Com 226 DATA COMMUNICATIONS

CHAPTER ONE

Computer Communications Networks

Introduction

The purpose of a computer communications network is to allow moving information from one point to another inside the network. The information could be stored on a device, such as a personal computer in the network, it could be generated live outside the network, such as speech, or could be generated by a process on another piece of information, such as automatic sales transactions at the end of a business day. The device does not necessarily have to be a computer; it could be a hard disk, a camera or even a printer on the network. Due to a large variety of information to be moved, and due to the fact that each type of information has its own conditions for intelligibility, the computer network has evolved into a highly complex system. Specialized knowledge from many areas of science and engineering goes into the design of networks. It is practically impossible for a single area of science or engineering to be entirely responsible for the design of all the components. Therefore, a study of computer networks branches into many areas as we go up from fundamentals to the advanced levels. Advancements in communication of speech have long been matured in the form of public switched telephone network (PSTN). However, design of store-and-forward type of networks, such as the Internet, is far from matured - perhaps due to proliferation of the ways in which such networks are used. The integration of the two types of networks is the culmination of telecommunications technology. It is not futuristic to imagine telecommunications networks meeting the needs of live traffic (e.g., phone calls) as well as store-and-forward data (e.g., email) traffic according to the desired quality of service. In this chapter, we look at a computer network as a whole, from both an application point of view and a design point of view. In fact, the design and application influence each other so much that a study of the fundamentals is practically impossible by leaving either one out.

Main Components

As described above, a computer network is composed of a number of independent components. Three main components are:

Data Networks There is a myriad of terms used to describe a computer communications network. Computer Networks, Networks, Communications Networks, Telecommunications Networks, Packet Switched Networks, Networking Systems and Data Networking Systems are all among the terms used. The terms data network used to be more descriptive when live speech and video could not be transmitted over such networks. At that time, data meant store and forward type of information that had no real-time content. Examples of such information are file, email and logon programs. With time, definition of data has changed , and so has the definition of data networks. A data network now includes all of the above. The telephone network could be seen as an exception, even though it is also has been used as data network in many of its applications. However, due to its circuit-switched nature, it is not projected to belong to computer networks.

The Computer System

Computer systems are stand-alone systems, along with peripheral devices, capable of performing information input, output, storage and processing. The study and design of computer systems is the

job of computer scientists and engineers. Computer systems usually consist of hardware (processor, memory, storage devices and input and output devices), system software for user interface and resource management, such as operating system and special purpose software such as programming languages, database management system, text-processing systems etc. Developments in microchip have led to the utilization of processor technology in everyday appliances, making all networkable devices operating like a computer system. Examples of computer systems are: personal computers, notebook computers, and data acquisition systems.

The Communications System

The communications systems provide a vehicle of carrying information from one point to another by conditioning it appropriately. The conditioning may include changing the actual shape of the information, or even adding to and removing parts of it. Example of changing the actual shape of the information is in speech communication devices that take speech signal in the form of mechanical energy and generate an equivalent electrical signal suitable for transmission media. Examples of adding to the information is error control coding in which extra information is added in order to combat errors that might have entered the information during its movement inside the network. Example of removing information is data compression in which the size of the information is reduced yet preserving the amount of intelligence it represents. The challenges in designing a communications system relate to the efficient usage of available network resources (bandwidth, etc.), reliable communication in the wake of channel noise, and special purpose requirements owing to applications that generate information or are the users of the information (e.g., security) or other conditions (e.g., wireless, underwater). A device known as MODEM (Modulator/demodulator) is an example of a communications system. In designing a MODEM for telephone line, the main challenge comes in utilizing the limited telephone bandwidth to transfer information at a maximum possible rate. The job of communication system design lies with the communications engineer. A communications engineer has to study the characteristics of information, the channel and the environment in order to design a system to meet specified performance criteria.

Some of the challenges faced by a network engineer include the efficient use of communications link (using, for example, multiplexing), study of the characteristics of the information to be exchanged and its peculiar requirements of timing and bandwidth. Networks are designed to share communications resources and network engineer designs switching mechanisms for this purpose. Because of link sharing, security of information becomes very critical in networking systems. In this way, solving one problem raises another. The area of network engineering started as a conglomerate of computer science and communications engineering. However, it has fully grown into a field of knowledge by itself. Sometimes it means different things to different types of people involved - users, providers, and designers.

Communication Systems Versus Networking Systems

The communication systems and networking systems are two different fields of study altogether. Sometimes, confusion may arise as to if the design of a component of a computer network is part of the networking system or the communications system due to close interaction of the two. Let's look at two examples of such components: one relating to a hardware component and the other relating to a function. A hardware example is the MODEM. A MODEM design is the job of communications engineer and not the network engineer. The network engineer is a user of modem. An example of a function is error recovery. A communications system is design to be robust in the presence of random errors that may interfere with electrical energy signals. The receiver design employs efficient signal detection mechanisms to minimize the probability of error. The communication

engineer usually adds additional hardware or mechanism called a CODEC (from Coder/Decoder) to substantiate recovery from errors. A network engineer can add other layers of coding or use different methods for error control, such as, retransmission of information with errors. However, the concern of a network engineer is not at signal detection, but after the signals have been received. In this way, the end product may have many layers of error control, as part of both the networking and communication systems.

Network Development Example

In this section, we will consider an example of a computer from a user point of view. Later, we will see how these requirements of user translate to a language suitable for the designer of such a network. Consider a multinational organization with offices and personnel computing needs in the following hierarchy.

1. Every employee (or at least office) needs to have a computer on every location.
2. Many people are working in multistory buildings at each location.
3. The company has many locations in metropolitan areas.
4. A large number of transactions with other national/international locations are carried out on a regular basis.
5. All computers must have some general-purpose software.
6. Many select computers require special purpose software.
7. Software and hardware sharing on each floor is desirable.
8. Software sharing among all floors in each building is desirable.
9. Some software sharing among all locations at the end-of-business-day is a requirement.
10. Some sharing in locations nationwide is desirable.
11. Transaction capability among international locations and with other businesses is also desirable.

Three Role Players

In order to solve the networking problem for the above business, three types of services/staff are involved. These are: user with the information technologists (IT staff), the network provider and the network vendor/designer.

User/IT Staff: The business in question is the user in this example. The IT staff is a permanent staff closely aware of the business needs and is expert of the available software for special and general purpose. The IT staff provides what is called the information system services.

Network Provider: Another company will provide those networking services that are not required to be owned or/and are too high level, technically, to maintain by the IT staff. The network providers may not design their own systems, in general - they could be simply a carrier or operator company. Instead, the system design is a separate task not related to the business organization directly.
Network Designers/Vendors: A variety of equipment and services vendors may be involved directly or indirectly. Some may not directly interact with the business organization. Such a vendor designs and manufactures equipment to be operated by user and network provider.

Network Design

With help from the IT staff and network provider, the company may end up with the hierarchy of networking systems shown in Figure 1-1. The double-sided arrows show a bi-directional communications capability for sending and receiving information. To each user, the network connection should look transparent and direct with all networking levels, as shown in Figure 1-2.

This transparency of intermediate networks is a very important issue in the design of data communication networks. It is taken care by many layers of software and hardware.

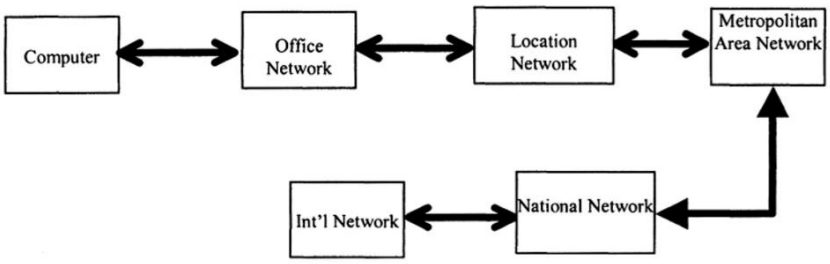


Figure 1-1. The hierarchy of networks to meet all the business needs for the example in section 1.2.

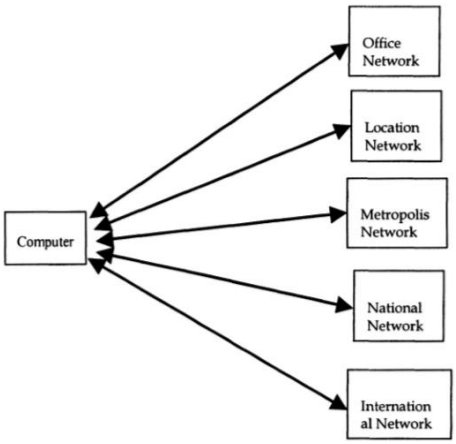


Figure 1-2. The logical connections between a user and the hierarchy of networks

User/IT Staff

User must own or be the sole controller of all the equipment responsible for internetworking of the computers and other devices within a floor, or even the whole building. This is for the security purposes as well as for the fact that things may change from time to time and computers and software may need updating frequently. Therefore, the challenging task for the user is to decide from which vendor to make purchases so that the current and foreseeable, or otherwise future, needs may also be incorporated in the network design. The IT staff is hired for this purpose.

Network Provider

The network connection from building to building, nationally and internationally may be too expensive and inefficient for the user to own. Many companies provide such connections on subscription or leasing bases. The challenge for the network provider is to provide some guarantee of a reliable service in a cost effective manner.

Network Designer/Vendor

The network vendor designs the equipment as required by network provider and user for a diverse user base. The challenge for the designer is to have ready-made designs to meet most or all of the current and near future needs. The research and development department of the vendor must be able to predict future needs and, possibly, prototype future equipment.

Standardization

Through the process of standardization of networking hardware and software, it is possible for the vendor to design equipment without consulting the network provider – such consultation is done during the standardization process. Standards for networks encourage competition among vendors by allowing enhanced services. They provide user with the chance to shop around every time a new need of software, hardware or service arises. Most of all, it allows for interoperability, the capability of the equipment by one vendor to interact with the equipment designed by another vendor.

Circuit Switching

In this switching mechanism, a circuit is allocated to every piece of complete information (called a call). This circuit allocation is all the way from the sending to the receiving computer or terminal. It stays in place throughout the duration of the call until the sending (or receiving) side signals that it is not needed any more. In more formal terms, we say that a fixed bandwidth is guaranteed throughout the communication session. Circuit switching can be used for voice or file communication. However, it is easily seen that for file communication, it is best to send one part of file at a time. These chunk or data blocks are called packets. Each packet of the file may be transmitted via the same or different route. This allows for the number of additional functions and procedures that can be performed on each packet. Moreover, it arises a new type of switching, called packet switching for obvious reasons.

Packet Switching

In this type of switching, data is broken into smaller data units, called packets. Inside the network, each packet may be treated as if it were a small, complete message. The bandwidth can either be guaranteed for all packets or not. It is best suited for file transfer. However, if enough guarantees can be provided about the inter-gap times of voice signals, it can be used for voice as well (called packet voice).

Classification of Networks

There are several ways of classifying data networks, such as, geographic scopes, protocol architectures and type of service. Following is a classification based on (roughly) the geographical scope.

Local Area Networks (LANs)

LANs are (usually) small networks that provide a high-speed physical and logical connection among a group of stations. They typically encompass a walk-able geographic area, owned and administered by the user and are mainly used either for hardware sharing or as access networks for greater geographical scale. Most commonly used LAN is the Ethernet.

Wide Area Networks (WANs)

WANs cover a general geographical area that may vary from a small office area to the whole world (or even more!). Usually, network providers and big businesses own such networks. WANs are mostly heterogeneous, meaning, a large variety of LANs and equipment or other WANs can constitute a single WAN. An example of a WAN is the Internet. Internet spans much of the populated world, is administered by different groups at different locations, and has many other WANs as part of it.

Metropolitan Area Networks (MANs)

MANs are networks between a LAN and WAN. They are a type of interconnecting networks for big businesses in a metropolitan area. Usually, they have interconnecting (switching) devices instead of user desktop computers as their nodes, but it is possible to have user computers directly attached to a MAN.

