Data Communication

Data communication is the exchange of data (In the form 0's and 1's) between two devices via some form of transmission medium (such as wire cable). This communication is considered local (devices in the same building) or remote if the devices are farther apart.

The effectiveness of a data communication system depends on three ***fundamental characteristics***:-

1- Delivery: the system must deliver data to the correct destination. Data must be received by the intended device or user and only by that device or user.

2- Accuracy: The system must deliver data accurately. Data that have been altered in transmission and left uncorrected are unusable.

3- Timeliness: The system must deliver data in timely manner. Data delivered late are useless.

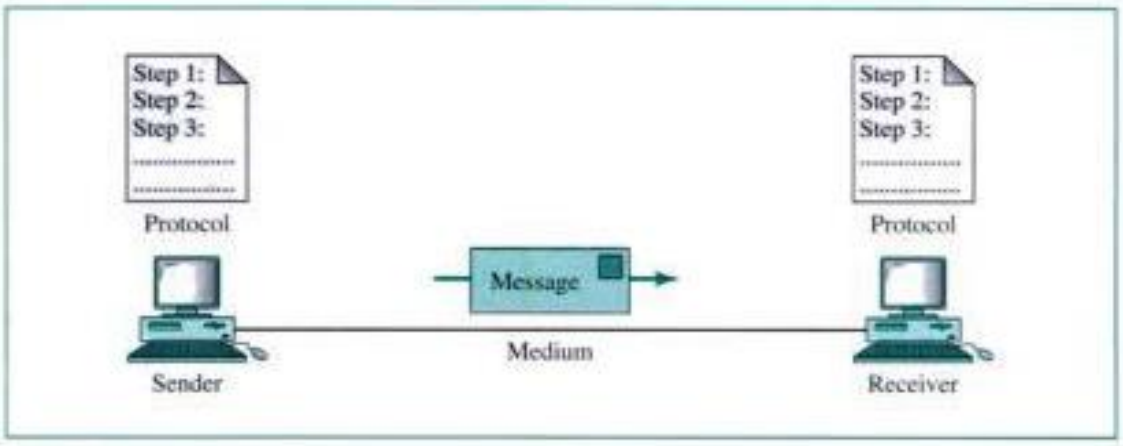
Real – time transmission: - delivery of data (video, audio, and voice data) as they are produced in the same order they are produced and without significant delay.

Components of a Data Communication System:

A data communication system is made up of 5 components:

A data communications system has five components (see Fig. 1.1).

Figure 1.1 *Five components of data communication*



1- ***Message:*** The message is the information (data) to be communicated. It can consist of text, numbers, sound, picture, or video, or any combination of these.
2- ***Sender:*** The sender is the device that sends the data message. It can be a computer, workstation, telephone handset, video camera, and so on.
3- ***Receiver:*** The receiver is the device that receives the message. It can be a computer, workstation, telephone, handset, television, and so on.
4- ***Medium:*** The transmission medium is the physical path by which a message travels from sender to receiver. It can consist of twisted pair wire, coaxial cable, fiber optic cable, laser, or radio waves.
5- ***Protocol:*** A protocol is a set of rules that govern data communication. It represents an agreement between the communicating devices. Without a protocol, two devices may be connected, but not communicating, just as a person speaking French cannot be understood by a person who speaks only Japanese.

Uses of Computer Networks

Before we start to examine the technical issues in detail, it is worth devoting some time to pointing out why people are interested in computer networks and what they can be used for.

1-Business Applications

Many companies have a substantial number of computers. For example, a company may have separate computers to monitor production, keep track of inventories, and do the payroll. Initially, each of these computers may have worked in isolation from the others, but at some point, management may have decided to connect them to be able to extract and correlate information about the entire company.

The issue here is resource sharing, and the goal is to make all programs, equipment, and especially data available to anyone on the network without regard to the physical location of the resource and the user. An obvious and widespread example is having a group of office workers share a common printer However; probably even more important than sharing physical resources such as printers, scanners, and CD burners, is sharing information. Every

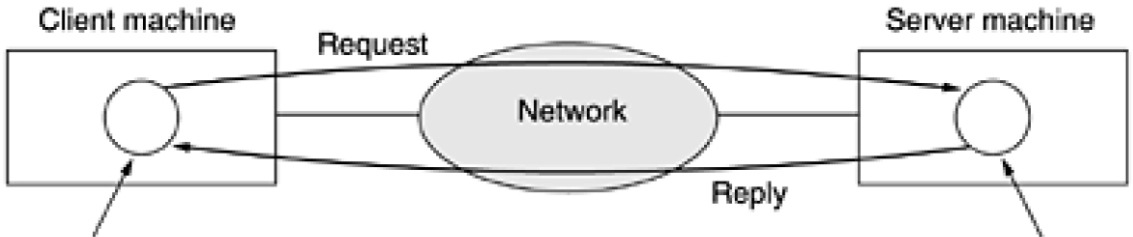
large and medium-sized company and many small companies are vitally dependent on computerized information.

Most companies have customer records, inventories, accounts receivable, financial statements, tax information, and much more online.

If all of its computers went down, a bank could not last more than five minutes. A modern manufacturing plant, with a computer-controlled assembly line, would not last even that long. In the simplest of terms, one can imagine a company's information system as consisting of one or more databases and some number of employees who need to access them remotely. In this model, the data are stored on powerful computers called servers. Often these are centrally housed and maintained by a system administrator. In contrast, the employees have simpler machines, called clients, on their desks, with which they access remote data, for example, to include in spreadsheets they are constructing.

This whole arrangement is called the client-server model. It is widely used and forms the basis of much network usage. It is applicable when the client and server are both in the same building (e.g., belong to the same company), but also when they are far apart. For example, when a person at home accesses a page on the World Wide Web, the same model is employed, with the remote Web server being the server and the user's personal computer being the client. Under most conditions, one server can handle a large number of clients.

If we look at the client-server model in detail, we see that two processes are involved, one on the client machine and one on the server machine. Communication takes the form of the client process sending a message over the network to the server process. The client process then waits for a reply message. When the server process gets the request, it performs the requested work or looks up the requested data and sends back a reply. These messages are shown in figure below.



A second goal of setting up a computer network has to do with people rather than information or even computers. A computer network can provide a powerful communication medium among employees.

A third goal for increasingly many companies is doing business electronically with other companies, especially suppliers and customers.

A fourth goal that is starting to become more important is doing business with consumers over the Internet. Airlines, bookstores, and music vendors have discovered that many customers like the convenience of shopping from home. Consequently, many companies provide catalogs of their goods and services online and take orders on-line. This sector is expected to grow quickly in the future. It is called e-commerce (electronic commerce).

2- Home Applications

Some of the more popular uses of the Internet for home users are as follows:

1. Access to remote information.
2. Person-to-person communication.
3. Interactive entertainment.
4. Electronic commerce.

3-Mobile Users

Mobile computers, such as notebook computers and personal digital assistants (PDAs), are one of the fastest-growing segments of the computer industry.

Although wireless networking and mobile computing are often related, they are not identical, as shown in figure below. Here we see a distinction between fixed wireless and mobile wireless.

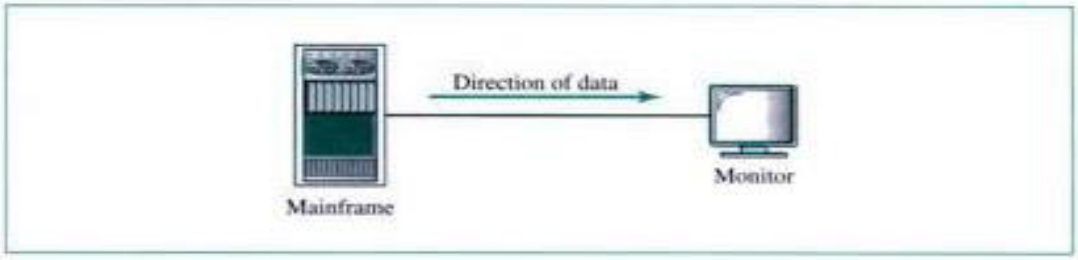
Wireless	Mobile	Applications
No	No	Desktop computers in offices
No	Yes	A notebook computer used in a hotel room
Yes	No	Networks in older, unwired buildings
Yes	Yes	Portable office; PDA for store inventory

Data Flow

Communication between two devices can be **simplex**, **half-duplex**, or **full-duplex**

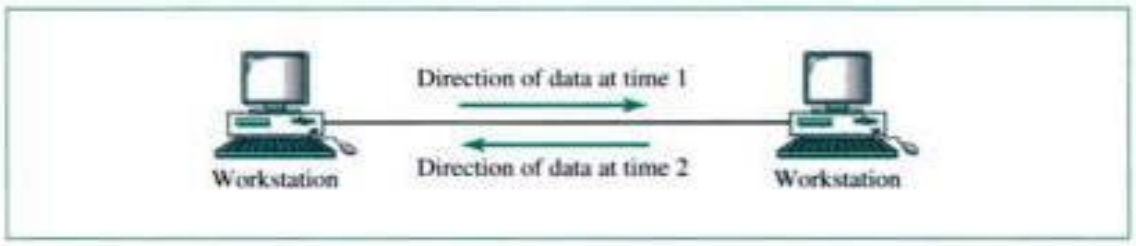
1. **Simplex:-** The communication is unidirectional as on, one way street, only one of the two stations on the link can transmit, the other can only receive. Ex: [keyboard can only introduce input, monitor can only accept output].

Figure 1.2 *Simplex*



2. **Half Duplex:-** in this mode, each station can both transmit and receive, but not at the same time. When one device sending, the other can only receive, and vice versa. The entire capacity of the channel is taken over by whichever of the two devices is transmitting at the time. Walkie – talkies is an example of half – duplex system.

