the payroll application should be accessible only to authorized users. Without directory-enabled authentication, access to each restricted component would require a separate password. In such a case, you would be forced to remember a different password for each restricted-access application, server, or other component, which would result in a network administrator's nightmare: constant calls for help when a password is forgotten, or worse, passwords taped onto monitors for all to see.

Directories solve this problem by providing "single sign-on." You log on to the network once with one password, and access to network components is controlled with information in the components' schema. For example, the schema for a server in Human Resources (HR) might indicate that only users in the Human Resources OU have the right to access it. If Marcia Whitehead in Production tries to log on to the HR server, the directory checks the server's schema to see if she is included in its access control list (ACL). Because no one in Production has rights she is denied access. The ACL is similar to a bouncer guarding the entrance to an exclusive club: if your name is not on his clipboard, you don't get in.

Directories allow access rights to be assigned on a large scale or on a very small scale. For example, if network administrators want to grant everyone in the company rights to print on a particular printer, they can drag the printer's object onto the Networking_Inc object and set the rights in one easy step. Also, they can assign rights with an extremely fine degree of granularity, such as determining that while a half-dozen users can see the contents of the personnel database, only one user can alter them.

This method of granting access rights makes for powerful security: if unauthorized users try to access the payroll application, for example, the directory will prevent them from doing so because such rights would not be listed in the application's ACL. They would also not have the ability to alter their access rights, making it next to impossible for them to access anything they should not.

An additional advantage of using a directory for authentication is that because the login information is stored centrally in the directory, you can log on to the network from any workstation on the network. With proper configuration, you can even log on through the Internet, which is extremely useful when you are away from your office.

Network Management

Directories also form the basis for improved network management. Just as access rights can be granted with a fine degree of control, administrative rights can also be granted on any scale. If Bonnie May needs control over every aspect of the network, her icon is dragged onto the root object and rights are granted (by someone else with similar rights). If network administrators do not have time to make changes to telephone numbers and addresses, they can grant rights to alter only those fields to one or more administrative assistants.

Principal among the advantages of using a directory to manage a network is that administrative control can be centralized. With directory-enabled management tools such as Novell's NetWare Administrator or ConsoleOne, network administrators can see and manage the entire network from one location. Instead of hiring four full complements of network administrators for each office, for example, Networking, Inc. could hire only one.

Although the directory enables management from a central location, a well-deployed directory does not itself reside in any one place. If the entire directory resided on a single server, the directory would be extremely vulnerable to failure: if that server went down, essential services provided by the directory such as authentication, security, and management would disappear.

For this reason, well-designed directories can be replicated or copied across the network to provide fault tolerance. For example, if in each of Networking, Inc.'s departmental OUs there were two servers, a copy of the directory could reside on each one. If one server went down, the directory's services would be provided by back-up copies on the other servers.

However, it is probably not practical to house a copy of the entire directory on each server. Not only would the directory take up too much disk space, it would also increase server traffic and reduce server performance. It is better to "partition" the directory along logical boundaries and store replicas of the partitions on different servers.

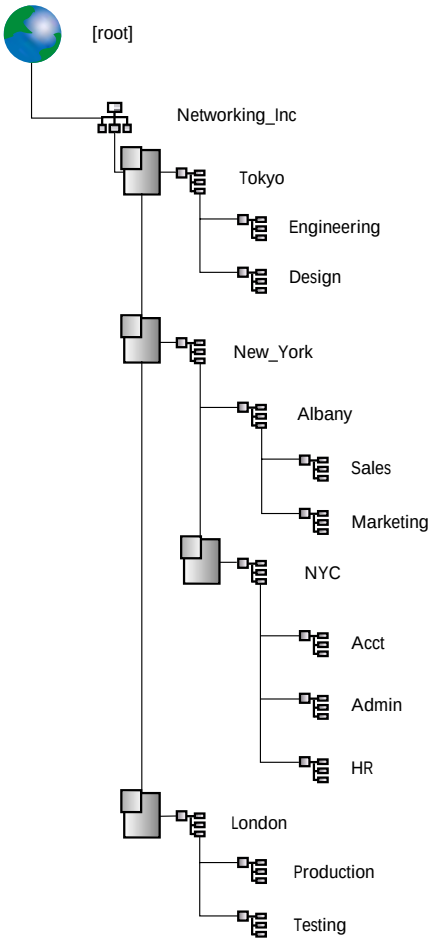


Figure 46

Directory partitions

Figure 46 shows the directory partitioned into four segments. One holds the Tokyo OU and its subordinates. The second holds all the objects in Albany, the third holds only the NYC OU, and the fourth consists of everything in the London OU. To alleviate server traffic, only a partition of the directory would be housed in each location: the London partition would reside on the London servers only, for example, and the Tokyo partition would reside on the Tokyo servers. In this arrangement, each directory partition would service only those objects that are physically closest to it, thereby reducing directory response time. Additionally, you would authenticate to the directory through the partition closest to you instead of authenticating across a slow or expensive WAN connection. And even with the directory divided and distributed across multiple servers, network administrators can still view and manage the directory as a single unit.