Network Protocol Architecture

In addition to classifying a network as LAN, MAN or WAN, there is a structured terminology to describe and identify various parts of the hardware and software making up a computer communication network. Three most important terms of this terminology are protocols, standards and network architecture.

Protocols

Protocols are rules of communication. It is through protocols that computers can exchange information. Just like humans obey certain rules of communications, so must the computers. Computers are specific about rules and cannot guess like humans. They have protocols as part of their software or hardware interaction and can't change that unless the software or hardware is changed or modified.

Standards

Standards are the protocols that have gone through a standardization process. They are documented by some agency or organization so that a large number of vendors can get those documents and design systems based on the same protocols. This takes care of the interoperability issue and helps both vendors and users. Examples of standardization agencies are; the Internet Society, International Organization for Standardization (ISO), Institute of Electrical and Electronic Engineers (IEEE) and American National Standards Institute (ANSI), European Telecommunications Standards Institute (ETSI) and International Telecommunications Union (ITU).

Layered Tasks

To reduce the design complexity, most of the networks are organized as a series of layers or levels, each one build upon one below it. The basic idea of a layered architecture is to divide the design into small pieces. Each layer adds to the services provided by the lower layers in such a manner that the highest layer is provided a full set of services to manage communications and run the applications. The benefits of the layered models are modularity and clear interfaces, i.e. open architecture and comparability between the different providers' components. A basic principle is to ensure independence of layers by defining services provided by each layer to the next higher layer without defining how the services are to be performed. This permits changes in a layer without affecting other layers. The basic elements of a layered model are services, protocols and interfaces. A service is a set of actions that a layer offers to another (higher) layer. Protocol is a set of rules that a layer uses to exchange information with a peer entity. These rules concern both the contents and the order of the messages used. Between the layers service interfaces are defined. The messages from one layer to another are sent through those interfaces.

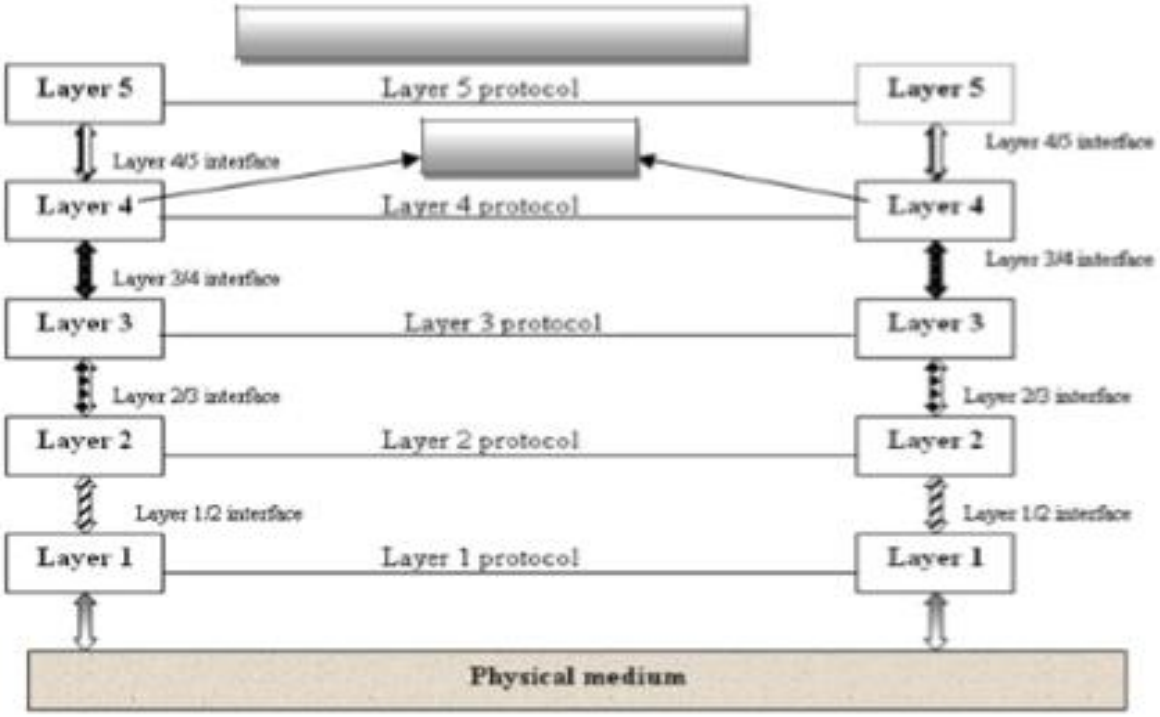


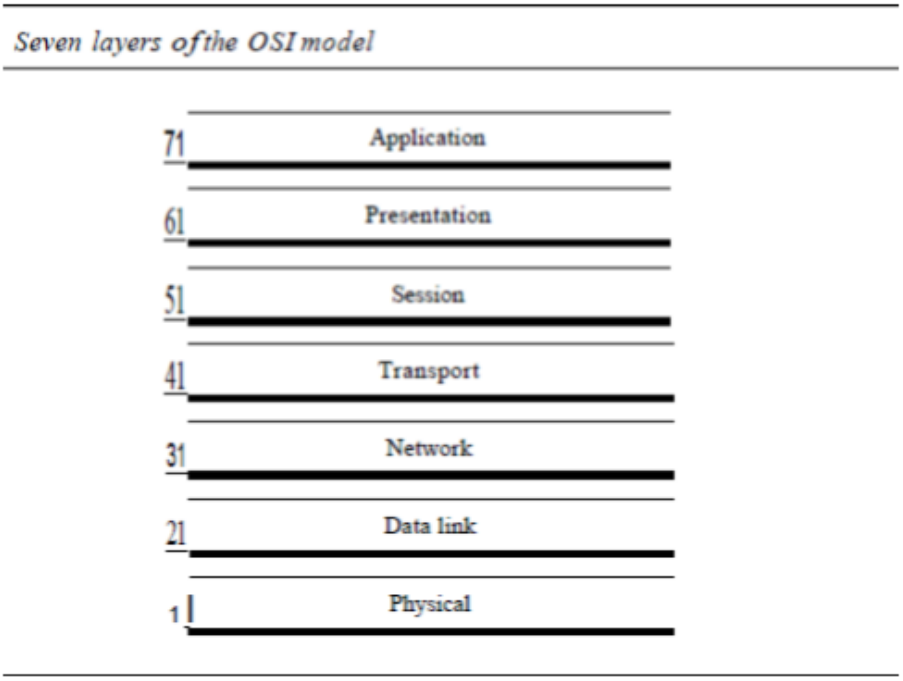
Figure: Layered Architecture

In a n-layer architecture, layer n on one machine carries on conversation with the layer n on other machine. The rules and conventions used in this conversation are collectively known as the layer-n protocol. Basically, a protocol is an agreement between the communicating parties on how communication is to proceed. Five-layer architecture is shown above; the entities comprising the corresponding layers on different machines are called peers. In other words, it is the peers that communicate using protocols. In reality, no data is transferred from layer n on one machine to layer n of another machine. Instead, each layer passes data and control information to the layer immediately below it, until the lowest layer is reached. Below layer-1 is the physical layer through which actual communication occurs. With layered architectures, communications between two corresponding layers requires a unit of data called a protocol data unit (PDU). A PDU can be a header added at the beginning of a message or a trailer appended to the end of a message. Data flows downward through the layers in the source system and upwards at the destination address. As data passes from one layer into another, headers and trailers are added and removed from the PDU. This process of adding or removing PDU information is called encapsulation/decapsulation. Between each pair of adjacent layers there is an interface. The interface defines which primitives operations and services the lower layer offers to the upper layer adjacent to it. A set of layers and protocols is known as network architecture. A list of protocols used by a certain system, one protocol per layer, is called protocol stack.

OSI MODEL

The OSI model is based on a proposal developed by the International Standards Organization (ISO) as a first step toward international standardization of the protocols used in the various layers (Day and Zimmermann, 1983). It was revised in 1995(Day, 1995). The model is called the ISO-OSI (Open Systems Interconnection) Reference Model because it deals with connecting open systems—that is, systems that are open for communication with other systems. The OSI model is a layered framework

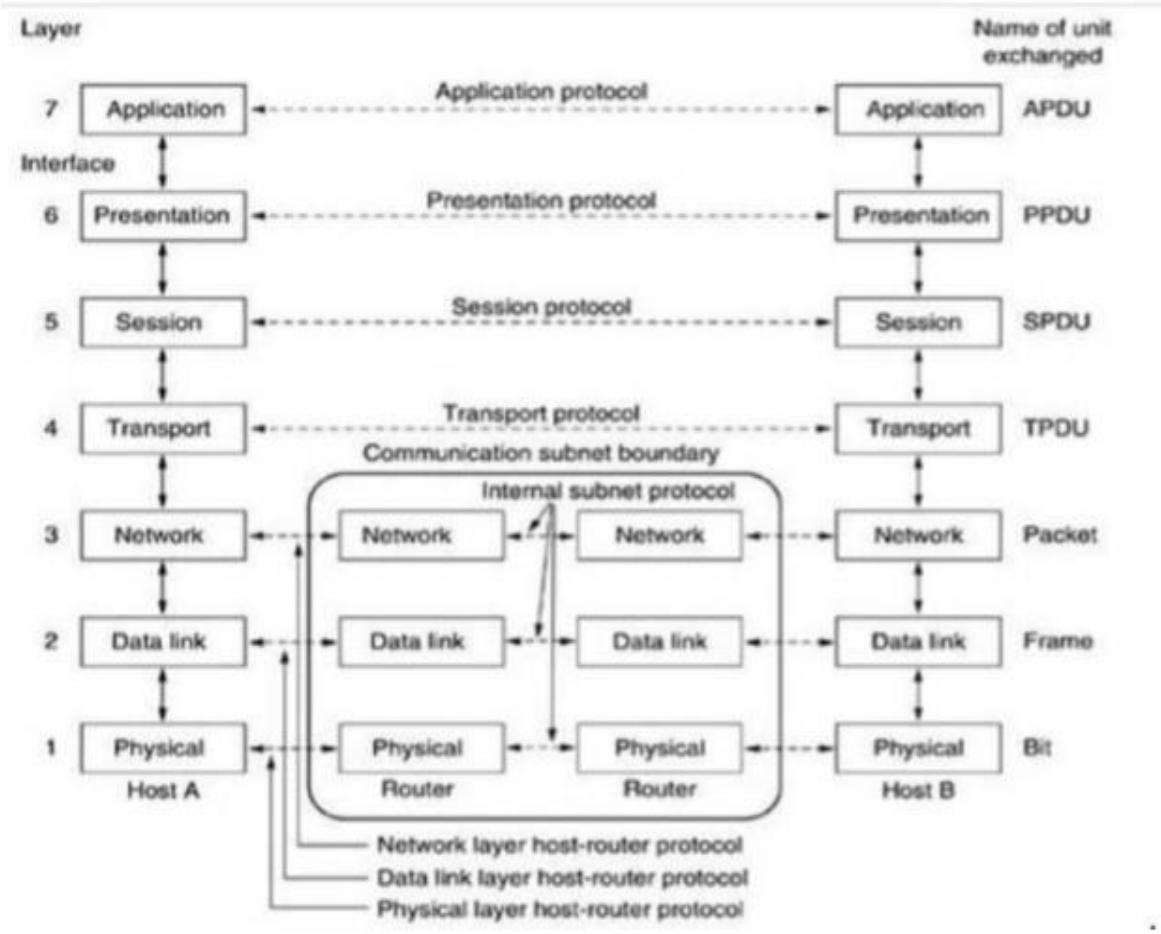
for the design of network systems that allows communication between all types of computer systems. It consists of seven separate but related layers, each of which defines a part of the process of moving information across a network.



The OSI model is composed of seven ordered layers: physical (layer 1), data link (layer 2), network (layer 3), transport (layer 4), session (layer 5), presentation (layer 6), and application (layer 7). Figure below shows the layers involved when a message is sent from device A to device B. As the message travels from A to B, it may pass through many intermediate nodes. These intermediate nodes usually involve only the first three layers of the OSI model.