Figure 1.3 *Half-duplex*



3-Full – Duplex: in full duplex mode, (also called duplex), both stations can transmit and receive simultaneously. Signals going in either direction share the capacity of the link. This sharing can occur in two ways:

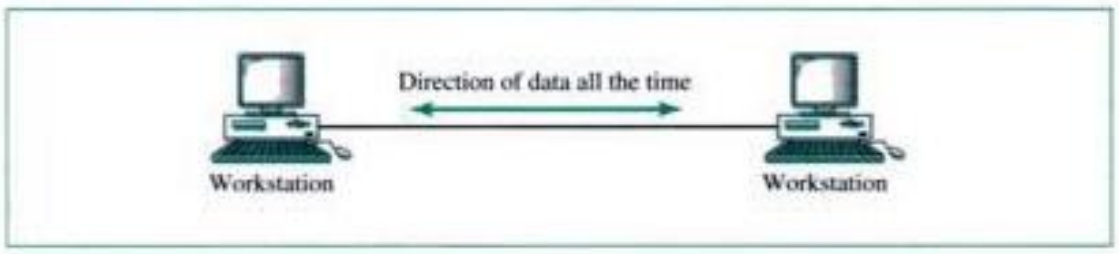
- The link must contain two physically separate transmission path, one for sending and other for receiving.

Or:

- The capacity of the channel is divided between signals traveling in both directions.

One common example of full-duplex communication is the telephone network. When two people are communicating by a telephone line, both can talk and listen at the same time.

Figure 1.4 *Full-duplex*



Networks:

A networks is a set of devices (often retired to as nodes) connected by media links. A node can be computer, printer, or any other device capable of sending and/or

receiving data generated by other nodes on the network. The links connecting the devices are often called communication channels.

Distributed processing:

Network use distributed processing, in which a task is divided among multiple computers, Instead of a single large machine being responsible for all aspects of a process, each separate computer handles a subset.

Network Criteria:

To be considered effective and efficient, a N.W must meet a number of criteria:

1- ***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 response.

The performance of N.W depends on a number of factors including:

- Number of users: Having a large number of concurrent users can slow response time in a N.W not designed to coordinate heavy traffic load. How a N.W response to loading is a measure of its performance.
- Type of transmission medium: The medium defines the speed at which data can travel through a connection (data rate).
- Hardware.
- Software.

2- ***Reliability:*** In addition to accuracy of delivery, network reliability is measured by:

- Frequency of failure
- Recovery time of a network after a failure.
- Catastrophe.

3- ***Security:*** Network security issues include protecting data from unauthorized access and viruses.

Network Hardware:

In general there are two dimension stands out as important to classify Network: ***Transmission Technology*** and Scale

Transmission Technology

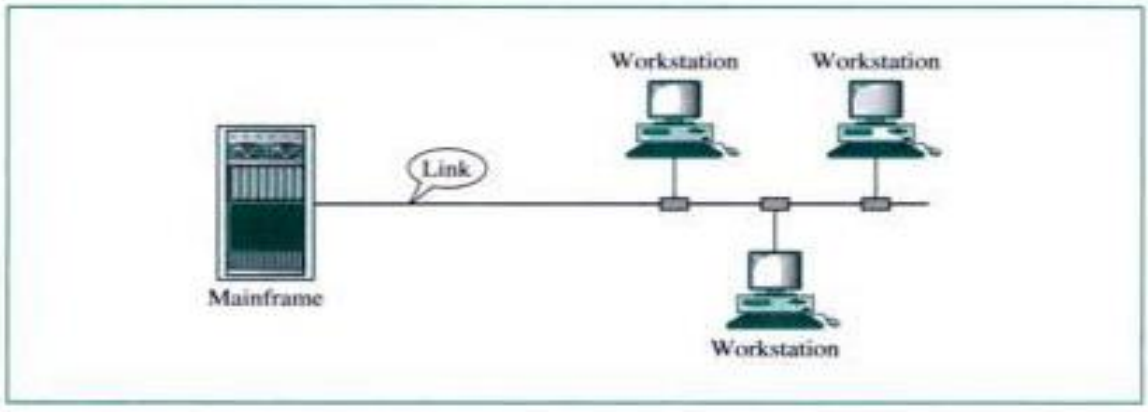
Broadly speaking, there are two types of transmission Technology:

1- Broadcast links (multipoint)

It is the network that has a single communication channel that is shared by all the machines on the network. In multipoint environment, the capacity of the channel is shared, either spatially or temporally. If several devices can use the link simultaneously, it is spatially shared. If user must take turn, it is a time shared line configuration.

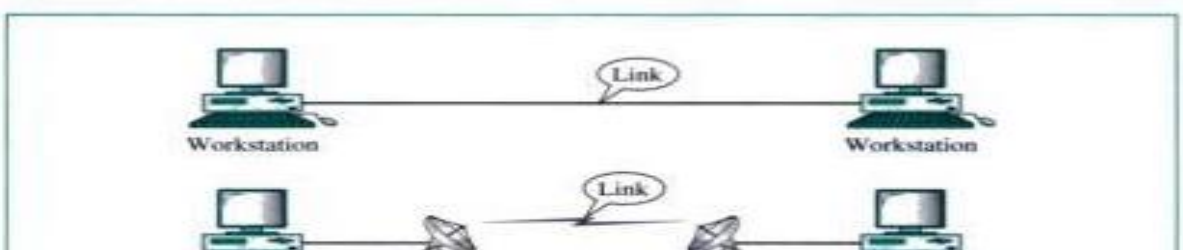
In Broadcast, short messages, called packets, sent by one machine are received by all the others. An address filed within the packet specifies the intended recipient.

Figure 1.6 *Multipoint connection*



2- Point – to – Point Network:

A point – to – point line configuration provides a dedicated link between two devices. The entire capacity of the channel is reserved for transmission between those two devices. Most point – to – point line configuration use an actual length or wire or cable to connect the two ends, but other options such as a microwave or satellite links are also possible.



An alternative for classifying network is their scale:-

	Interprocessor distance	processor located in same
Personal Area N.W	1M	Square meter
	10 M	Room
LAN	100M	Building
	1 km	Campus
MAN	10 km	City
	100 km	Country
WAN	1000km	Continent
The internet	10,000km	Planet

- The personal Area Network it networks that meant for person. Ex (mouse to computer / wireless).
- Longer range networks can be divided into local, Metropolitans, and wide area Network.
- The connection of two or more networks is called internetwork.

Ex (Internet).

Topology:

- The term topology refers to the way a network is laid out, either physically or logically.
- Two 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 to each other.
- The topology describe how the devices in a networks are interconnected rather than their physical arrangement. For example, having a star topology doesn't mean that all of the computer in the network must be places physically around a hub in a star shape. A consideration when choosing a topology is the relative statues of the devices to be linked.

There are four basic topologies possible:-

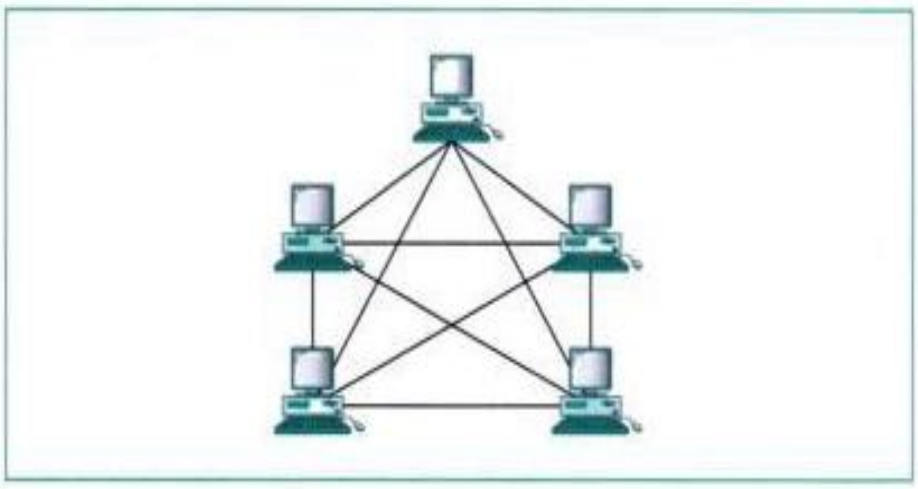
Mesh, Star, Ring, Bus

1- Mesh Topology:

In a mesh topology, every device has a dedicated point – to point link to every other devices. The term dedicated means that the link carries traffic only between the two devices it connects.

- A fully connected mesh network has (**n (n-1)/2**) physical channels to link **n** devices.
- Every device on the network must have (**n-1**) input / output port.
- Advantages of mesh topology

1. Eliminating the traffic problems that can occur when links must be shared by multiple devices.
2. A mesh topology is robust, if one like becomes unusable, it doesn't incapacitate the entire system.
3. Security, when every message sent travels along a dedicated line, only the intended recipient sees it.
4. Point – to – point links make fault identification and fault isolation easy.



A fully connected mesh topology (five devices)

The main disadvantages of mesh are related to the amount of cabling and the number of I/O ports required.

1- Because every device must be connected to every other device, installation and reconfiguration are difficult.

2- The sheer bulk of the wiring can be greater that the available space can accommodate.

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

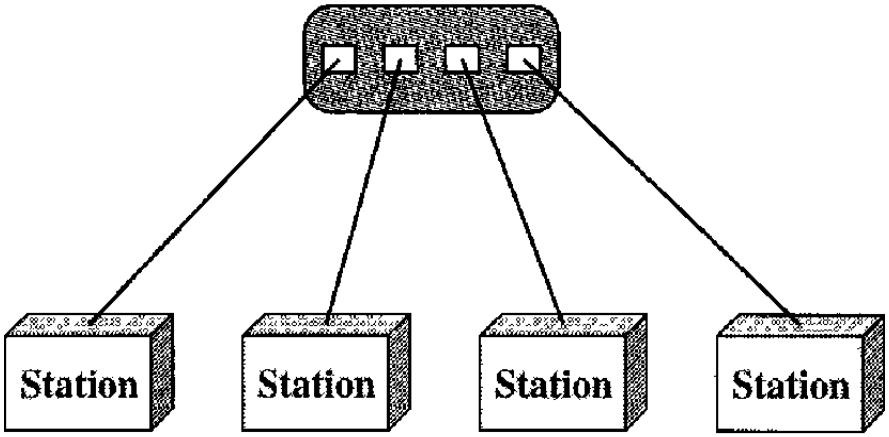
For these reasons, a mesh topology is usually implemented in a limited fashion.

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**.

- Devices are not linked to each other.

- Star topology doesn't allow direct traffic between devices.
- The controller acts as an exchanger: If one device wants to send data to another, it sends to the controller which then relays the data to the other connected devices.
- A star topology is less expensive than mesh topology and easier to install and reconfigure.
- Star topology is robust, such that if one link fails, only that link is affected, all other links remain active.
- Easy fault identification and isolation.