The Directory and E-Business

The security and management capabilities of directories have been exploited for quite some time, but the full utility of directories as data stores is only beginning to be explored. The hierarchical arrangement, extensible schema, and fine degree of control make directories the ideal foundation for applications that require flexible methods to store and organize data. A large number of these applications are "e-business" or electronic business applications. E-business is the practice of conducting traditional business processes by electronic means, often using an intranet or the Internet. These processes include customer service, electronic store fronts, employee provisioning, supply-chain management, and other kinds of collaboration between businesses.

To demonstrate how a directory can make business processes more efficient, consider the following scenario. New employees hired by Networking, Inc. need to be "provisioned" or supplied with what their new job requires—workstations, network passwords, security clearance—and the company needs to put the employees' personal information into their various systems—payroll, insurance, 401K. Without a directory-enabled system, the employee must visit each department one by one to obtain the necessary supplies and must fill out a multitude of forms that will be manually keyed in by each department into its own separate system. The risk of human error is high, and when the employee's information changes, each department must be notified separately, further increasing the likelihood that some departments will have outdated information.

With a directory-enabled provisioning application the employee's information is entered once into the computer and stored in the directory. Each department's application would have access to that single data store. As illustrated in Figure 47, a directory-enabled process ensures that employees can be provisioned quickly and efficiently—even before they walk in the door their first day.

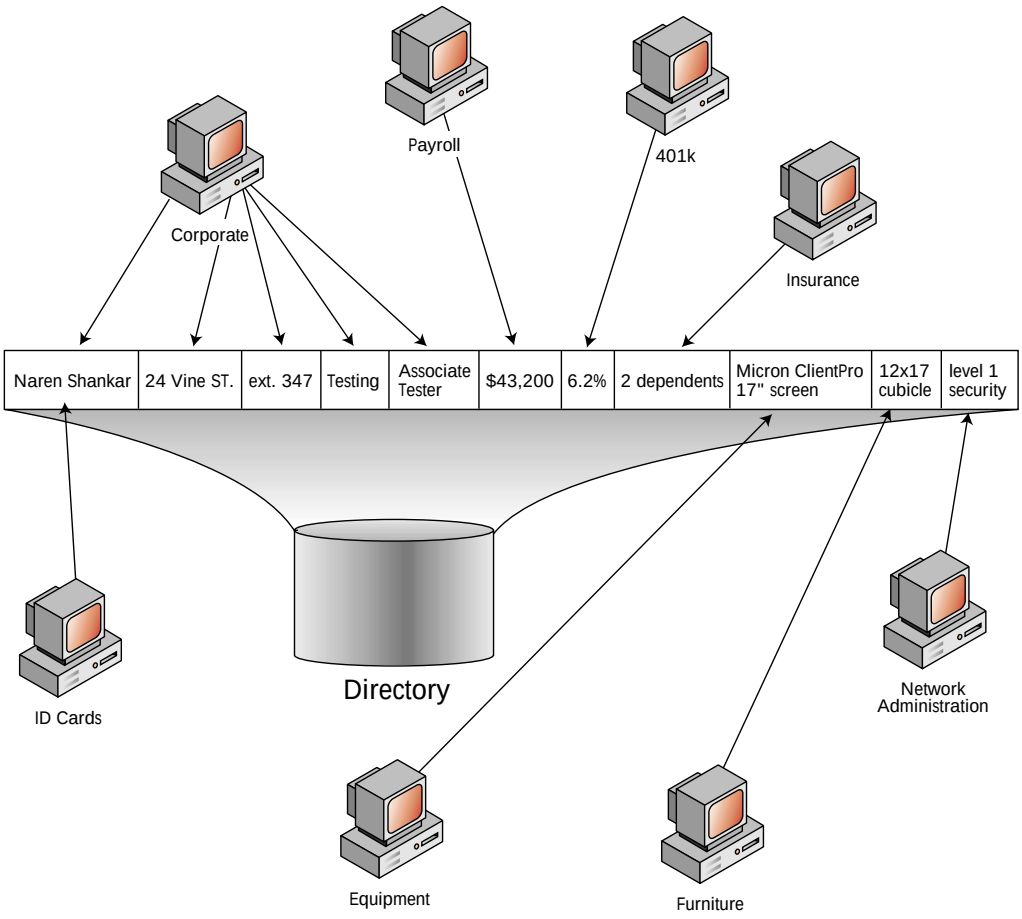


Figure 47

Directory-enabled provisioning software simplifies the provisioning process.

The figure shows a directory entry for Naren Shankar. Each of the computer icons represents departments that need to know information about him so that he can be properly provisioned. For example, Payroll would need to know his salary, Equipment would need to know what kind of workstation he will use, and Network Administration would need to know at which level he should be granted access rights. The arrows in Figure 47 point to the information that would be of special interest to each department, but in reality the departments would have access to several fields such as name and phone extension. It should be noted that none of the departments would have or need access to all the information: for example, Network Administration would not be able access to salary information, and Insurance would not require security clearance information. Just as a directory can grant and restrict user access rights, it can also restrict which fields applications access. And when Naren leaves the company, his information will be deleted only once and the change will take effect across the entire network.