1. **Physical Layer**

   The physical layer coordinates the functions required to carry a bit stream over a physical medium. It deals with the mechanical and electrical specifications of the interface and transmission medium. It also defines the procedures and functions that physical devices and interfaces have to perform for transmission to occur.



The physical layer is also concerned with the following:

- **Physical characteristics of interfaces and medium**. The physical layer defines the characteristics of the interface between the devices and the transmission medium. It also defines the type of transmission medium.
- **Representation of bits**. The physical layer data consists of a stream of bits (sequence of Os or 1s) with no interpretation. To be transmitted, bits must be encoded into signals--electrical or optical. The physical layer defines the type of encoding.
- **Data rate**. The transmission rate-the number of bits sent each second-is also defined by the physical layer. In other words, the physical layer defines the duration of a bit, which is how long it lasts.
- **Synchronization of bits.** The sender and receiver not only must use the same bit rate but also must be synchronized at the bit level. In other words, the sender and the receiver clocks must be synchronized.

  Line configuration. The physical layer is concerned with the connection of devices to the media. In a point-to-point configuration, two devices are connected through a dedicated link. In a multipoint configuration, a link is shared among several devices.

  Physical topology. The physical topology defines how devices are connected to make a network. Devices can be connected by using a mesh topology (every device is connected to every other device), a star topology (devices are

connected through a central device), a ring topology (each device is connected to the next, forming a ring), a bus topology (every device is on a common link), or a hybrid topology (this is a combination of two or more topologies).

- **Transmission mode**. The physical layer also defines the direction of transmission between two devices: simplex, half-duplex, or full-duplex. In simplex mode, only one device can send; the other can only receive. The simplex mode is a one-way communication. In the half-duplex mode, two devices can send and receive, but not at the same time. In a full-duplex (or simply duplex) mode, two devices can send and receive at the same time.

2. Data Link Layer

The data link layer transforms the physical layer, a raw transmission facility, to a reliable link. It makes the physical layer appear error-free to the upper layer (network layer). Other responsibilities of the data link layer include the following:

Framing. The data link layer divides the stream of bits received from the network layer into manageable data units called frames.

Physical addressing. If frames are to be distributed to different systems on the network, the data link layer adds a header to the frame to define the sender and/or receiver of the frame. If the frame is intended for a system outside the sender's network, the receiver address is the address of the device that connects the network to the next one.

Flow control. If the rate at which the data are absorbed by the receiver is less than the rate at which data are produced in the sender, the data link layer imposes a flow control mechanism to avoid overwhelming the receiver.

Error control. The data link layer adds reliability to the physical layer by adding mechanisms to detect and retransmit damaged or lost frames. It also uses a mechanism to recognize duplicate frames. Error control is normally achieved through a trailer added to the end of the frame.

Access control. When two or more devices are connected to the same link, data link layer protocols are necessary to determine which device has control over the link at any given time.

3 **Network Layer** The network layer is responsible for the source-to-destination delivery of a packet, possibly across multiple networks (links). Whereas the data link layer oversees the delivery of the packet between two systems on the same network (links), the network layer ensures that each packet gets from its point of origin to its final destination. If two systems are connected to the same link, there is usually no need for a network layer. However, if the two systems are attached to different networks (links) with connecting devices between the networks (links), there is often a need for the network layer to accomplish source-to-destination delivery. Other responsibilities of the network layer include the following:

Logical addressing. The physical addressing implemented by the data link layer handles the addressing problem locally. If a packet passes the network boundary, we need another addressing system to help distinguish the source and destination systems. The network layer adds a header to the packet coming from the upper layer that, among other things, includes the logical addresses of the sender and receiver.

Routing. When independent networks or links are connected to create intemetworks (network of networks) or a large network, the connecting devices (called routers or switches) route or switch the packets to their final destination. One of the functions of the network layer is to provide this mechanism.

4 **Transport Layer** The transport layer is responsible for process-to-process delivery of the entire message. A process is an application program running on a host. Whereas the network layer

oversees source- to-destination delivery of individual packets, it does not recognize any relationship between those packets. It treats each one independently, as though each piece belonged to a separate message, whether or not it does. The transport layer, on the other hand, ensures that the whole message arrives intact and in order, overseeing both error control and flow control at the source- to-destination level. Other responsibilities of the transport layer include the following:

Service-point addressing. Computers often run several programs at the same time. For this reason, source-to-destination delivery means delivery not only from one computer to the next but also from a specific process (running program) on one computer to a specific process (running program) on the other. The transport layer header must therefore include a type of address called a service-point address (or port address). The network layer gets each packet to the correct computer; the transport layer gets the entire message to the correct process on that computer.

Segmentation and reassembly. A message is divided into transmittable segments, with each

- segment containing a sequence number. These numbers enable the transport layer to reassemble the message correctly upon arriving at the destination and to identify and replace packets that were lost in transmission.

Connection control. The transport layer can be either connectionless or connection oriented. A connectionless transport layer treats each segment as an independent packet and delivers it to the transport layer at the destination machine. A connection oriented transport layer makes a connection with the transport layer at the destination machine first before delivering the packets. After all the data are transferred, the connection is terminated. o Flow control. Like the data link layer, the transport layer is responsible for flow control. However, flow control at this layer is performed end to end rather than across a single link.

Error control. Like the data link layer, the transport layer is responsible for error control. However, error control at this layer is performed process-to-process rather than across a single link. The sending transport layer makes sure that the entire message arrives at the receiving transport layer without error (damage, loss, or duplication). Error correction is usually achieved through retransmission.

5 **Session Layer** The services provided by the first three layers (physical, data link, and network) are not sufficient for some processes. The session layer is the network dialog controller. It establishes, maintains, and synchronizes the interaction among communicating systems. Specific responsibilities of the session layer include the following:

Dialog control. The session layer allows two systems to enter into a dialog. It allows the communication between two processes to take place in either half duplex (one way at a time) or full-duplex (two ways at a time) mode.

Synchronization. The session layer allows a process to add checkpoints, or synchronization points, to a stream of data.

6 **Presentation Layer** The presentation layer is concerned with the syntax and semantics of the information exchanged between two systems. Specific responsibilities of the presentation layer include the following:

Translation. The processes (running programs) in two systems are usually exchanging information in the form of character strings, numbers, and so on. The information must be changed to bit streams before being transmitted. Because different computers use different

- encoding systems, the presentation layer is responsible for interoperability between these different encoding methods. The presentation layer at the sender changes the information

from its sender-dependent format into a common format. The presentation layer at the receiving machine changes the common format into its receiver-dependent format.

Encryption. To carry sensitive information, a system must be able to ensure privacy. Encryption means that the sender transforms the original information to another form and sends the resulting message out over the network. Decryption reverses the original process to transform the message back to its original form. o Compression. Data compression reduces the number of bits contained in the information. Data compression becomes particularly important in the transmission of multimedia such as text, audio, and video.

7. Application Layer The application layer enables the user, whether human or software, to access the network. It provides user interfaces and support for services such as electronic mail, remote file access and transfer, shared database management, and other types of distributed information services. Specific services provided by the application layer include the following:

Network virtual terminal. A network virtual terminal is a software version of a physical terminal, and it allows a user to log on to a remote host.

File transfer, access, and management. This application allows a user to access files in a remote host (to make changes or read data), to retrieve files from a remote computer for use in the local computer, and to manage or control files in a remote computer locally.

Mail services. This application provides the basis for e-mail forwarding and storage. Directory services. This application provides distributed database sources and access for global information about various objects and services.

TCP/IP PROTOCOL SUITE The TCPIIP protocol suite was developed prior to the OSI model. Therefore, the layers in the TCP/IP protocol suite do not exactly match those in the OSI model. The original TCP/IP protocol suite was defined as having four layers: host-to-network, internet, transport, and application. However, when TCP/IP is compared to OSI, we can say that the host-to-network layer is equivalent to the combination of the physical and data link layers. The internet layer is equivalent to the network layer, and the application layer is roughly doing the job of the session, presentation, and application layers with the transport layer in TCP/IP taking care of part of the duties of the session layer. TCP/IP is a hierarchical protocol made up of interactive modules, each of which provides a specific functionality; however, the modules are not necessarily interdependent. Whereas the OSI model specifies which functions belong to each of its layers, the layers of the TCP/IP protocol suite contain relatively independent protocols that can be mixed and matched depending on the needs of the system. The term hierarchical means that each upper-level protocol is supported by one or more lower-level protocols. At the transport layer, TCP/IP defines three protocols: Transmission Control Protocol (TCP), User Datagram Protocol (UDP), and Stream Control Transmission Protocol (SCTP). At the network layer, the main protocol defined by TCP/IP is the Internetworking Protocol (IP); there are also some other protocols that support data movement in this layer.

CONNECTING DEVICES

1. **Repeaters**

   A repeater is a device that operates only in the physical layer. Signals that carry information within a network can travel a fixed distance before attenuation endangers the integrity of the data. A repeater receives a signal and, before it becomes too weak or

corrupted, regenerates the original bit pattern. The repeater then sends the refreshed signal. A repeater can extend the physical length of a LAN, as shown in Figure 1.1.

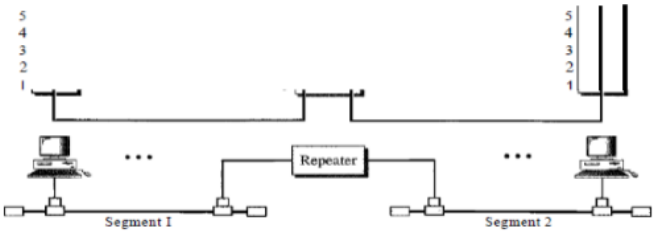


Figure 1.1 a repeater connecting two segments of a LAN

A repeater does not actually connect two LANs; it connects two segments of the same LAN. The segments connected are still part of one single LAN. A repeater is not a device that can connect two LANs of different protocols. It is tempting to compare a repeater to an amplifier, but the comparison is inaccurate. An amplifier cannot discriminate between the intended signal and noise; it amplifies equally everything fed into it. A repeater does not amplify the signal; it regenerates the signal. When it receives a weakened or corrupted signal, it creates a copy, bit for bit, at the original strength.

2. Bridges or Link Layer Switches

A bridge or Link layer switch (or simply Switch) operates in both the physical and the data link layer. As a physical layer device, it regenerates the signal it receives. As a data link layer device, the bridge can check the physical (MAC) addresses (source and destination) contained in the frame. A bridge has filtering capability. It can check the destination address of a frame and decide if the frame should be forwarded or dropped. If the frame is to be forwarded, the decision must specify the port. A bridge has a table that maps addresses to ports. Let us give an example. In Figure 15.5, two LANs are connected by a bridge. If a frame destined for station 712B13456142 arrives at port 1, the bridge consults its table to find the departing port. According to its table, frames for 7l2B13456142 leave through port 1; therefore, there is no need for forwarding, and the frame is dropped. On the other hand, if a frame for 712B13456141 arrives at port 2, the departing port is port 1 and the frame is forwarded. In the first case, LAN 2 remains free of traffic; in the second case, both LANs have traffic. In our example, we show a two-port bridge; in reality a bridge usually has more ports.

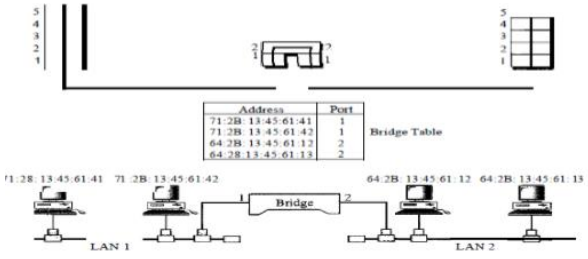


Figure 1.2 A bridge connecting two LANs

3. Hubs

A Hub is a device that operates only in the physical layer. Signals that carry information within a network can travel a fixed distance before attenuation endangers the integrity of the data. A repeater receives a signal and, before it becomes too weak or corrupted, regenerates and retimes the original bit pattern. The repeater then sends the refreshed signal. In a star topology , a repeater is a multiport device , often called as hub, that can

be used to serve as the connecting point and at the same time function as a repeater. Figure below shows that when a packet from station A to station B arrives at the hub , the signal representing the frame is regenerated to remove any possible corrupting noise , but the hub forwards the packet from all outgoing ports except the one from which the signal was received. In other words, the frame is broadcast. All the stations in the LAN receive the frame , but only station B keeps it. The rest of the stations discard it . A hub is a physical layer device. They do not have a link layer address and they do not check the link layer address of the received frame. They just regenerate the corrupted bits and send them out from every port.