A star topology connecting four stations

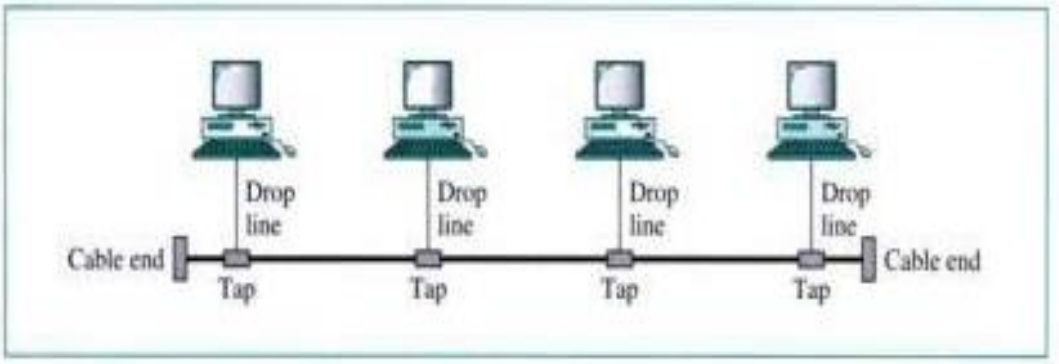
One big disadvantage of a star topology is the dependency of the whole topology on one single point, the hub. If the hub goes down, the whole system is dead.

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).

The star topology is used in local-area networks (LANs), LANs often use a star topology with a central hub.

3- Bus Topology:

- A bus topology is a multipoint.
- One long cable acts as a backbone to link all the devices in the network.



Bus Topology

- 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.
- As a signals travels along the backbone, some of its energy is transformed into heat, therefore, it becomes weaker and weaker the farther it has to travel. For this reason there is a limit on the number of taps a bus can support and on distance between those taps.
- Both ends of the bus must be terminated with a resistive load known as a **terminating resistor**. These resistors serve to prevent signal bounce.

- Advantages of a bus topology include:-

1- Ease of installation.
2- Bus using less cabling than mesh, star, tree topologies.

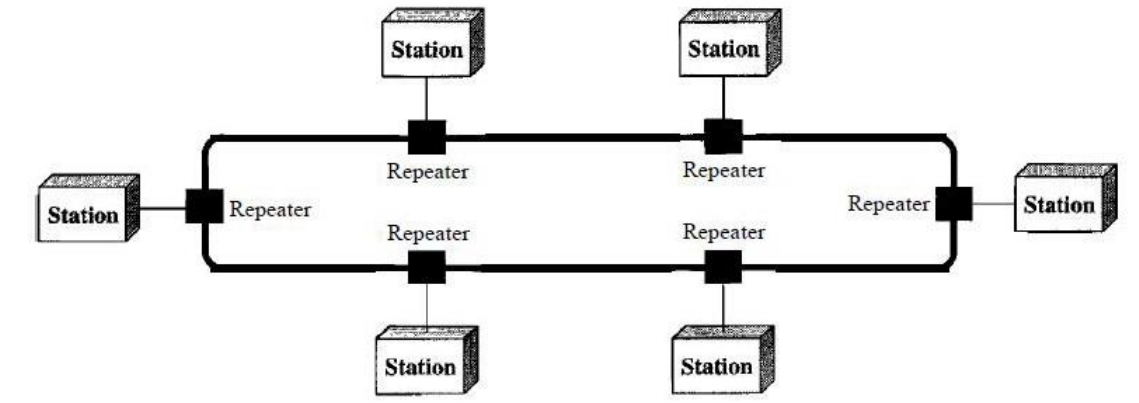
- Disadvantages of bus topology include:-

1. Difficult reconfiguration and fault isolation.
2. Difficult to add new devices. Adding new devices may therefore require modification or replacement of the backbone.
3. A fault or break in the bus cable stops all transmission, even between devices on the same side of the problem.
4. Bus topology was the one of the first topologies used in the design of early local area networks. Ethernet LANs can use a bus topology, but they are less popular now.

4- Ring Topology:

In a ring topology, each device has a dedicated point – to – point line configuration only with 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 repeats, regenerates the bits and passes them along.



Ring Topology

- Advantages of Ring Topology:

- Ring is relatively easy to install and reconfigure.
- Fault isolation is simplified. Generally in a ring a signal is circulating at all times. If one device doesn't receive a signal within a specified

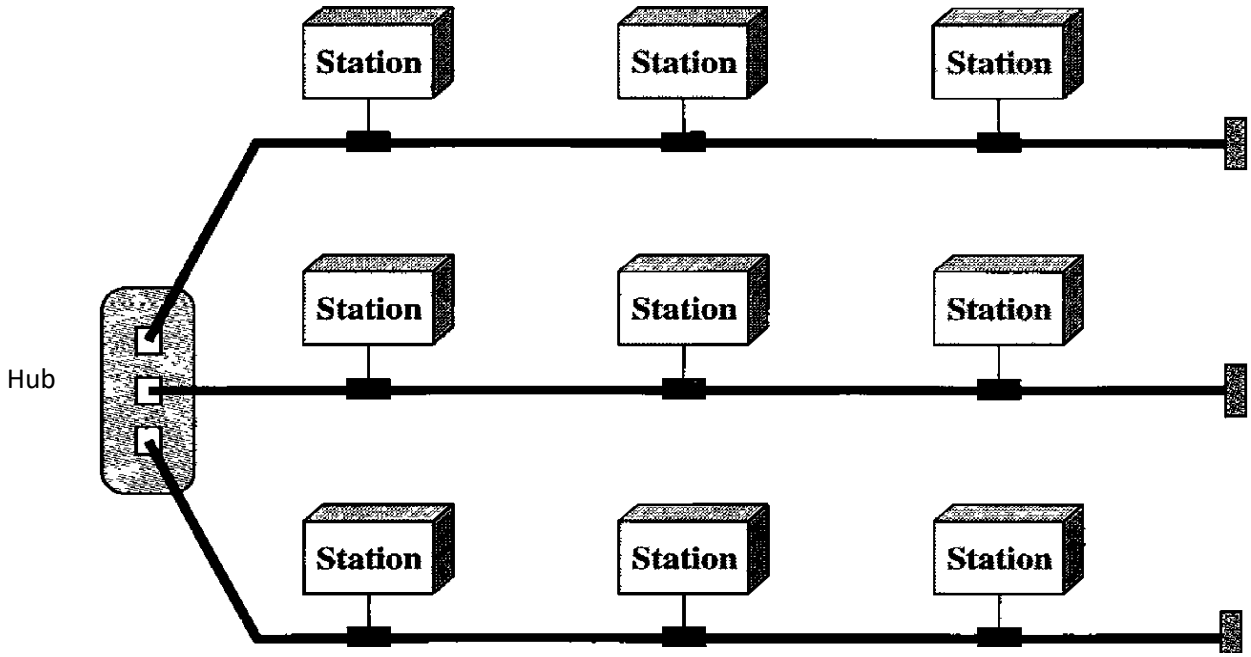
period, it can issue an alarm. The alarm alerts the network operator to the problem and its location.

Disadvantages:

Unidirectional traffic. In a simple ring, a break in the ring can disable the entire network. This weakness can be solved by using a dual ring.

5- *Hybrid Topology*

A network can be hybrid. For example, we can have a main star topology with each branch connecting several stations in a bus topology as shown in Figure below.



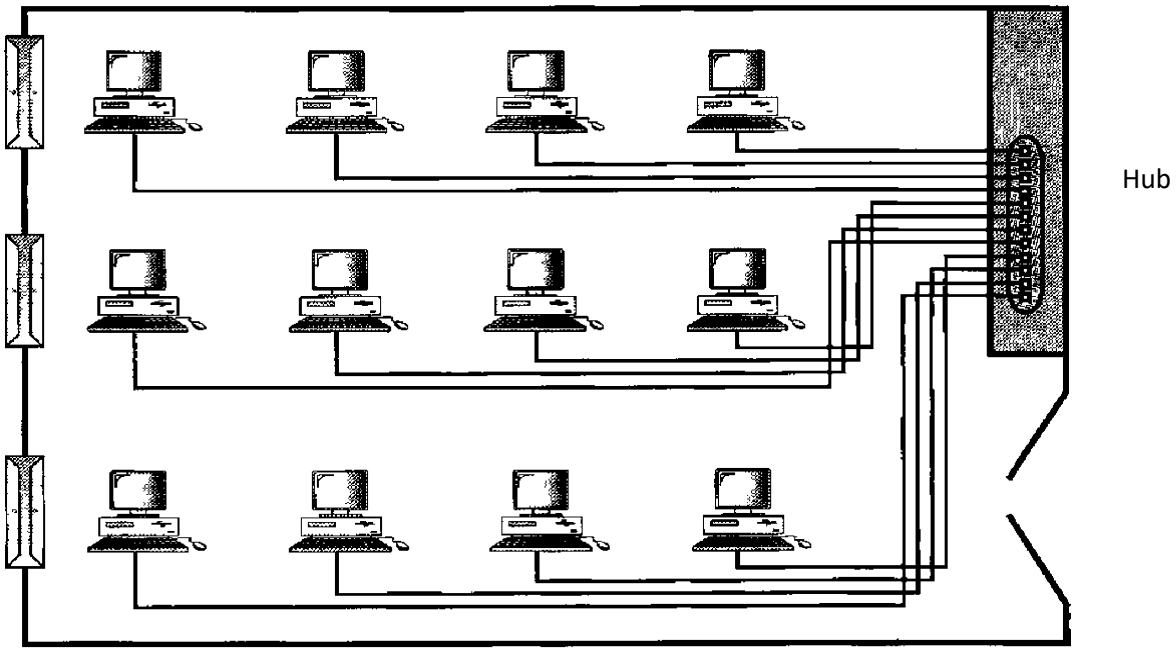
Hybrid Topology

Categories Of Networks:-

Today when we speak of networks, we are generally referring to three primary categories: **local-area networks**, **metropolitan area networks** and **wide-area networks**. The category into which a network falls is determined by its size.

1- Local Area Networks (LAN):-

A local area network (LAN) is usually privately owned and links the devices in a single office, building, or campus (see Figure below). Depending on the needs of an organization and the type of technology used, a LAN can be as simple as two PCs and a printer in someone's home office; or it can extend throughout a company and include audio and video peripherals. Currently, LAN size is limited to a few kilometers.