Novell Directory Services (NDS) is the best directory on the market today. NDS eDirectory, Novell's latest directory product, represents the culmination of more than a decade of development. Based on X.500 and LDAP, NDS has become the de facto standard for directories. eDirectory is highly scalable, which means it can hold more than one billion objects. eDirectory is integrated with the NetWare operating system, and with Novell Account Management you can use eDirectory to manage other NOSs such as Windows 2000, Windows NT, Solaris, and Linux.

Information Management

Even networks that are not managed by a central directory nevertheless contain many directories. Many applications such as e-mail applications create directories specific to their needs. Most companies also maintain one or more databases that often contain overlapping or redundant information; for example, several directories might store a person's address and telephone number, or product prices would appear in one or more databases. Keeping the information in these directories and databases up-to-date can occupy inordinate amounts of time and effort, and even the best efforts cannot prevent incorrect information from circulating.

The best solution would be to create one huge directory that contains every scrap of information used by every last application and every single department; however, unless you are building your company from the ground up, starting today, this approach is as impractical as it is expensive. The next best thing is to synchronize the directories, or in other words, connect them in such a way that when information is changed in one directory, the change is reflected in all the directories.

The latest technology to provide this capability is XML, which stands for eXtensible Markup Language. A language similar to HyperText Markup Language (HTML), XML enables translation between incompatible file formats. Novell has developed a solution called DirXML that uses XML to create "metadirectories"—directories that function as a single directory but that in fact are made up of several otherwise incompatible directories or databases. A metadirectory enables synchronization between directories so that when a change is made to one, it is made to all. It also makes it possible to search a number of directories at once from a single interface.

For more information about directories and DirXML, please visit the Novell World Wide Web site at http://www.novell.com.

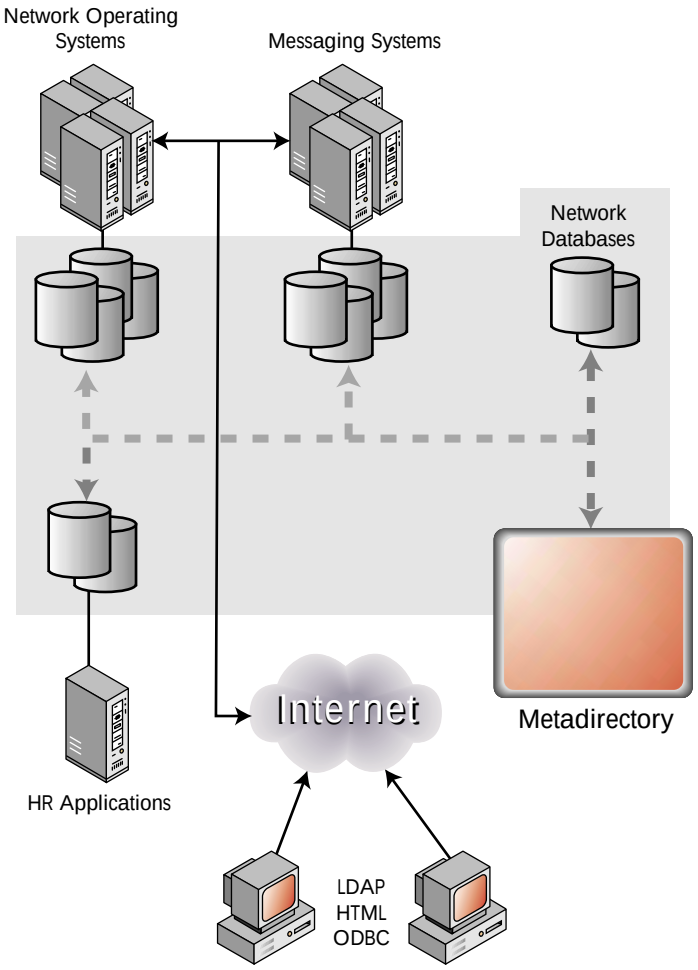


Figure 48

A metadirectory connects users to information stored on any network database.

Choosing a Network Implementation

Before designing a network, the complete assessment of a company's networking needs is in order. The tasks that will need to be automated or otherwise made more efficient must be identified, as must the business applications that are currently supported and those that are being considered for the future. Does the company need to provide shared access to word processing files, or does it have multi-user databases to support? Is electronic mail a necessity? What type of Web server and platform combination will best service the company's Web site requirements?

Once all of the current business tasks and functions that the company expects to support have been determined, and the best guess regarding future requirements has been made, prioritization is the next step. As the networking plan is deployed, the company should consider which parts of that plan—such as those that impact critical business functions—should be implemented immediately and which can be addressed at a later time. This process will allow the company to diffuse its expenditures over time and also give employees sufficient time to adapt to an updated networking environment.