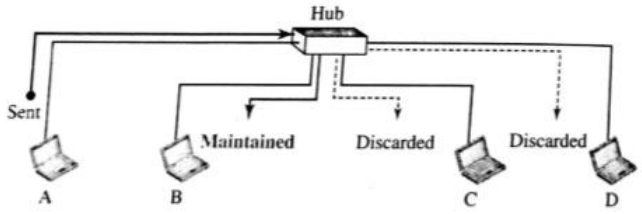


Figure 1.3 A hub

4. Routers

A router is a three-layer device that routes packets based on their logical addresses (host-to-host addressing). A router normally connects LANs and WANs in the Internet and has a routing table that is used for making decisions about the route. The routing tables are normally dynamic and are updated using routing protocols. Figure 1.4 shows a part of the Internet that uses routers to connect LANs and WANs.

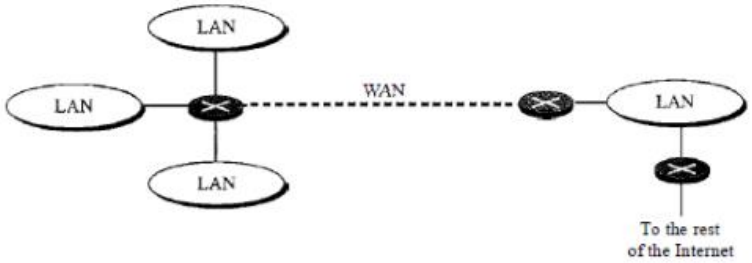


Figure 1.4 Routers connecting independent LANs and WANs

A router is a three-layer device. It operates in the physical, data link and network layers. As a physical layer device, it regenerates the signal it receives. As a link layer device the router checks the physical addresses contained in the packet. As a network layer device, a router checks the network layer addresses. A router can connect networks. In other words, a router is an internetworking device; it connects independent networks to form an internetwork. There are three major differences between a router and a repeater or switch.

a. A router has a physical and logical address for each of its interfaces.

b. A router acts only on those packets in which the link layer destination address matches the address of the interface at which the packet arrives.

c. A router changes the link layer address of the packet when it forwards the packet.

5. Gateways

A gateway is normally a computer that operates in all five layers of the Internet or seven layers of OSI model. A gateway takes an application message, reads it, and interprets it. This means that it can be used as a connecting device between two internetworks that use different models. For example, a network designed to use the OSI model can be connected to another network using the Internet model. The gateway connecting the

two systems can take a frame as it arrives from the first system, move it up to the OSI application layer, and remove the message.

CHAPTER TWO

NETWORKS

A network is a set of devices (often referred to as nodes) connected by communication links. A node can be a computer, printer, or any other device capable of sending and/or receiving data generated by other nodes on the network.

Network Criteria A network must be able to meet a certain number of criteria. The most important of these are performance, reliability, and security.

Performance: Performance can be measured in many ways, including transit time and response time. Transit time is the amount of time required for a message to travel from one device to another. Response time is the elapsed time between an inquiry and a response. The performance of a network depends on a number of factors, including the number of users, the type of transmission medium, the capabilities of the connected hardware, and the efficiency of the software.

Reliability: Network reliability is measured by the frequency of failure, the time it takes a link to recover from a failure, and the network's robustness in a catastrophe.

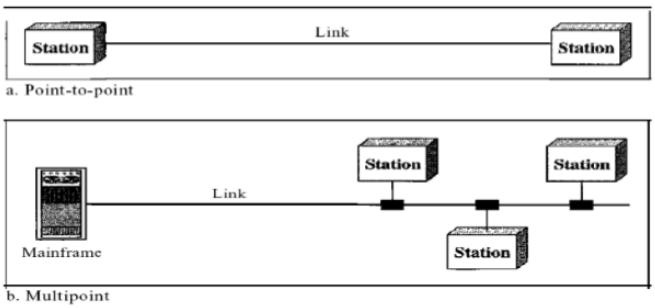
Security: Network security issues include protecting data from unauthorized access, protecting data from damage and development, and implementing policies and procedures for recovery from breaches and data losses.

PHYSICAL STRUCTURES

TYPES OF CONNECTIONS: A network is two or more devices connected through links. A link is a communications pathway that transfers data from one device to another. There are two possible types of connections: point-to-point and multipoint.

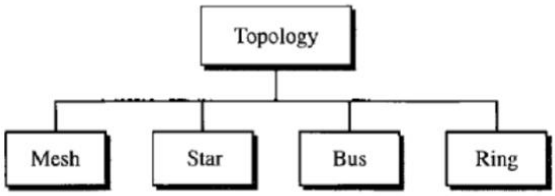
Point-to-Point A point-to-point connection provides a dedicated link between two devices. The entire capacity of the link is reserved for transmission between those two devices. Most point-to-point connections use an actual length of wire or cable to connect the two ends, but other options, such as microwave or satellite links, are also possible. When you change television channels by infrared remote control, you are establishing a point-to-point connection between the remote control and the television's control system.

Multipoint A multipoint (also called multidrop) connection is one in which more than two specific devices share a single link. In a multipoint environment, the capacity of the channel is shared, either spatially or temporally. If several devices can use the link simultaneously, it is a spatially shared connection. If users must take turns, it is a timeshared connection.



PHYSICAL TOPOLOGY:

The term physical topology refers to the way in which a network is laid out physically. One or more devices connect to a link; two or more links form a topology. The topology of a network is the geometric representation of the relationship of all the links and linking devices (usually called nodes) to one another. There are four basic topologies possible: mesh, star, bus, and ring.



1. **Mesh:** In a mesh topology, every device has a dedicated point-to-point link to every other device. The term dedicated means that the link carries traffic only between the two devices it connects. To find the number of physical links in a fully connected mesh network with n nodes, we first consider that each node must be connected to every other node. Node 1 must be connected to n - I nodes, node 2 must be connected to n – 1 nodes, and finally node n must be connected to n - 1 nodes. We need n(n - 1) physical links. However, if each physical link allows communication in both directions (duplex mode), we can divide the number of links by 2. In other words, we can say that in a mesh topology, we need n(n -1) /2 duplex-mode links. To accommodate that many links,

every device on the network must have n – 1 input/output ports to be connected to the other n - 1 stations.

Advantages:

1. The use of dedicated links guarantees that each connection can carry its own data load, thus eliminating the traffic problems that can occur when links must be shared by multiple devices.

2. A mesh topology is robust. If one link becomes unusable, it does not incapacitate the entire system.

3. There is the advantage of privacy or security. When every message travels along a dedicated line, only the intended recipient sees it. Physical boundaries prevent other users from gaining access to messages.

4. Point-to-point links make fault identification and fault isolation easy. Traffic can be routed to avoid links with suspected problems. This facility enables the network manager to discover the precise location of the fault and aids in finding its cause and solution.

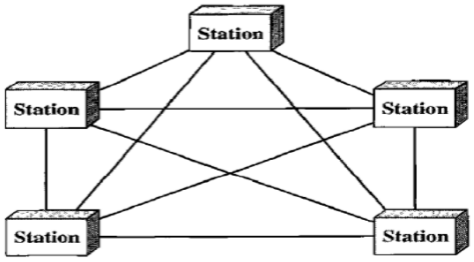
Disadvantages:

1. Disadvantage of a mesh are related to the amount of cabling because every device must be connected to every other device.

2. Installation and reconnection are difficult.

3. The sheer bulk of the wiring can be greater than the available space (in walls, ceilings, or floors) can accommodate.

4. The hardware required to connect each link (I/O ports and cable) can be prohibitively expensive.



2. Star Topology: In a star topology, each device has a dedicated point-to-point link only to a central controller, usually called a hub. The devices are not directly linked to one another. Unlike a mesh topology, a star topology does not allow direct traffic between devices. The controller acts as an exchange: If one device wants to send data to another, it sends the data to the controller, which then relays the data to the other connected device .

Advantages:

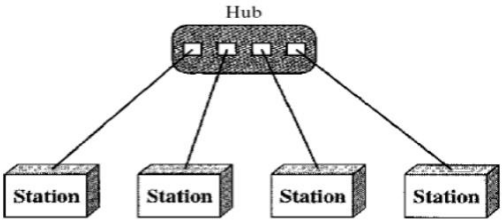
1. A star topology is less expensive than a mesh topology. In a star, each device needs only one link and one I/O port to connect it to any number of others.

2. Easy to install and reconfigure.

3. Far less cabling needs to be housed, and additions, moves, and deletions involve only one connection: between that device and the hub.

4. Other advantage include robustness. If one link fails, only that link is affected. All other links remain active. This factor also lends itself to easy fault identification and fault isolation. As long as the hub is working, it can be used to monitor link problems and bypass defective links. Disadvantages: One big disadvantage of a star topology is the dependency of the

whole topology on one single point, the hub. If the hub goes down, the whole system is dead. Although a star requires far less cable than a mesh, each node must be linked to a central hub. For this reason, often more cabling is required in a star than in some other topologies (such as ring or bus).



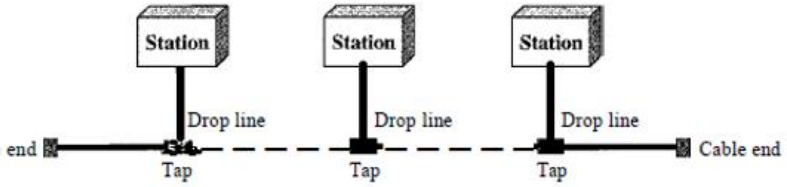
2. **BUS:** A bus topology is multipoint. One long cable acts as a backbone to link all the devices in a network. Nodes are connected to the bus cable by drop lines and taps. A drop line is a connection running between the device and the main cable. A tap is a connector that either splices into the main cable or punctures the sheathing of a cable to create a contact with the metallic core. As a signal travels along the backbone, some of its energy is transformed into heat. Therefore, it becomes weaker and weaker as it travels farther and farther. For this reason there is a limit on the number of taps a bus can support and on the distance between those taps.

   Advantages:

   Advantages of a bus topology include ease of installation. Backbone cable can be laid along the most efficient path, then connected to the nodes by drop lines of various lengths. In this way, a bus uses less cabling than mesh or star topologies. In a star, for example, four network devices in the same room require four lengths of cable reaching all the way to the hub. In a bus, this redundancy is eliminated. Only the backbone cable stretches through the entire facility. Each drop line has to reach only as far as the nearest point on the backbone.

   Disadvantages:

   Disadvantages include difficult reconnection and fault isolation. A bus is usually designed to be optimally efficient at installation. It can therefore be difficult to add new devices. Signal reflection at the taps can cause degradation in quality. This degradation can be controlled by limiting the number and spacing of devices connected to a given length of cable. Adding new devices may therefore require modification or replacement of the backbone. In addition, a fault or break in the bus cable stops all transmission, even between devices on the same side of the problem. The damaged area reflects signals back in the direction of origin, creating noise in both directions.