- Ethernet, for example, is a bus – based broadcast network with decentralized control. Computers on Ethernet can transmit wherever they want to, if two or more packets collide,

each computer just waits a random time and tries again later. LANs are designed to allow resources to be shared between personal computers or workstations. The resources to be shared can include hardware (e.g., a printer), software (e.g., an application program), or data. A common example of a LAN, found in many business environments, links a workgroup of task-related computers, for example, engineering workstations or accounting PCs. One of the computers may be given a large capacity disk drive and may become a server to clients. Software can be stored on this central server and used as needed by the whole group. In this example, the size of the LAN may be determined by licensing restrictions on the number of users per copy of software, or by restrictions on the number of users licensed to access the operating system.

In addition to size, LANs are distinguished from other types of networks by their **transmission media** and **topology**. In general, a given LAN will use only one type of transmission medium. The most common LAN topologies are bus, ring, and star.

Early LANs had data rates in the 4 to 16 megabits per second (Mbps) range. Today, however, speeds are normally 100 or 1000 Mbps.

2- Metropolitan Area Networks (MAN) :-

A metropolitan area network (MAN) is a network with a size between a LAN and a WAN. It normally covers the area inside a town or a city. It is designed for customers who need a high-speed connectivity, normally to the Internet, and have endpoints spread over a city or part of city. A good example of a MAN is the part of the telephone company network that can provide a high-speed DSL line to the customer.

Another example is the cable TV network that originally was designed for cable TV, but today can also be used for high-speed data connection to the Internet.

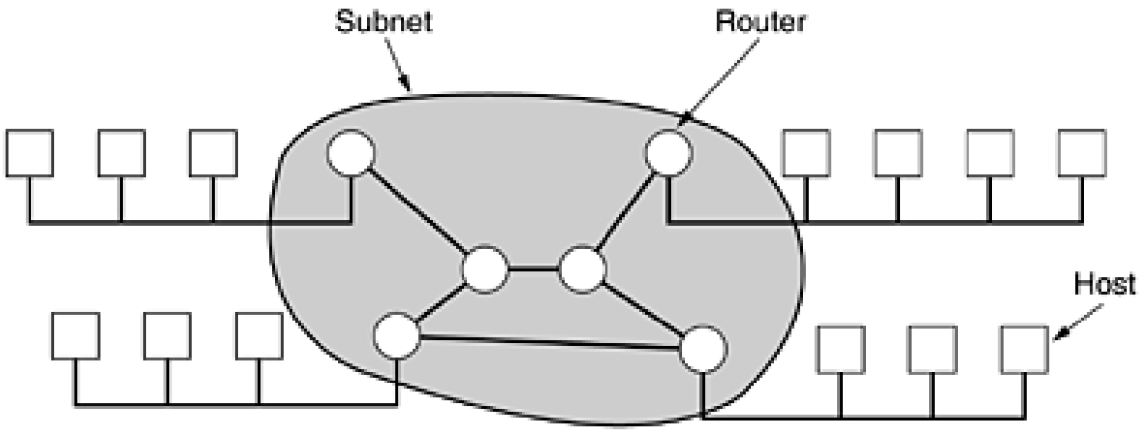
3- Wide Area Networks (WAN) :-

- WAN spans a large geographical (country or continent).
- It contains a collections of machines intended for running user programs (this machine traditionally called Host).
- Hosts are connected by a communication subnet.
- The hosts are owned by the customers, whereas the communication subnet is typically owned and operated by a telephone company or internet service provider.
- The job of the subnet is to carry messages from host to host.
- The subnet consists of distinct components:-

a) Transmission lines:- moves bits between machines (copper wire, optical fiber, or radio link).

b) Switching elements:- are specialized computers that connect three or more transmission lines. When data arrive on an incoming line, the switching elements must choose an outgoing line on which to forward them. The name **Router** is used for switching element.

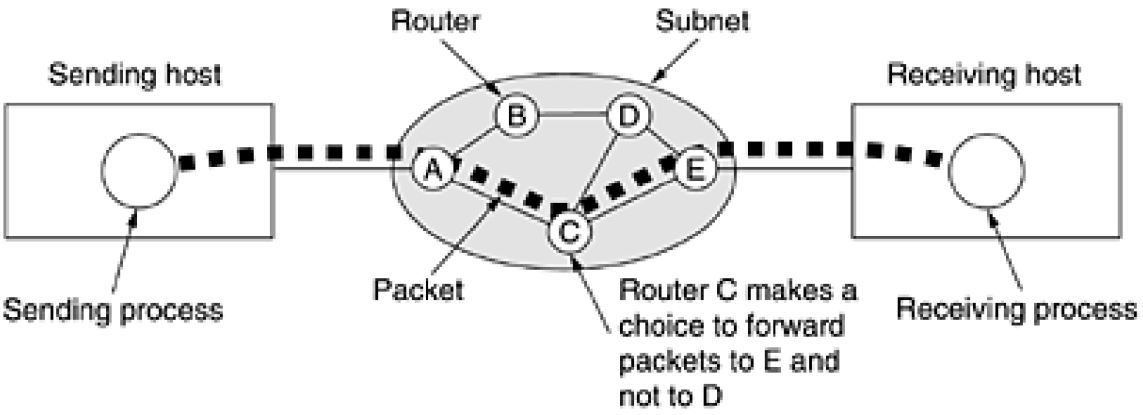
- The collection of communication lines and routers (but not the host) from the subnet.



Relation between hosts on LANs and the subnet.

- In most WANs, the network contains numerous transmission lines, each one connecting a pair of routers. If two Routers that don't share a transmission line wish to communicate, they must do this indirectly, via other routers. When a packet is sent from one router to anther via one or more intermediate routers the packet is received at each intermediate router in it is entirety, stored there until the required output line is free, and then forwarded. A subnet organized according to this principle is called **Store – and – forward** or **Packet switched subnet** . A good example of a switched WAN is the asynchronous transfer mode (ATM) network, which is a network with fixed-size data unit packets called cells.

The principle of a packet-switched WAN is so important, when a process on some host has a message to be sent to a process on some other host, the sending host first cuts the message into packets, each one bearing its number in the sequence. These packets are then injected into the network one at a time in quick succession. The packets are transported individually over the network and deposited at the receiving host, where they are reassembled into the original message and delivered to the receiving process. A stream of packets resulting from some initial message is illustrated in figure below.



A stream of packets from sender to receiver.

In this figure, all the packets follow the route ACE, rather than ABDE or ACDE. In some networks all packets from a given message must follow the same route; in others each

packet is routed separately. Of course, if ACE is the best route, all packets may be sent along it, even if each packet is individually routed.

Routing decisions are made locally. When a packet arrives at router **A**, it is up to **A** to decide if this packet should be sent on the line to **B** or the line to **C**. How **A** makes that decision is called the routing algorithm, many of them exist.

A second type of WAN can be as simple as a dial-up line that connects a home computer to the Internet. We normally refer to this type a point-to-point WAN

The point-to-point WAN is normally a line leased from a telephone or cable TV provider that connects a home computer or a small LAN to an Internet service provider (lSP). This type of WAN is often used to provide Internet access.

4- Wireless Networks:-

Wireless N.Ws can be divided into 3 main categories:-

A- System Interconnection:- is all about interconnecting the components of a computer using short – range radio. A short – range wireless network called Bluetooth used to connect components to computers without wires. In the simplest form, system interconnection networks use the master – slave paradigm. The system unit is normally the master, talking to the mouse, keyboard etc. as slave. The master tells the slaves what addresses to use, when they can broadcast, how long they can transmit, what frequencies they can use and so on.

B- Wireless LANs :- These are systems in which every computers has a radio modem and antenna with which it can communicate with other system. Wireless LANs are becoming increasingly common in small offices and homes where installing Ethernet is considered too much trouble.

C- Wireless WANs :- The radio network used for cellular telephone is an example of a low – bandwidth wireless system. This system has already gone through three generations. The first was analog and for voice only. The second was digital and for voice only. The third is digital and is for both voice and data. In a certain sense, cellular wireless network are like wireless LANs except that the distances involved are much greater and the bit rates much slower. Wireless LANs can operate at rates up to about 50 mbps over distances of ten meters. Cellular systems operate below 1 mbps over distances measured in kms.

5- Home Networks:-

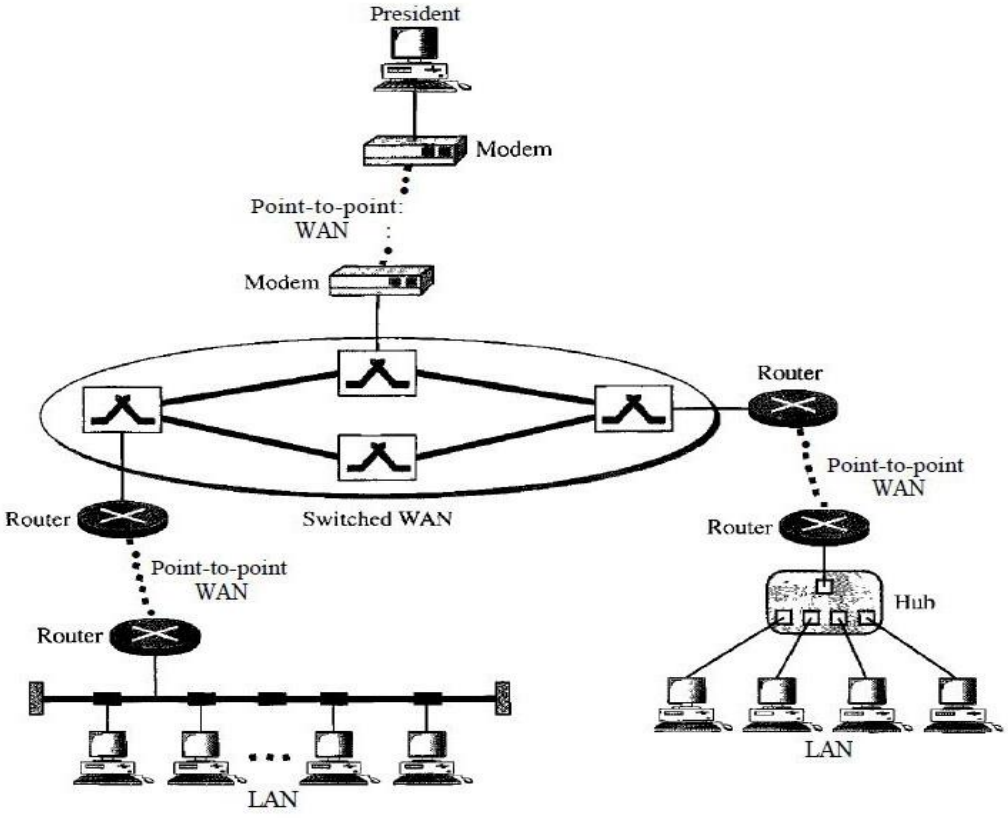
Home networks are on the Horizon. The fundamental idea is that in the future most home will be setup for networking. Every device in the home will be capable of communicating with every other device and all of them will be accessible over the Internet.