3. **RING:**

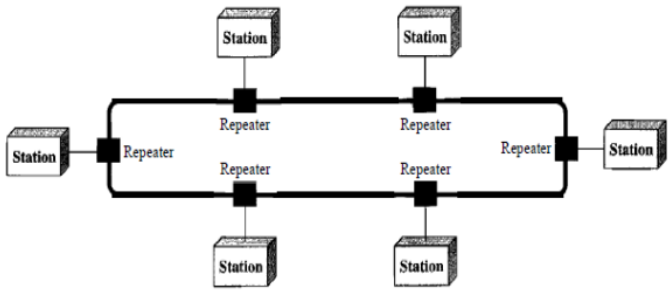
   In a ring topology, each device has a dedicated point-to-point connection with only the two devices on either side of it. A signal is passed along the ring in one direction, from device to device, until it reaches its destination. Each device in the ring incorporates a repeater. When a device receives a signal intended for another device, its repeater regenerates the bits and passes them along.

   Advantages:

A ring is relatively easy to install and reconfigure. Each device is linked to only its immediate neighbors (either physically or logically). To add or delete a device requires changing only two connections. The only constraints are media and traffic considerations (maximum ring length and number of devices). In addition, fault isolation is simplified. Generally in a ring, a signal is circulating at all times. If one device does not receive a signal within a specified period, it can issue an alarm. The alarm alerts the network operator to the problem and its location.

Disadvantages:

Unidirectional traffic can be a disadvantage. In a simple ring, a break in the ring (such as a disabled station) can disable the entire network. This weakness can be solved by using a dual ring or a switch capable of closing off the break. Ring topology was prevalent when IBM introduced its local-area network Token Ring. Today, the need for higher-speed LANs has made this topology less popular.



CHAPTER THREE

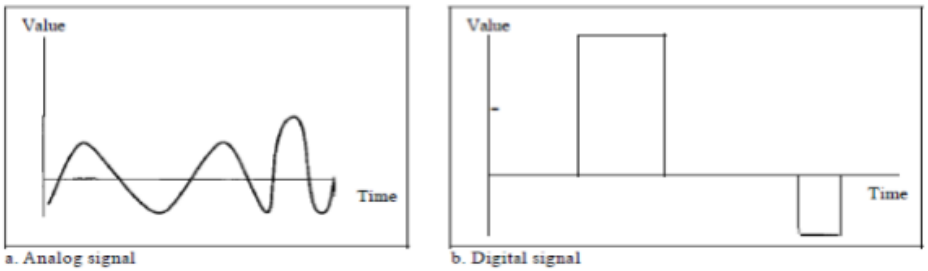
ANALOG AND DIGITAL

Analog Data: The term analog data refers to information that is continuous; For example, an analog clock that has hour, minute, and second hands gives information in a continuous form; the movements of the hands are continuous. Analog data, such as the sounds made by a human voice, take on continuous values. When someone speaks, an analog wave is created in the air. This can be captured by a microphone and converted to an analog signal or sampled and converted to a digital signal.

Digital Data: Digital data refers to information that has discrete states. For example, a digital clock that reports the hours and the minutes will change suddenly from 8:05 to 8:06. Digital data takes on discrete values. For example, data are stored in computer memory in the form of Os and 1s. They can be converted to a digital signal or modulated into an analog signal for transmission across a medium.

Analog and Digital Signals: Like the data they represent, signals can be either analog or digital. An analog signal has infinitely many levels of intensity over a period of time. As the wave moves from value A to value B, it passes through and includes an infinite number of values along its path. A digital signal, on the other hand, can have only a limited number of defined values. Although each value can be any number, it is often as simple as 1 and O. The simplest way to show signals is by plotting them on a pair of perpendicular axes. The vertical axis represents the value or strength of a signal. The horizontal axis represents time. Figure below illustrates an analog signal and a digital signal. The curve representing the analog signal passes through an infinite number of points. The vertical lines of the digital signal, however, demonstrate the sudden jump that the signal makes from value to value.

Figure 3.1 *Comparison of analog and digital signals*



a. Analog signal b. Digital signal

Periodic and Nonperiodic Signals:

A periodic signal completes a pattern within a measurable time frame, called a period, and repeats that pattern over subsequent identical periods. The completion of one full pattern is called a cycle. A nonperiodic signal changes without exhibiting a pattern or cycle that repeats over time.

PERIODIC ANALOG SIGNALS:

Periodic analog signals can be classified as simple or composite. A simple periodic analog signal, a sine wave, cannot be decomposed into simpler signals. A composite periodic analog signal is composed of multiple sine waves. Sine Wave The sine wave is the most fundamental form of a periodic analog signal. When we visualize it as a simple oscillating curve, its change over the course of a cycle is smooth and consistent, a continuous, rolling flow. Figure below shows a sine wave. Each cycle consists of a single arc above the time axis followed by a single arc below it.

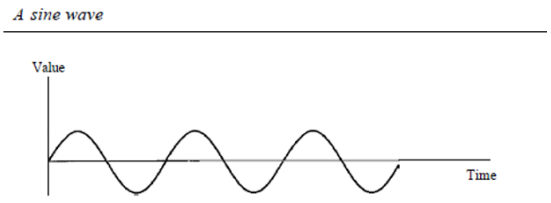


Figure 3.2

Characteristics of Signals:

1. **Peak Amplitude**

The peak amplitude of a signal is the absolute value of its highest intensity, proportional to the energy it carries. For electric signals, peak amplitude is normally measured in volts. Figure below shows two signals and their peak amplitudes.

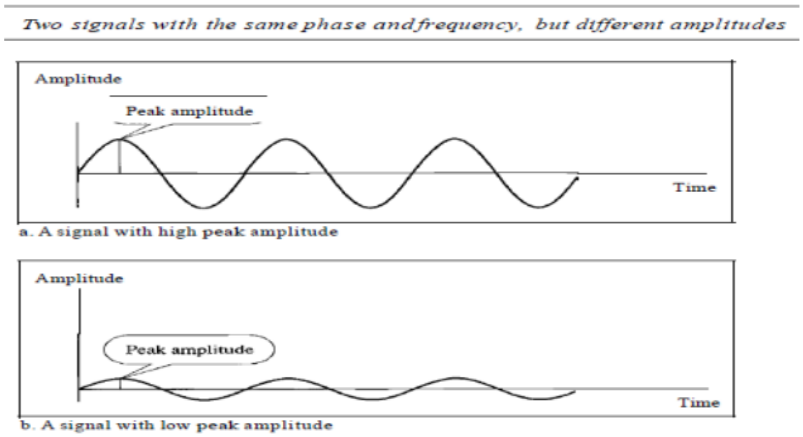


Figure 3.3

2. **Period and Frequency**

Period refers to the amount of time, in seconds, a signal needs to complete 1 cycle. Frequency refers to the number of periods in I s. Note that period and frequency are just one characteristic defined in two ways. Period is the inverse of frequency, and frequency is the inverse of period, as the following formulas show. f=1/T and T=1/f

Period is formally expressed in seconds. Frequency is formally expressed in Hertz (Hz), which is cycle per second.

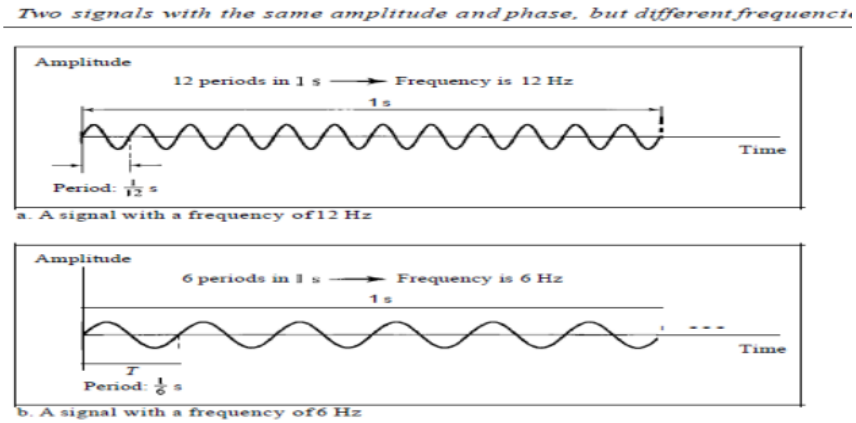


Figure 3.4

3. Phase

The term phase describes the position of the waveform relative to time O. If we think of the wave as something that can be shifted backward or forward along the time axis, phase describes the amount of that shift. It indicates the status of the first cycle. Phase is measured in degrees or radians [360° is 2n rad; 1° is 2n/360 rad, and 1 rad is 360/(2n)]. A phase shift of 360° corresponds to a shift of a complete period; a phase shift of 180° corresponds to a shift of onehalf of a period; and a phase shift of 90° corresponds to a shift of one-quarter of a period.

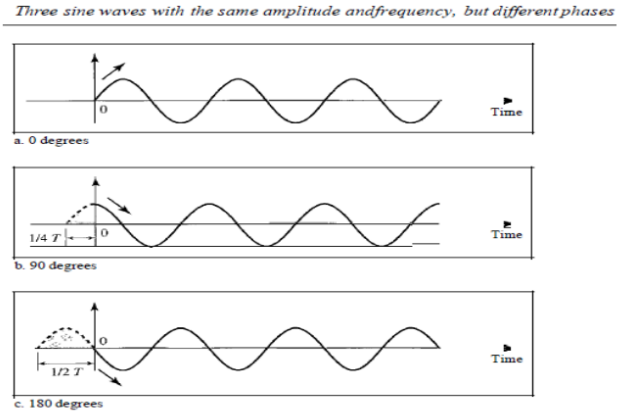


Figure 3.5

I. A sine wave with a phase of 0° starts at time 0 with a zero amplitude. The amplitude is increasing.

II. II. A sine wave with a phase of 90° starts at time 0 with a peak amplitude. The amplitude is decreasing.

III. III. A sine wave with a phase of 180° starts at time 0 with a zero amplitude. The amplitude is decreasing.

4. **Wavelength**

Wavelength is another characteristic of a signal traveling through a transmission medium. Wavelength binds the period or the frequency of a simple sine wave to the propagation speed of the medium. While the frequency of a signal is independent of the medium, the wavelength depends on both the frequency and the medium. Wavelength is a property of any type of signal. In data communications, we often use wavelength to describe the transmission of light in an optical fiber. The wavelength is the distance a simple signal can travel in one period. Wavelength can be calculated if one is given the propagation speed (the speed of light) and the period of the signal. However, since period and frequency are related to each other, if we represent wavelength by λ, propagation speed by c (speed of light), and frequency by f, we get Wavelength=Propagation speed * Period = propagation speed/frequency λ=c/f The wavelength is normally measured in micrometers (microns) instead of meters.

Bandwidth

The range of frequencies contained in a composite signal is its bandwidth. The bandwidth is normally a difference between two numbers. For example, if a composite signal contains frequencies between 1000 and 5000, its bandwidth is 5000 - 1000, or 4000. Figure 3.6 shows the concept of bandwidth. The figure depicts two composite signals, one periodic and the other nonperiodic. The bandwidth of the periodic signal contains all integer frequencies between 1000 and 5000 (1000,

100 I, 1002, ...). The bandwidth of the nonperiodic signals has the same range, but the frequencies are continuous.

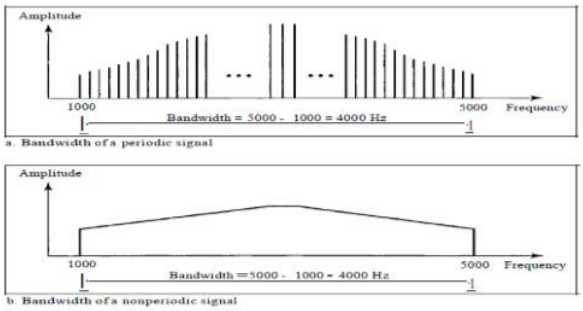


Figure 3.6 The bandwidth of periodic and nonperiodic composite signals

DIGITAL SIGNALS

In addition to being represented by an analog signal, information can also be represented by a digital signal. For example, a 1 can be encoded as a positive voltage and a 0 as zero voltage. A digital signal can have more than two levels. In this case, we can send more than 1 bit for each level. Figure 3.7 shows two signals, one with two levels and the other with four.

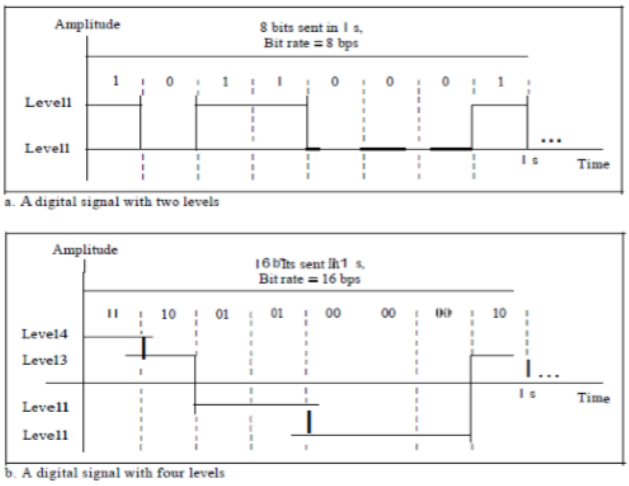


Figure 3.7 Two digital signals: one with two signal levels and the other with four signal levels

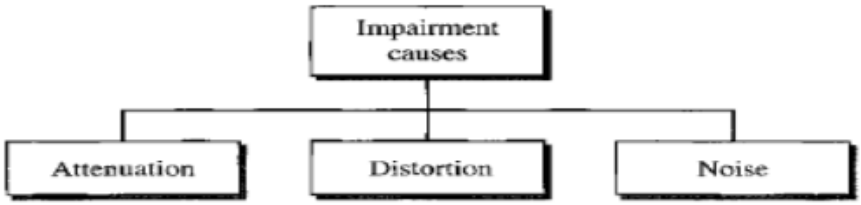
Bit Rate

Most digital signals are nonperiodic, and thus period and frequency are not appropriate characteristics. Another term-bit rate is used to describe digital signals. The bit rate is the number of bits sent in 1s, expressed in bits per second (bps). Figure 3.7 shows the bit rate for two signals.

TRANSMISSION IMPAIRMENT

Signals travel through transmission media, which are not perfect. The imperfection causes signal impairment. This means that the signal at the beginning of the medium is not the same as the signal at the end of the medium. What is sent is not what is received. Three causes of impairment are attenuation, distortion, and noise.

Causes of impairment



1. **Attenuation**

   Attenuation means a loss of energy. When a signal, simple or composite, travels through a medium, it loses some of its energy in overcoming the resistance of the medium. That is why a wire carrying electric signals gets warm, if not hot, after a while. Some of the electrical energy in the signal is converted to heat. To compensate for this loss, amplifiers are used to amplify the signal. Attenuation is measured in terms of Decibels. The decibel (dB) measures the relative strengths of two signals or one signal at two different points. Note that the decibel is negative if a signal is attenuated and positive if a signal is amplified. dB=10log10 P2/P1

   Variables PI and P2 are the powers of a signal at points 1 and 2, respectively.

2. **Distortion:**

   Distortion means that the signal changes its form or shape. Distortion can occur in a composite signal made of different frequencies. Each signal component has its own propagation speed through a medium and, therefore, its own delay in arriving at the final destination. Differences in delay may create a difference in phase if the delay is not exactly the same as the period duration. In other words, signal components at the receiver have phases different from what they had at the sender. The shape of the composite signal is therefore not the same. Figure 3.8 shows the effect of distortion on a composite signal.

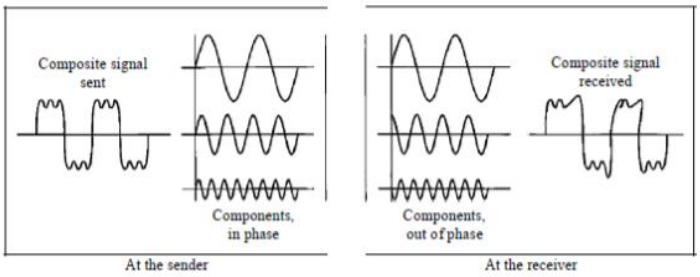


Figure 3.8 Distortion

3. **Noise**

   Noise is another cause of impairment. Several types of noise, such as thermal noise, induced noise, crosstalk, and impulse noise, may corrupt the signal. Thermal noise is the random motion of electrons in a wire which creates an extra signal not originally sent by the transmitter. Induced noise comes from sources such as motors and appliances. These devices act as a sending antenna, and the transmission medium acts as the receiving antenna. Crosstalk is the effect of one wire on the other. One wire acts as a sending antenna and the other as the receiving antenna. Impulse noise is a spike (a signal with high energy in a very short time) that comes from power lines, lightning, and so on.

Signal-to-Noise Ratio (SNR)

The signal-to-noise ratio is defined as SNR= Average Signal power / Average Noise Power SNR is actually the ratio of what is wanted (signal) to what is not wanted (noise). A high SNR means the signal is less corrupted by noise; a low SNR means the signal is more corrupted by noise. Because SNR is the ratio of two powers, it is often described in decibel units, SNR dB, defined as SNRdB = l0logl0 SNR

DATA RATE LIMITS

A very important consideration in data communications is how fast we can send data, in bits per second. over a channel. Data rate depends on three factors:

1. The bandwidth available
2. The level of the signals we use

3. The quality of the channel (the level of noise) Two theoretical formulas were developed to calculate the data rate: one by Nyquist for a noiseless channel. another by Shannon for a noisy channel.

Noiseless Channel: Nyquist Bit Rate

For a noiseless channel, the Nyquist bit rate formula defines the theoretical maximum bit rate BitRate = 2 x bandwidth x 10g2 L In this formula, bandwidth is the bandwidth of the channel, L is the number of signal levels used to represent data, and BitRate is the bit rate in bits per second. According to the formula, we might think that, given a specific bandwidth, we can have any bit rate we want by increasing the number of signa11eve1s. Although the idea is theoretically correct, practically there is a limit. When we increase the number of signal1eve1s, we impose a burden on the receiver. If the number of levels in a signal is just 2, the receiver can easily distinguish between a 0 and a 1. If the level of a signal is 64, the receiver must be very sophisticated to distinguish between 64 different levels. In other words, increasing the levels of a signal reduces the reliability of the system.

Noisy Channel: Shannon Capacity

In reality, we cannot have a noiseless channel; the channel is always noisy. In 1944, Claude Shannon introduced a formula, called the Shannon capacity, to determine the theoretical highest data rate for a noisy channel: Capacity =bandwidth X log2 (1 +SNR) In this formula, bandwidth is the bandwidth of the channel, SNR is the signal-to-noise ratio, and capacity is the capacity of the channel in bits per second. Note that in the Shannon formula there is no indication of the signal level, which means that no matter how many levels we have, we cannot achieve a data rate higher than the capacity of the channel. In other words, the formula defines a characteristic of the channel, not the method of transmission.

Bandwidth in Bits per Seconds

The term bandwidth can also refer to the number of bits per second that a channel, a link, or even a network can transmit. For example, one can say the bandwidth of a Fast Ethernet network is a maximum of 100 Mbps. This means that this network can send 100 Mbps.

DIGITAL-TO-ANALOG CONVERSION

Digital-to-analog conversion is the process of changing one of the characteristics of an analog signal based on the information in digital data. Figure 3.9 shows the relationship between the digital information, the digital-to-analog modulating process, and the resultant analog signal.

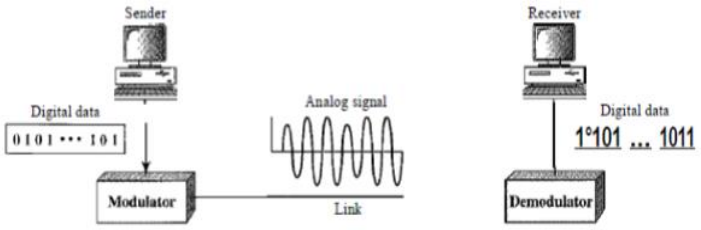


Figure 3.9 Analog to Digital Conversion

A sine wave is defined by three characteristics: amplitude, frequency, and phase. When we vary anyone of these characteristics, we create a different version of that wave. So, by changing one characteristic of a simple electric signal, we can use it to represent digital data. Any of the three characteristics can be altered in this way, giving us at least three mechanisms for modulating digital data into an analog signal: amplitude shift keying (ASK), frequency shift keying (FSK), and phase shift keying (PSK).

Aspects of Digital-to-Analog Conversion

Before we discuss specific methods of digital-to-analog modulation, two basic issues must be reviewed: bit and baud rates and the carrier signal.

Data Element Versus Signal Element

Data element as the smallest piece of information to be exchanged, the bit. We also defined a signal element as the smallest unit of a signal that is constant.

Data Rate Versus Signal Rate

We can define the data rate (bit rate) and the signal rate (baud rate). The relationship between them is S= N/r baud where N is the data rate (bps) and r is the number of data elements carried in one signal element. The value of r in analog transmission is r =log2 L, where L is the type of signal element, not the level.

Carrier Signal

In analog transmission, the sending device produces a high-frequency signal that acts as a base for the information signal. This base signal is called the carrier signal or carrier frequency. The receiving device is tuned to the frequency of the carrier signal that it expects from the sender. Digital information then changes the carrier signal by modifying one or more of its characteristics (amplitude, frequency, or phase). This kind of modification is called modulation (shift keying).

1. **Amplitude Shift Keying (ASK)**

   In amplitude shift keying, the amplitude of the carrier signal is varied to create signal elements. Both frequency and phase remain constant while the amplitude changes. Although we can have several levels (kinds) of signal elements, each with a different amplitude, ASK is normally implemented using only two levels. This is referred to as binary amplitude shift keying or on-off keying (OOK). The peak amplitude of one signal level is 0; the other is

the same as the amplitude of the carrier frequency. Figure 3.10 gives a conceptual view of binary ASK.

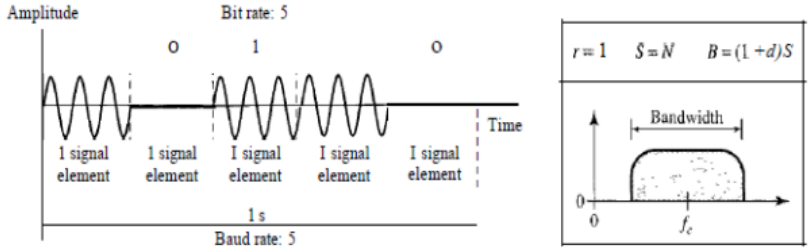


Figure 3.10 Binary Amplitude Shift Keying

2. **Frequency Shift Keying (FSK)**

In frequency shift keying, the frequency of the carrier signal is varied to represent data. The frequency of the modulated signal is constant for the duration of one signal element, but changes for the next signal element if the data element changes. Both peak amplitude and phase remain constant for all signal elements. One way to think about binary FSK (or BFSK) is to consider two carrier frequencies. In Figure 3.11, we have selected two carrier frequencies,f1 andf2. We use the first carrier if the data element is 0; we use the second if the data element is 1. However, note that this is an unrealistic example used only for demonstration purposes. Normally the carrier frequencies are very high, and the difference between them is very small.

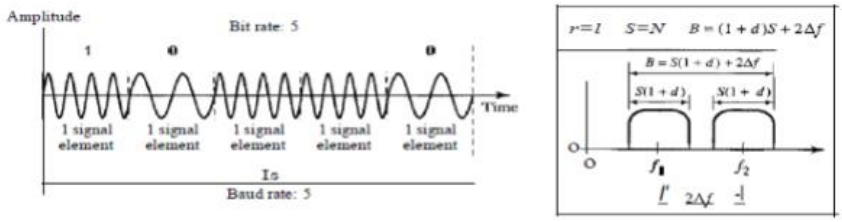


Figure 3.11 Binary Frequency Shift Keying

3. **Phase Shift Keying (PSK)**

In phase shift keying, the phase of the carrier is varied to represent two or more different signal elements. Both peak amplitude and frequency remain constant as the phase changes. Today, PSK is more common than ASK or FSK. The simplest PSK is binary PSK, in which we have only two signal elements, one with a phase of 0°, and the other with a phase of 180°. Figure 3.12 gives a conceptual view of PSK.