6- Interconnection of Networks: Internetwork :-

When two or more N.Ws are connected, they become an internetwork or an internet. Today, it is very rare to see a LAN, a MAN, or a LAN in isolation; they are connected to one another. When two or more networks are connected, they become an internetwork, or internet.

As an example, assume that an organization has two offices, one on the east coast and the other on the west coast. The established office on the west coast has a bus topology LAN; the newly opened office on the east coast has a star topology LAN. The president of the company lives somewhere in the middle and needs to have control over the

company from her horne. To create a backbone WAN for connecting these three entities (two LANs and the president's computer), a switched WAN (operated by a service provider such as a telecom company) has been leased. To connect the LANs to this switched WAN, however, three point-to-point WANs are required. These point-to-point WANs can be a high-speed DSL line offered by a telephone company or a cable modern line offered by a cable TV provider as shown in Figure below.



Network Software:-

The first computer N.W were designed with the hardware as the main concern and the software as an after through. This strategy no longer works. Network software is now highly structured.

PROTOCOLS AND STANDARDS

In this section, we define two widely used terms: protocols and standards. First, we define protocol, which is synonymous with rule. Then we discuss standards, which are agreed-upon rules.

Protocols

In computer networks, communication occurs between entities in different systems. An entity is anything capable of sending or receiving information. However, two entities cannot simply send bit streams to each other and expect to be understood. For communication to occur, the entities must agree on a protocol. A protocol is a set of rules that govern data communications. A protocol defines what is communicated, how it is communicated, and when it is communicated. The key elements of a protocol are syntax, semantics, and timing.

- Syntax. The term syntax refers to the structure or format of the data, meaning the order in which they are presented. For example, a simple protocol might expect the first 8 bits of data to be the address of the sender, the second 8 bits to be the address of the receiver, and the rest of the stream to be the message itself.

- Semantics. The word semantics refers to the meaning of each section of bits. How is a particular pattern to be interpreted, and what action is to be taken based on that interpretation? For example, does an address identify the route to be taken or the final destination of the message?

- Timing. The term timing refers to two characteristics: when data should be sent and how fast they can be sent. For example, if a sender produces data at 100 Mbps but the receiver can process data at only 1 Mbps, the transmission will overload the receiver and some data will be lost.

Standards

Standards are essential in creating and maintaining an open and competitive market for equipment manufacturers and in guaranteeing national and international interoperability of data and telecommunications technology and processes. Standards provide guidelines to manufacturers, vendors, government agencies, and other service providers to ensure the kind of interconnectivity necessary in today's marketplace and in international communications.

Protocol Hierarchies:-

To reduce their design complexity, most N.Ws is organized as a stack of layers or levels, each one built upon one below it. The purpose of each layer is to:-

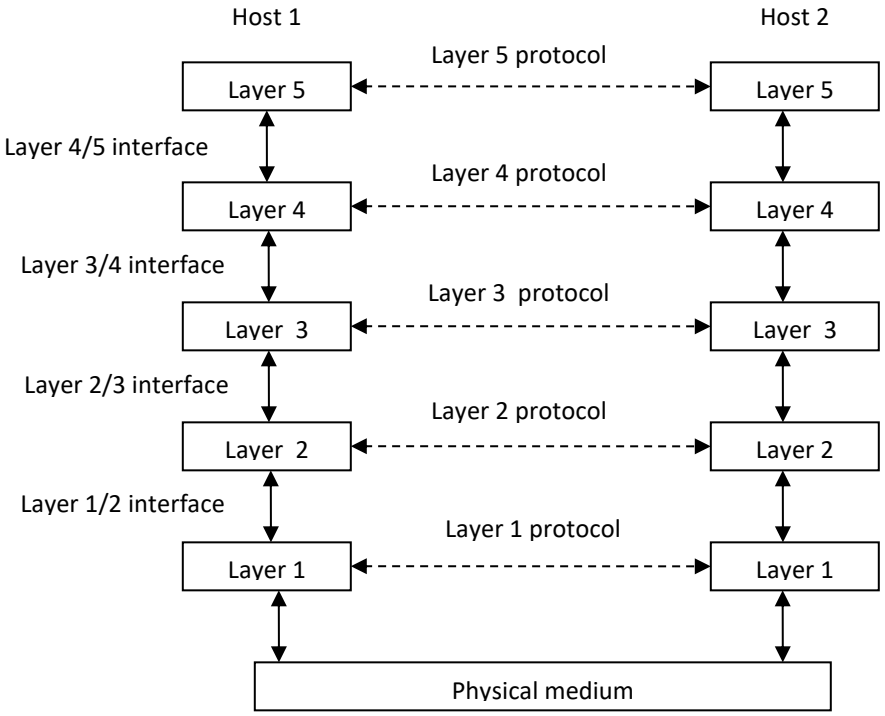
1- Offer certain services to the higher layers.
2- Shielding those layers from the details of how the offered services are actually implemented.

In a sense, each layer is a kind of virtual machine offering certain services to the layer above it.

The fundamental idea is that a particular piece of software (or hardware) provides a service to its user but keeps the details of its internal state and algorithms hidden from them.

Layer n on one machine carries on a conversation with a layer n on another machine. The rules and conventions used in this conversation are collectively known as the layer n protocol.

Figure below show five – layers networks and layers and protocols and interfaces of these N.W.



The entities comprising the corresponding layers on different machines are called peers. The peers may be processes, hardware, devices, or even human being. In the words, it is the peers that communicate by using the protocol.

In reality, no data are directly transferred from layer n on one machine to layer n on another machine, instead, each layer passes data and control the lowest layer is reached. Below layer 1 is the physical medium through which actual communications occurs.

Between each pair of adjacent layers is an interface. The interface defines which primitive operations and services the lower layers makes available to the upper one. Network designer must define clear interface between the layers, Doing this requires that each layer performer a specific collection of well – understood functions. In addition to minimizing the amount

of information that must be passed between layers. Clear – cut interface make it simpler to replace the implementation of one layer with a completely different implementation (e.g.:- all the telephone lines are replaces by satellite channels).

A set of layers and protocols is called network architecture.

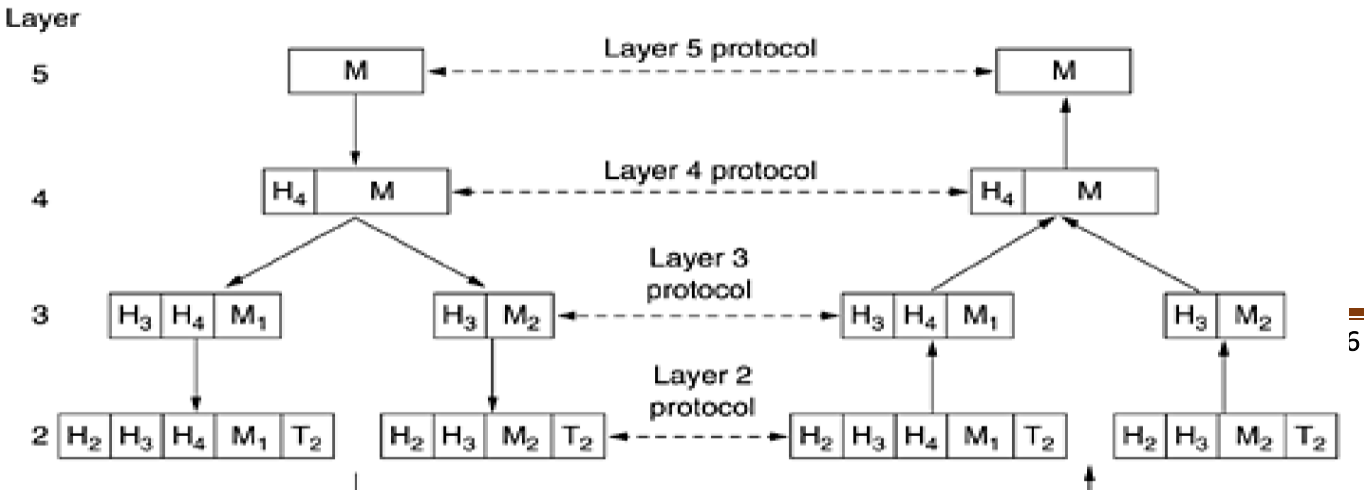
Now consider following technical example:-

How to provide communication on the top layer of the five – layer network. ?

A message **M** is produced by an application process running in layer 5 and given to layer 4 for transmission.

Layer 4 puts header in front of the message to identify the message and passes the result to layer 3. The header include control information, such as sequence numbers, to allow layer 4 on the destination machine to deliver messages in the right order if the lower layers don't maintain sequence.

In some layers, headers can also contain size, times, and other control informational fields.



Always there is limit to the size of messages transmitted by layer 3, so layer 3 must break incoming messages into smaller units, packets, prep ending a layer 3 header to each packet. At the receiving machine the messages moves upward from layer to layer with headers being stripped off as it progresses. None of the headers for layers below n are passed up to layer n.

Connection – oriented and Connectionless Services:-

Layers can offer two different types of services to the layers above them:-

Connection – oriented services:- is modeled after the telephone system. To talk to someone, you pick up the phone, dial the number, talk and then hang up. Similarly, to use a connection – oriented services, the service user first establishes a connection, uses the connection, and then releases the connection. The essential aspect of a connection is that it acts like a tube: the sender pushes object (bits) in at one end, and the receiver takes them out at the other end. In most cases the order is preserved so that the bits arrive in the order they were sent.

Connectionless Services:- is modeled after postal system. Each message (letter) carries the full destination address, and each one is routed through the system independent of all the others.

Each service can be characterized by a quality is Service. Some services are reliable in the sense that they never lose data. Usually a reliable services is implemented by having the receiver acknowledge the receipt of each message so the sender is sure that it arrived. The acknowledgment process introduces overhead and delays. A typical situation in which reliable connection – oriented service is appropriate is file transfer.

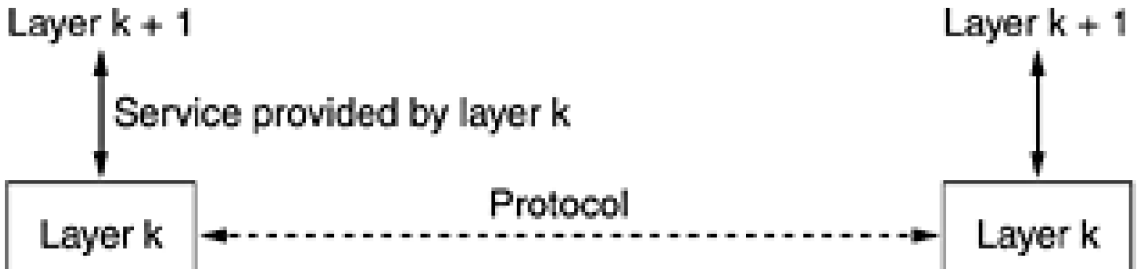
For some application, the delays by an acknowledgment are unacceptable. One such application is digitized voice traffic. Similarly video conference.

The Relationship of Services to Protocols

Services and protocols are distinct concepts, although they are frequently confused. This distinction is so important, however, that we emphasize it again here. A service is a set of primitives (operations) that a layer provides to the layer above it. The service defines what operations the layer is prepared to perform on behalf of its users, but it says nothing at all about how these operations are implemented. A service relates to an interface between two layers, with the lower layer being the service provider and the upper layer being the service user.

A protocol, in contrast, is a set of rules governing the format and meaning of the packets, or messages that are exchanged by the peer entities within a layer. Entities use protocols to implement their service definitions. They are free to change their protocols at will, provided they do not change the service visible to their users. In this way, the service and the protocol are completely decoupled.

In other words, services relate to the interfaces between layers, as illustrated in figure below. In contrast, protocols relate to the packets sent between peer entities on different machines, it is important not to confuse the two concepts.



An analogy with programming languages is worth making. A service is like an abstract data type or an object in an object-oriented language. It defines operations that can be performed on an object but does not specify how these operations are implemented. A protocol relates to the implementation of the service and as such is not visible to the user of the service.

Reference Models:-

There are two important network architecture, the **OSI** reference model and the **TCP/IP** reference model.

- The OSI Model :-

This model is based on a proposed developed by the international standard organization (ISO). The model is called OSI (Open System Interconnecting) because it deals with connecting open system [Systems that are open for communication with other systems]. An open system is a set of protocols that allows any two different systems to communicate regardless of their underlying architecture.

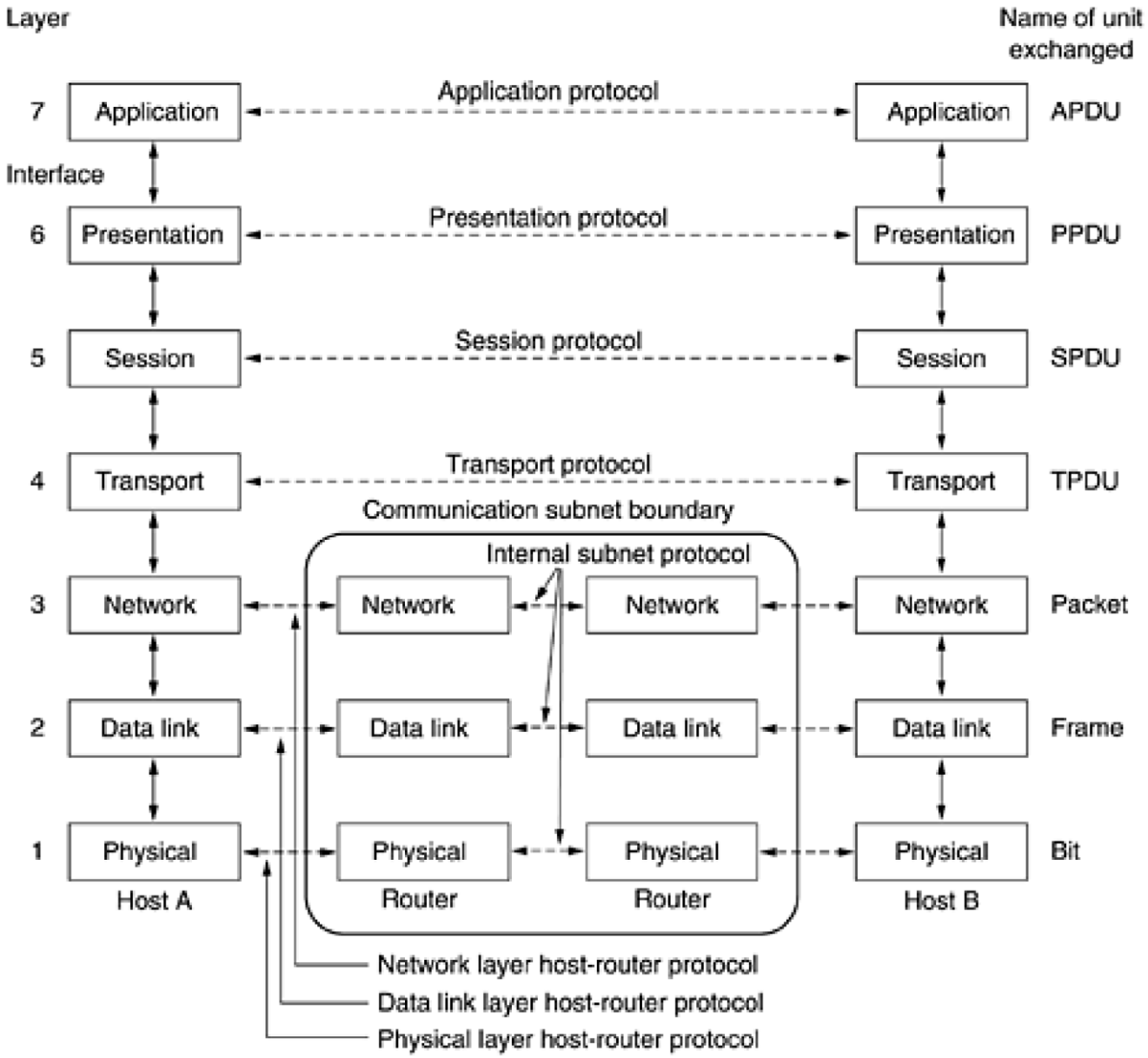
The purpose of the OSI model is to show how to facilitate communication between different systems without requiring changes to the logic of the underlying hardware and software. The OSI model is not a protocol; it is a model for understanding and designing a network architecture that is flexible, robust, and interoperable.

The OSI model is built of seven ordered layers. As shown in figure (OSI reference model) , as the message travels from device A to device B it may pass through many intermediate nodes. These intermediate nodes usually involve only the first three layers of the OSI model.

The seven layers can be thought of as belonging to three subgroups. Layers (1, 2, and 3) are the network support layers, they deals with the physical aspects of moving data from one device to another (such as electrical specifications, physical connections, physical addressing, and transport timing and reliability).

Layers (5, 6, and 7) can be thought of as the user support layers, the allow interpretability among unrelated software systems. Layers 4, the transport layer, links the two subgroups and ensures that what the lower layers have transmitted is in a form that the upper layers can use. The upper OSI layers are almost always implemented in software;

lower layers are a combination of hardware and software except for the physical layer, which is mostly hardware.



OSI reference model

Function of The Layers :-

1- Physical layer :-

The physical layer coordinates the functions required to transmit a bit stream over physical medium. It deals with the mechanical and electrical specifications of the primary connections, such as cables, connectors, and signaling options that physically ink two nodes on a network.

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 0s 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 (how Os and I s are changed to signals).
- 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.

- ❖ Transmission mode. The physical layer also defines the direction of transmission between two devices: simplex, half-duplex, or full-duplex.

2- Data link layer :-

The main task of Data link layer is to transmission facility into a line that appears free of undetected transmission errors to the network layer. It accomplishes this task by having the sender break up the data into data frames, and transmit the frames sequentially. If the service is reliable, the receiver confirms correct receipt of each frame by sending back an acknowledgment frame.

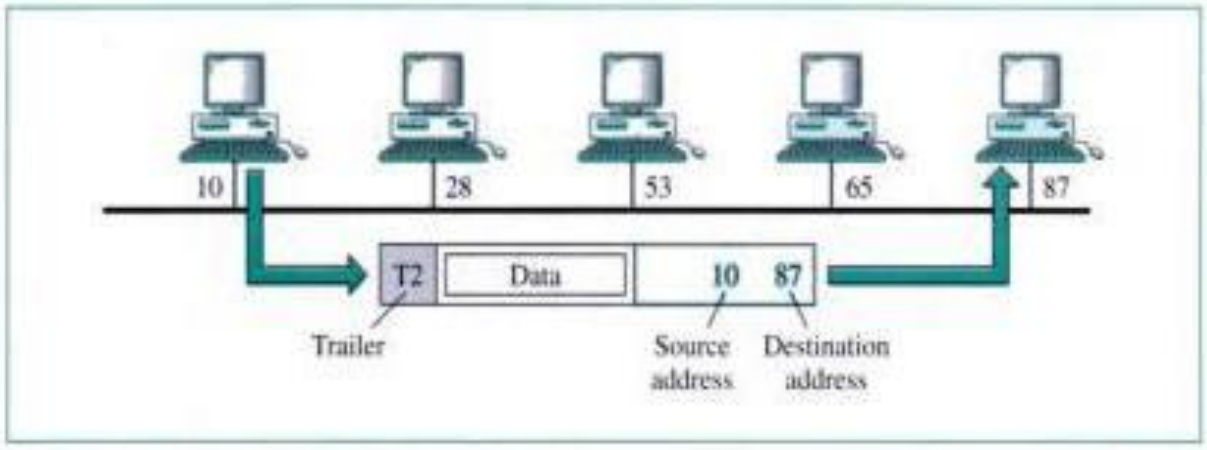
Specific responsibilities of the data link layer include the following :-

- o Node – to – Node delivery.
- o Framing. The data link layer divides the stream of bits received from the network layer into manageable data units called frames.
- o 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.
- o Access control :- When two or more devices are connected to the same link, it necessary to determine which device has control over the line at any given time.
- o Flow control :- To avoid overwhelming the receiver, the data link layer regulates the amount of data that can be transmitted at one time. It adds identifying numbers to enable the receiving node to control the ordering of the frames.
- o Error handling : Data link layer provide for data recovery, usually by having the entire frame retransmitted.