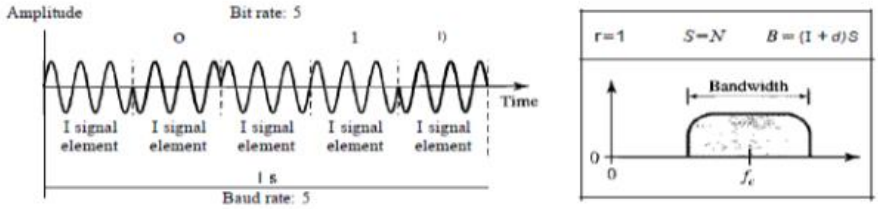


Figure 3.12 Binary Phase Shift Keying

CHAPTER FOUR

MULTIPLEXING

Whenever the bandwidth of a medium linking two devices is greater than the bandwidth needs of the devices, the link can be shared. Multiplexing is the set of techniques that allows the simultaneous transmission of multiple signals across a single data link. In a multiplexed system, n lines share the bandwidth of one link. Figure 4.1 shows the basic format of a multiplexed system. The lines on the left direct their transmission streams to a multiplexer (MUX), which combines them into a single stream (many-to-one). At the receiving end, that stream is fed into a demultiplexer (DEMUX), which separates the stream back into its component transmissions (one-to-many) and directs them to their corresponding lines. In the figure, the word link refers to the physical path. The word channel refers to the portion of a link that carries a transmission between a given pair of lines. One link can have many (n) channels.

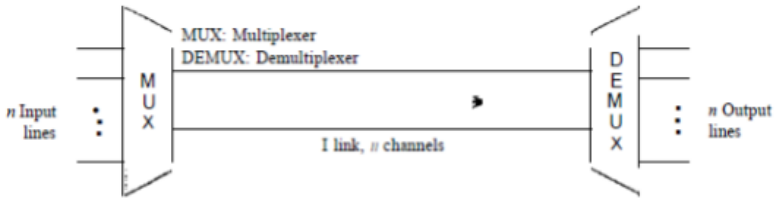


Figure 4.1 Dividing a link into channels

There are three basic multiplexing techniques: frequency-division multiplexing, wavelengthdivision multiplexing, and time-division multiplexing. The first two are techniques designed for analog signals, the third, for digital signals.

1. **Frequency-Division Multiplexing**

   Frequency-division multiplexing (FDM) is an analog technique that can be applied when the bandwidth of a link (in hertz) is greater than the combined bandwidths of the signals to be transmitted. In FDM, signals generated by each sending device modulate different carrier frequencies. These modulated signals are then combined into a single composite signal that can be transported by the link. Carrier frequencies are separated by sufficient bandwidth to accommodate the modulated signal. These bandwidth ranges are the channels through which the various signals travel. Channels can be separated by strips of unused bandwidth-guard bands-to prevent signals from overlapping. In addition, carrier frequencies must not interfere with the original data frequencies.

   Multiplexing Process

   Figure 4.2 is a conceptual illustration of the multiplexing process. Each source generates a signal of a similar frequency range. Inside the multiplexer, these similar signals modulates different carrier frequencies. The resulting modulated signals are then combined into a single composite signal that is sent out over a media link that has enough bandwidth to accommodate it.

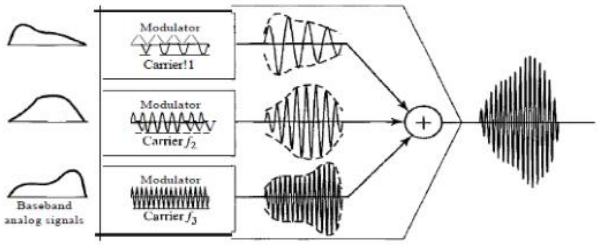


Figure 4.2 FDM Process

Demultiplexing Process

The demultiplexer uses a series of filters to decompose the multiplexed signal into its constituent component signals. The individual signals are then passed to a demodulator that separates them from their carriers and passes them to the output lines. Figure 4.3 is a conceptual illustration of demultiplexing process.

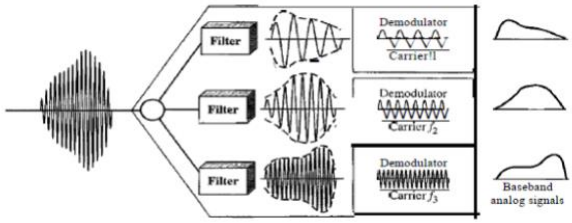


Figure 4.3 FDM Demultiplexing Example

2. **Wavelength-Division Multiplexing**

Wavelength-division multiplexing (WDM) is designed to use the high-data-rate capability of fiber-optic cable. The optical fiber data rate is higher than the data rate of metallic transmission cable. Using a fiber-optic cable for one single line wastes the available bandwidth. Multiplexing allows us to combine several lines into one. WDM is conceptually the same as FDM, except that the multiplexing and demultiplexing involve optical signals transmitted through fiber-optic channels. The idea is the same: We are combining different signals of different frequencies. The difference is that the frequencies are very high. Figure 4.4 gives a conceptual view of a WDM multiplexer and demultiplexer. Very narrow bands of light from different sources are combined to make a wider band of light. At the receiver, the signals are separated by the demultiplexer.

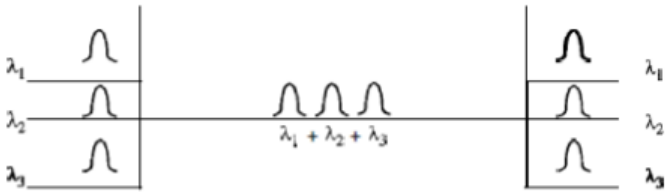


Figure 4.4 wavelength- Division multiplexing

Although WDM technology is very complex, the basic idea is very simple. We want to combine multiple light sources into one single light at the multiplexer and do the reverse at the demultiplexer. The combining and splitting of light sources are easily handled by a prism.

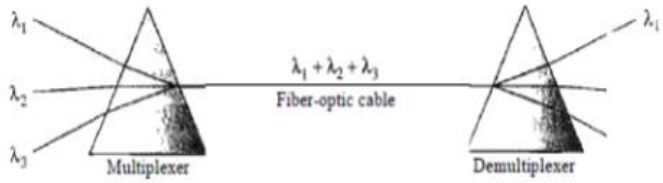


Figure 4.5 Prisms in multiplexing and demultiplexing

One application of WDM is the SONET network in which multiple optical fiber lines are multiplexed and demultiplexed. 3. Time-Division Multiplexing Time-division multiplexing (TDM) is a digital process that allows several connections to share the high bandwidth of a linle Instead of sharing a portion of the bandwidth as in FDM, time is shared. Each connection occupies a portion of time in the link. Figure 4.6 gives a conceptual view of TDM. Note that the same link is used as in FDM; here, however, the link is shown sectioned by time rather than by frequency. In the figure, portions of signals 1,2,3, and 4 occupy the link sequentially.

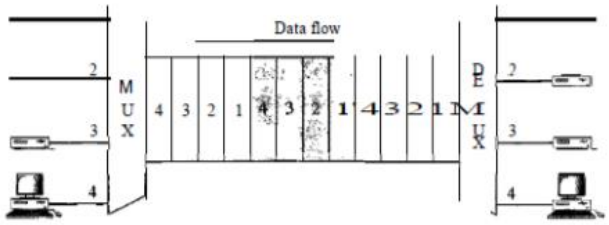


Figure 4.6 TDM

Note that in Figure 4.6 we are concerned with only multiplexing, not switching. This means that all the data in a message from source 1 always go to one specific destination, be it 1, 2, 3, or 4. The delivery is fixed and unvarying, unlike switching.

CHAPTER FIVE

TRANSMISSION MEDIA

A transmission medium can be broadly defined as anything that can carry information from a source to a destination. For example, the transmission medium for two people having a dinner conversation is the air. The air can also be used to convey the message in a smoke signal or semaphore. For a written message, the transmission medium might be a mail carrier, a truck, or an airplane. In data communications the definition of the information and the transmission medium is more specific. The transmission medium is usually free space, metallic cable, or fiber-optic cable. The information is usually a signal that is the result of a conversion of data from another form.

Guided Media

Guided media, which are those that provide a conduit from one device to another, include twisted-pair cable, coaxial cable, and fiber-optic cable. A signal traveling along any of these media is directed and contained by the physical limits of the medium. Twisted-pair and coaxial cable use metallic (copper) conductors that accept and transport signals in the form of electric current. Optical fiber is a cable that accepts and transports signals in the form of light.

1. Twisted-Pair Cable A twisted pair consists of two conductors (normally copper), each with its own plastic insulation, twisted together, as shown in Figure 5.1.

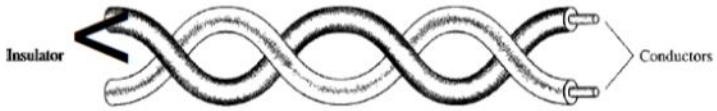


Figure 5.1 Twisted pair cable

One of the wires is used to carry signals to the receiver, and the other is used only as a ground reference. The receiver uses the difference between the two. In addition to the signal sent by the sender on one of the wires, interference (noise) and crosstalk may affect both wires and create unwanted signals. If the two wires are parallel, the effect of these unwanted signals is not the same in both wires because they are at different locations relative to the noise or crosstalk sources (e,g., one is closer and the other is farther). This results in a difference at the receiver. By twisting the pairs, a balance is maintained. For example, suppose in one twist, one wire is closer to the noise source and the other is farther; in the next twist, the reverse is true. Twisting makes it probable that both wires are equally affected by external influences (noise or crosstalk). This means that the receiver, which calculates the difference between the two, receives no unwanted signals. The unwanted signals are mostly canceled out. From the above discussion, it is clear that the number of twists per unit of length (e.g., inch) has some effect on the quality of the cable. Applications Twisted-pair cables are used in telephone lines to provide voice and data channels. The local loop-the line that connects subscribers to the central telephone office-commonly consists of unshielded twisted-pair cables. The DSL lines that are used by the telephone companies to provide high-data-rate connections also use the high-bandwidth capability of unshielded twistedpair cables. Local-area networks, such as lOBase-T and lOOBase-T, also use twisted-pair cables.

2. Coaxial Cable Coaxial cable (or coax) carries signals of higher frequency ranges than those in twisted pair cable, in part because the two media are constructed quite differently. Instead of having

two wires, coax has a central core conductor of solid or stranded wire (usually copper) enclosed in an insulating sheath, which is, in turn, encased in an outer conductor of metal foil, braid, or a combination of the two. The outer metallic wrapping serves both as a shield against noise and as the second conductor, which completes the circuit. This outer conductor is also enclosed in an insulating sheath, and the whole cable is protected by a plastic cover (see Figure 5.2).

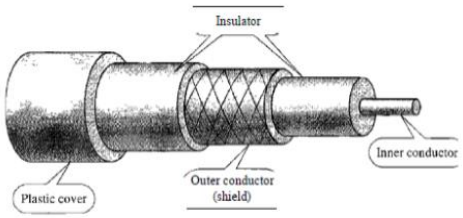


Figure 5.2 Coaxial Cable

Applications

Coaxial cable was widely used in analog telephone networks where a single coaxial network could carry 10,000 voice signals. Later it was used in digital telephone networks where a single coaxial cable could carry digital data up to 600 Mbps. However, coaxial cable in telephone networks has largely been replaced today with fiber-optic cable. Cable TV networks also use coaxial cables. In the traditional cable TV network, the entire network used coaxial cable. Later, however, cable TV providers replaced most of the media with fiber-optic cable; hybrid networks use coaxial cable only at the network boundaries, near the consumer premises. Cable TV uses RG-59 coaxial cable. Another common application of coaxial cable is in traditional Ethernet LANs. Because of its high bandwidth, and consequently high data rate, coaxial cable was chosen for digital transmission in early Ethernet LANs.

3. Fiber Optic Cable: A fiber-optic cable is made of glass or plastic and transmits signals in the form of light. To understand optical fiber, we first need to explore several aspects of the nature of light. Light travels in a straight line as long as it is moving through a single uniform If a ray of light traveling through one substance suddenly enters another substance (of a different density), the ray changes direction. Figure 5.3 shows how a ray of light changes direction when going from a more dense to a less dense substance.

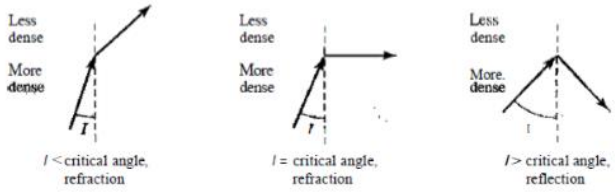


Figure 5.3 Bending of light ray

As the figure shows, if the angle of incidence I (the angle the ray makes with the line perpendicular to the interface between the two substances) is less than the critical angle, the ray refracts and moves closer to the surface. If the angle of incidence is equal to the critical angle, the light bends along the interface. If the angle is greater than the critical angle, the ray reflects (makes a turn) and travels again in the denser substance. Note that the critical angle is a property of the substance, and its value differs from one substance to another. Optical fibers use reflection to guide light through a channel. A glass or plastic core is surrounded by a cladding of less dense glass or plastic. The

difference in density of the two materials must be such that a beam of light moving through the core is reflected off the cladding instead of being refracted into it. See Figure 5.4.

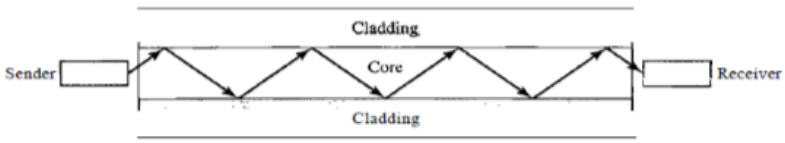


Figure 5.4 Optical Fiber

Cable Composition

Figure 5.5 shows the composition of a typical fiber-optic cable. The outer jacket is made of either PVC or Teflon. Inside the jacket are Kevlar strands to strengthen the cable. Kevlar is a strong material used in the fabrication of bulletproof vests. Below the Kevlar is another plastic coating to cushion the fiber. The fiber is at the center of the cable, and it consists of cladding and core.

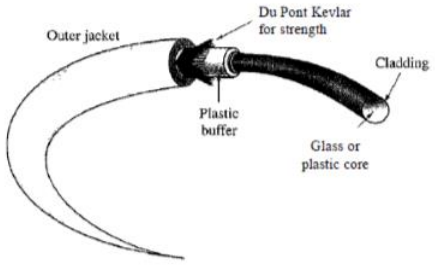


Figure 5.5 fiber Construction

Applications

Fiber-optic cable is often found in backbone networks because its wide bandwidth is costeffective. Today, with wavelength-division multiplexing (WDM), we can transfer data at a rate of 1600 Gbps. The SONET network provides such a backbone. Some cable TV companies use a combination of optical fiber and coaxial cable, thus creating a hybrid network. Optical fiber provides the backbone structure while coaxial cable provides the connection to the user premises. This is a cost-effective configuration since the narrow bandwidth requirement at the user end does not justify the use of optical fiber. Local-area networks such as 100Base-FX network (Fast Ethernet) and 1000Base-X also use fiber-optic cable.

Advantages and Disadvantages of Optical Fiber

Advantages Fiber-optic cable has several advantages over metallic cable (twisted pair or coaxial).

a. Higher bandwidth. Fiber-optic cable can support dramatically higher bandwidths (and hence data rates) than either twisted-pair or coaxial cable. Currently, data rates and bandwidth utilization over fiber-optic cable are limited not by the medium but by the signal generation and reception technology available.

b. Less signal attenuation. Fiber-optic transmission distance is significantly greater than that of other guided media. A signal can run for 50 km without requiring regeneration. We need repeaters every 5 km for coaxial or twisted-pair cable.

c. Immunity to electromagnetic interference. Electromagnetic noise cannot affect fiber-optic cables.
d. Resistance to corrosive materials. Glass is more resistant to corrosive materials than copper.

e. Light weight. Fiber-optic cables are much lighter than copper cables.

f. Greater immunity to tapping. Fiber-optic cables are more immune to tapping than copper cables. Copper cables create antenna effects that can easily be tapped.

Disadvantages There are some disadvantages in the use of optical fiber.

a. Installation and maintenance. Fiber-optic cable is a relatively new technology. Its installation and maintenance require expertise that is not yet available everywhere.

b. Unidirectional light propagation. Propagation of light is unidirectional. If we need bidirectional communication, two fibers are needed.

c. Cost. The cable and the interfaces are relatively more expensive than those of other guided media. If the demand for bandwidth is not high, often the use of optical fiber cannot be justified.

UNGUIDED MEDIA: **WIRELESS**

Unguided media transport electromagnetic waves without using a physical conductor. This type of communication is often referred to as wireless communication. Signals are normally broadcast through free space and thus are available to anyone who has a device capable of receiving them. Unguided signals can travel from the source to destination in several ways: ground propagation, sky propagation, and line-of-sight propagation, as shown in Figure 5.6. In ground propagation, radio waves travel through the lowest portion of the atmosphere, hugging the earth. These lowfrequency signals emanate in all directions from the transmitting antenna and follow the curvature of the planet. Distance depends on the amount of power in the signal: The greater the power, the greater the distance. In sky propagation, higher-frequency radio waves radiate upward into the ionosphere where they are reflected back to earth. This type of transmission allows for greater distances with lower output power. In line-or-sight propagation, very high-frequency signals are transmitted in straight lines directly from antenna to antenna. Antennas must be directional, facing each other, and either tall enough or close enough together not to be affected by the curvature of the earth. Line-of-sight propagation is tricky because radio transmissions cannot be completely focused.

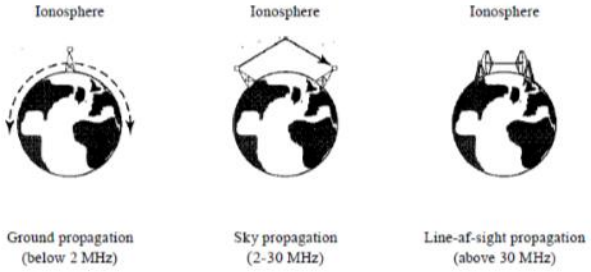


Figure 5.6 Propagation methods

1. **Radio Waves**

   Waves ranging in frequencies between 3 kHz and 1 GHz are called radio waves. Radio waves, for the most part, are omnidirectional. When an antenna transmits radio waves, they are propagated in all directions. This means that the sending and receiving antennas do not have to be aligned. A sending antenna sends waves that can be received by any receiving antenna. The omnidirectional property has a disadvantage, too. The radio waves transmitted by one antenna are susceptible to interference by another antenna that may send signals using the same frequency or band. Radio waves, particularly those waves that propagate in the sky mode, can travel long distances. This makes radio waves a good candidate for long-

distance broadcasting such as AM radio. Radio waves, particularly those of low and medium frequencies, can penetrate walls. This characteristic can be both an advantage and a disadvantage. It is an advantage because, for example, an AM radio can receive signals inside a building. It is a disadvantage because we cannot isolate a communication to just inside or outside a building. The radio wave band is relatively narrow, just under 1 GHz, compared to the microwave band. When this band is divided into sub bands, the sub bands are also narrow, leading to a low data rate for digital communications.

Omnidirectional Antenna

Radio waves use omnidirectional antennas that send out signals in all directions. Based on the wavelength, strength, and the purpose of transmission, we can have several types of antennas. Figure 5.7 shows an omnidirectional antenna.

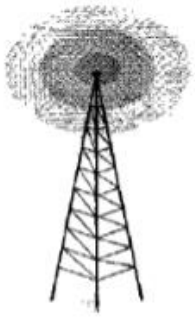


Figure 5.7 Omnidirectional Antenna

Applications

The omnidirectional characteristics of radio waves make them useful for multicasting, in which there is one sender but many receivers. AM and FM radio, television, maritime radio, cordless phones, and paging are examples of multicasting.

2. **Microwaves**

Electromagnetic waves having frequencies between I and 300 GHz are called microwaves. Microwaves are unidirectional. When an antenna transmits microwave waves, they can be narrowly focused. This means that the sending and receiving antennas need to be aligned. The unidirectional property has an obvious advantage. A pair of antennas can be aligned without interfering with another pair of aligned antennas. The following describes some characteristics of microwave propagation:

a. Microwave propagation is line-of-sight. Since the towers with the mounted antennas need to be in direct sight of each other, towers that are far apart need to be very tall. The curvature of the earth as well as other blocking obstacles do not allow two short towers to communicate by using microwaves. Repeaters are often needed for long distance communication.

b. Very high-frequency microwaves cannot penetrate walls. This characteristic can be a disadvantage if receivers are inside buildings.

c. The microwave band is relatively wide, almost 299 GHz. Therefore wider sub bands can be assigned, and a high data rate is possible

d. Use of certain portions of the band requires permission from authorities.

Unidirectional Antenna

Microwaves need unidirectional antennas that send out signals in one direction. Two types of antennas are used for microwave communications: the parabolic dish and the horn (see Figure 7.21). A parabolic dish antenna is based on the geometry of a parabola: Every line parallel to the line of symmetry (line of sight) reflects off the curve at angles such that all the

lines intersect in a common point called the focus. The parabolic dish works as a funnel, catching a wide range of waves and directing them to a common point. In this way, more of the signal is recovered than would be possible with a single-point receiver. Outgoing transmissions are broadcast through a horn aimed at the dish. The microwaves hit the dish and are deflected outward in a reversal of the receipt path. A horn antenna looks like a gigantic scoop. Outgoing transmissions are broadcast up a stem (resembling a handle) and deflected outward in a series of narrow parallel beams by the curved head. Received transmissions are collected by the scooped shape of the horn, in a manner similar to the parabolic dish, and are deflected down into the stem.

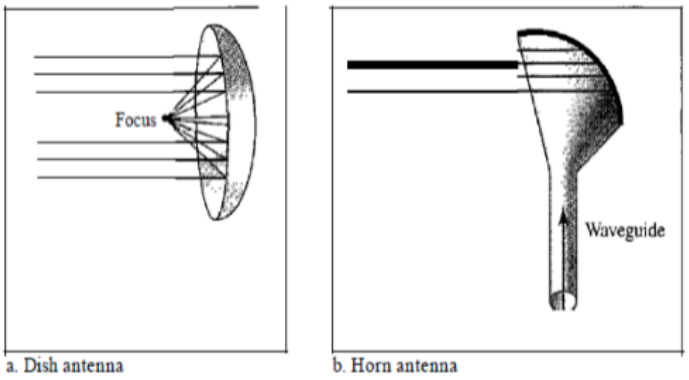


Figure 5.8 Unidirectional Antennas

3. **Infrared**

Infrared waves, with frequencies from 300 GHz to 400 THz (wavelengths from 1 mm to 770 nm), can be used for short-range communication. Infrared waves, having high frequencies, cannot penetrate walls. This advantageous characteristic prevents interference between one system and another; a short-range communication system in one room cannot be affected by another system in the next room. When we use our infrared remote control, we do not interfere with the use of the remote by our neighbors. However, this same characteristic makes infrared signals useless for long-range communication. In addition, we cannot use infrared waves outside a building because the sun's rays contain infrared waves that can interfere with the communication. Applications The infrared band, almost 400 THz, has an excellent potential for data transmission. Such a wide bandwidth can be used to transmit digital data with a very high data rate. The Infrared Data Association (IrDA), an association for sponsoring the use of infrared waves, has established standards for using these signals for communication between devices such as keyboards, mice, PCs, and printers. For example, some manufacturers provide a special port called the IrDA port that allows a wireless keyboard to communicate with a PC. The standard originally defined a data rate of 75 kbps for a distance up to 8 m. The recent standard defines a data rate of 4 Mbps. Infrared signals defined by IrDA transmit through line of sight; the IrDA port on the keyboard needs to point to the PC for transmission to occur.