- Synchronization :- Headers contain bits to alert the receiving station that a frame is arriving. These bits allow the receiver to synchronize its timing to that of the transmission (Know duration of each bit). Trailers contain bits for error control and also bits that indicate the frame has ended, and that anything to follow is either a new frame or an idle channel.

The Data link layer is subdivided into two sub layers : **Logical Link Control (LLC)**, and **Media Access Control (MAC)**.

Example:- in figure below, node with physical address 10 sends a frame to a node with physical address 87. The two nodes are connected by a link. At the data link level, this frame contains physical addresses in the header. These are the only addresses needed. The rest of the header contains other information needed at this level. The trailer usually contains extra bits needed for error detection.



3- Network Layers :-

The network layer is responsible for the source – to – destination delivery of a packet across multiple network links.

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.

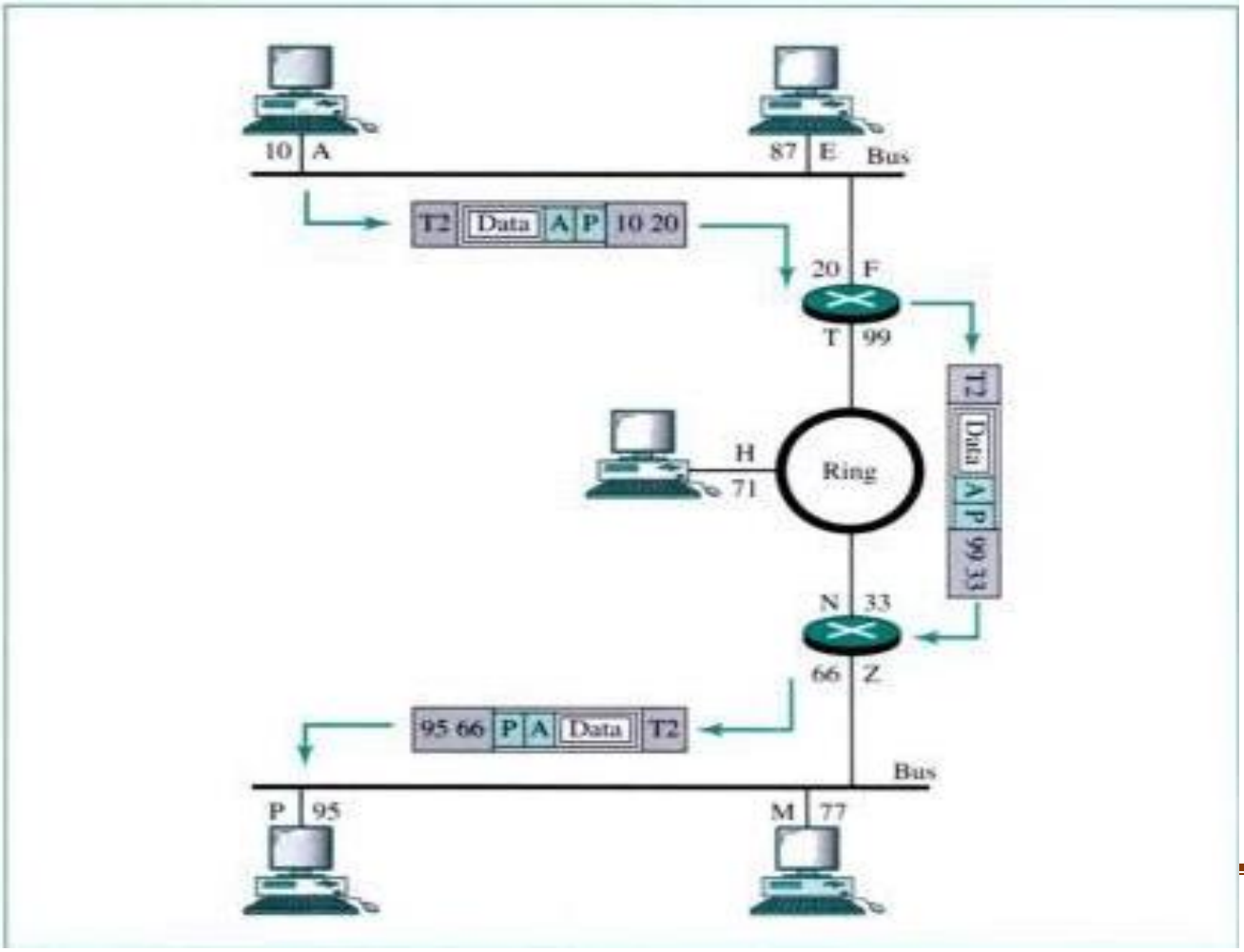
The network layer provider two related services : **Switching** and **Routing**.

Switching refers to temporary connections between physical links, resulting in longer links for network transmission [e.g. telephone conversation].

Routing means selecting the best path for sending a packet from one point to another when more than one path is available.

Routing and switching require the addition of a header that includes, the source and destination addresses of the packet. Data link addresses (physical Addresses) changes as a frame moves from one node to the next. While Network layer addresses [Logical Addresses] don't change during transmission.

Example: in the figure below, we want to send data from a node with network address **A** and physical address **10** located on one LAN, to a node with network address **P** and physical address **95** located on another LAN.



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.

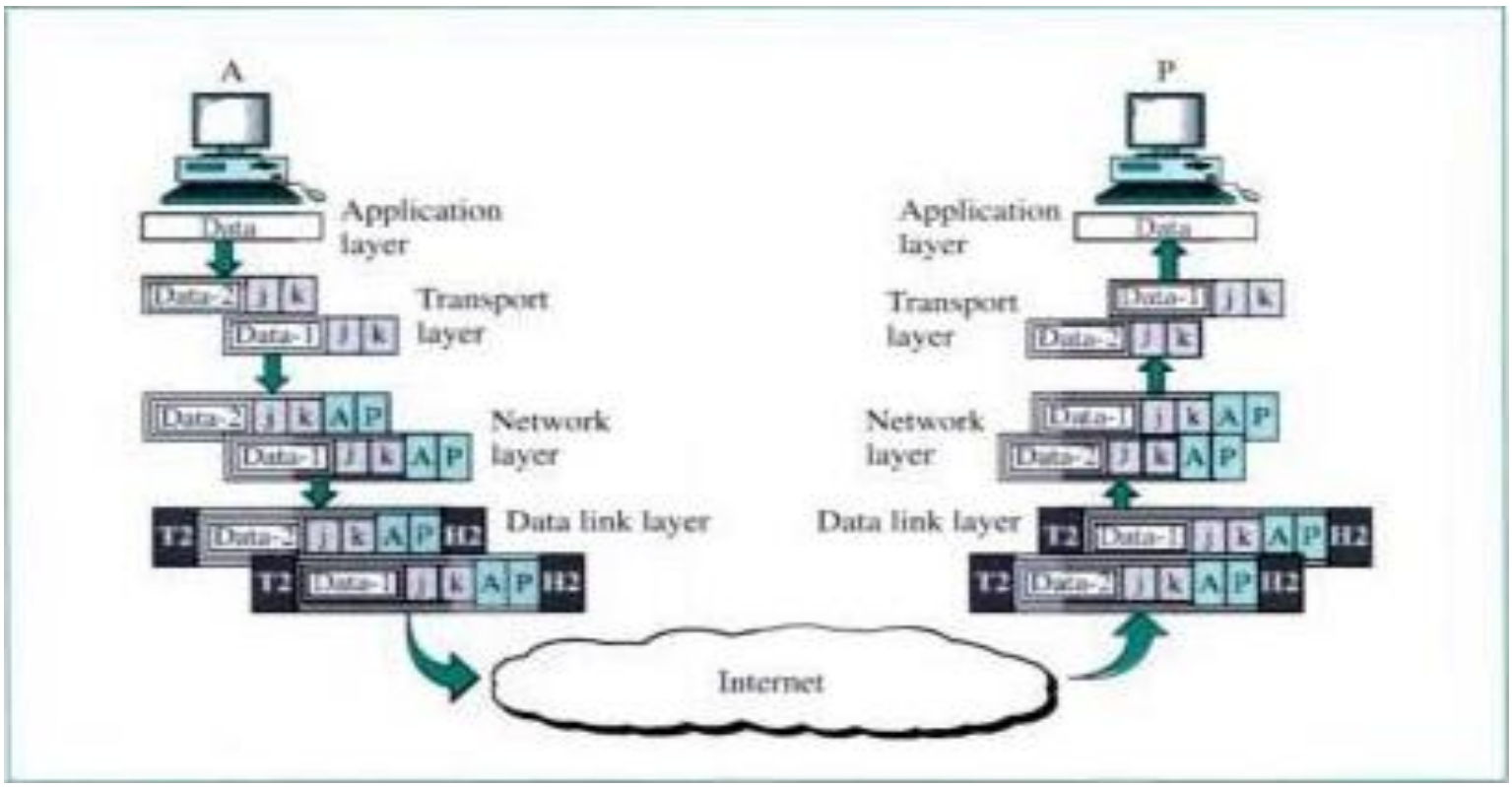
Specific responsibilities of the transport layer include the following :-

- End – to – End message delivery. Overseeing the transmission and arrival of all packets of a message at the destination point.
- 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.

- 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.

Example:- figure below shows example of transport layer communication, data coming from an upper layers have port addresses **j** and **k** , (**j** as address for sending process, and **k** is the address of receiving process). Since the data are larger than the network layer can handle, the data are split into two packets, each packet retaining the port addresses (**j** and **k**), then in the network layer, network addresses **A** and **P** are added to each packet. The packets can travel on different path and arrive on the destination either on order or out of order.



5- Session Layer :-

The session layer allows users on different machines to establish sessions between them.

Sessions offer various services, including:-

- dialog control (keeping track of whose turn it is to transmit).
- token management (preventing two parties from attempting the same critical operation at the same time).
- Synchronization. The session layer allows a process to add checkpoints, or synchronization points, to a stream of data. For example, if a system is sending a file of 2000 pages, it is advisable to insert checkpoints after every 100 pages to ensure that each 100-page unit is received and acknowledged independently. In this case, if a crash happens during the transmission of page 523, the only pages that need to be resent after system recovery are pages 501 to 523. Pages previous to 501 need not be resent.
- Graceful Close: - Ensuring that the exchange has been completed appropriately before the session closes.

6- Presentation layer :-

The presentation layer is concerned with the syntax and semantics of the information exchanged between two systems.

Specific responsibilities of this layer include:-

- Translation: - changing the format of a message from that used by sender into one mutually acceptable for transmission. Then at the destination changing the format into the one understood by the receiver.
- Encryption: - encryption and decryption of data for security purposes.
- Compression :- Compression and decompressing data to make transmission more efficient.
- Security :- Validating passwords and log – in codes.

7- Application layer:

The application layer enables the user, whether human or software, to access the network. It provides user interfaces and support for some services.

The application layer contains a variety of protocols that are commonly needed by users. One widely-used application protocol is HTTP (HyperText Transfer Protocol), which is the basis for the World Wide Web. When a browser wants a Web page, it sends the name of the page it wants to the server using HTTP. The server then sends the page back. Other application protocols are used for file transfer, electronic mail, and network news.

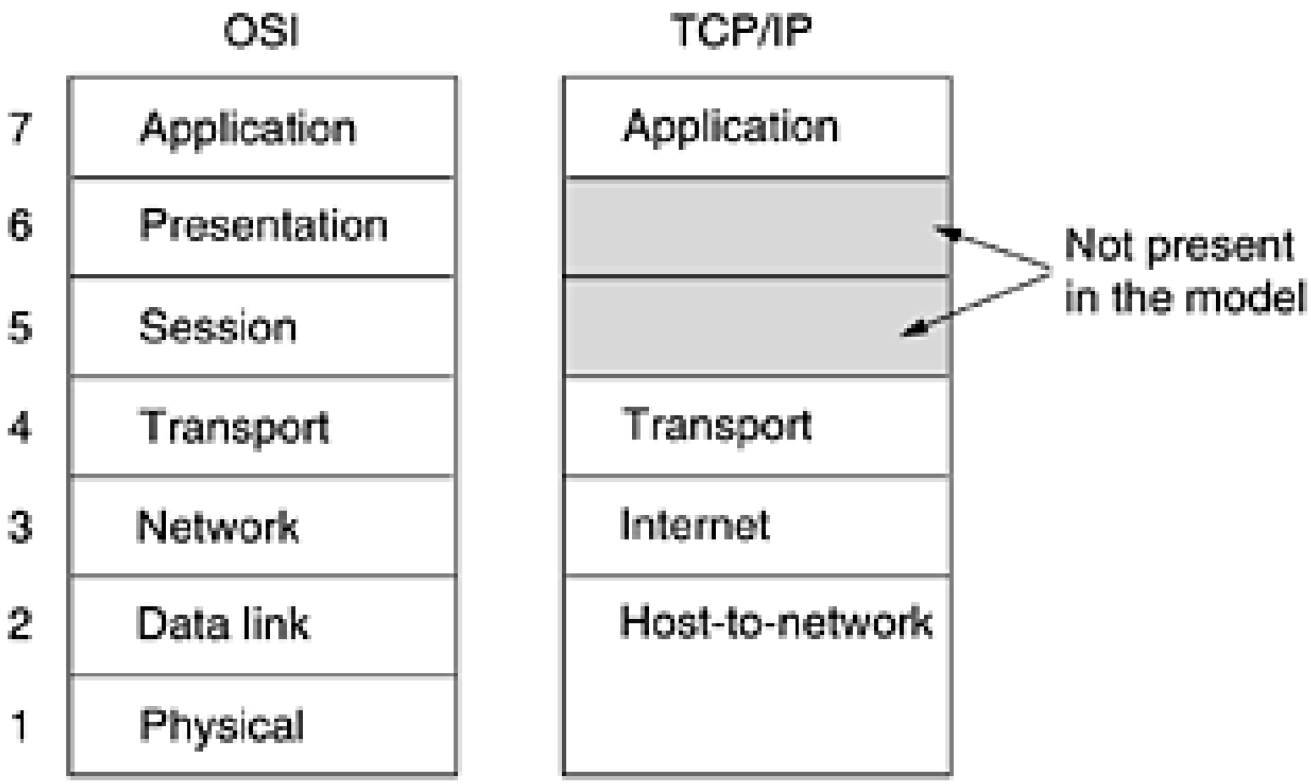
The TCP/IP Reference Model

Let us now turn from the OSI model to the reference mode; used in the grandparent of all wide area computer networks. The ARPANET, and its successor, the world wide internet.

The ARPENET was research network sponsored by the DoD (U.S Department of Defense). It eventually connected hundreds of universities and government installation, using leased telephone lines. When satellite and radio networks were added later, the exiting protocols had trouble inter working with them, so a new reference architecture was needed. Thus, the ability to connect multiple networks in a seamless way was one of the major design goals from the very beginning.

This architecture later became known as the TCP/IP Reference Model, after its two primary protocols. The TCP/IP 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.



The Host-to-Network Layer

Below the internet layer is a great void. The TCP/IP reference model does not really say much about what happens here. Except to point out that the host has to connect to the network using some protocol so it can send IP packets to it. This protocol is not defined and varies from host has to host and network to network. Books and papers about the TCP/IP model rarely discuss it.

The Internet Layer

internet layer, is the linchpin that holds the whole architecture together. Its job is to permit hosts to inject packets into any network and have them travel independently to the destination (potentially on a different network). They may even arrive in a different order than they were sent, in which case it is the job of higher layers to rearrange them. If in-order

delivery is desired. Note that "internet" is used here in a generic sense, even though this layer is in the internet.

The internet layer defines an official packet format and protocol called IP (Internet Protocol). IP, in turn, uses four supporting protocols: ARP, RARP, ICMP, and IGMP.

The Internetworking Protocol (IP)

The job of the internet layer is to deliver IP packets where they are supposed to go. Packet routing is clearly the major issue here, as is avoiding congestion. For these reasons, it is reasonable to say that the TCP/IP internet layer is similar in functionality to the OSI network layer.

IP is an unreliable and connectionless protocol, -a best-effort delivery service.

The term best effort means that IP provides no error checking or tracking. IP assumes the unreliability of the underlying layers and does its best to get a transmission through to its destination, but with no guarantees.

IP transports data in packets called datagrams, each of which is transported separately.

Datagrams can travel along different routes and can arrive out of sequence or be duplicated.

IP does not keep track of the routes and has no facility for reordering datagrams once they arrive at their destination. The limited functionality of IP should not be considered a weakness.

Address Resolution Protocol

The Address Resolution Protocol (ARP) is used to associate a logical address with a physical address.

Reverse Address Resolution Protocol

The Reverse Address Resolution Protocol (RARP) allows a host to discover its Internet address when it knows only its physical address.

Internet Control Message Protocol

The Internet Control Message Protocol (ICMP) is a mechanism used by hosts and gateways to send notification of datagram problems back to the sender.

Internet Group Message Protocol

The Internet Group Message Protocol (IGMP) is used to facilitate the simultaneous transmission of a message to a group of recipients.

The Transport Layer

The layer above the internet layer in the TCP/IP model is now usually called the transport layer. It is designed to allow entities in the source and destination hosts to carry on a conversation. Traditionally the transport layer was represented in TCP/IP by two protocols: TCP and UDP. IP is a host-to-host protocol, meaning that it can deliver a packet from one physical device to another. UDP and TCP are transport level protocols responsible or delivery of a message from a process (running program) to another process.

1- TCP (Transmission Control Protocol)

Is a reliable connection – oriented protocol that allows a byte stream originating on one machine to be delivered without error on any other machine in the internet. It fragments the incoming byte steam into discrete messages called segments, Each segment includes a sequence number for reordering after receipt, together with an acknowledgment number for the segments received. Segments are carried across the internet inside of IP datagrams. At the receiving end, TCP collects each datagram as it comes in and reorders the transmission based on sequence numbers.TCP also handles flow control to make sure a fast sender cannot swamp a slow receiver with more messages it can handle.

2- UDP (User Datagram Protocol)

The User Datagram Protocol (UDP) is the simpler of the two standard TCP/IP transport protocols. UDP (User Datagram Protocol), is an unreliable, connectionless protocol for applications that do not want TCP's sequencing or flow control and wish to provide their own. It also widely used for one – shot, silent – server – type request – reply queries and applications in which prompt delivery is more important than accurate delivery, such as transmitting speech or video. It is a process-to-process protocol that adds only port addresses, checksum error control, and length information to the data from the upper layer.

The application Layer

The TCP/IP model does not have session or presentation layers. No need for them was perceived, so they were not included. Experience with the OSI model has proven this view correct: they are of little use to most applications.

On top of the transport layer is the application layer. It contains the higher level protocols. The early ones included virtual terminal (TELNET), file transfer (FTP), and electronic mail (SMTP).

The virtual terminal protocol allows a user on one. Machine to log onto a distant machine and work there. The file transfer protocol provides a way to move data efficiently from one machine to another. Electronic mail was originally just a kind of the file transfer, but later a specialized protocol (SMTP) was developed for it. Many other protocols have been added to these over the years. The Domain Name System (DNS) for mapping host names onto their network addresses, NNTP, the protocol for moving USENET news articles around, and HTTP, the protocol for fetching pages on the World Wide Web, and many others.

A comparison of the OSI and TCP/IP Reference Models

The OSI and TCP/IP reference models have much in common. Both are based on the concept of a stack of independent protocols. Also the functionality of the layers is roughly similar. For example, in both models the layers up through and including the transport layer are there to provide an end-to-end, network – independent transport service to processes wishing to communicate. These layers from the transport provider. Again in both models, the layers above transport are application-oriented users of the transport service.

Despite these fundamental similarities, the two models also have many differences. In this section we will focus on the key differences between the two references model. It is important to note we are comparing the reference model here. Not the corresponding protocol stacks. The protocol themselves will be discussed later. For an entire book comparing and contrasting TCP/IP and OSI, see (piscitello and Chapin, 1993).

There concepts are central to the OSI model:

1. services
2. interfaces
3. protocols.

Probably the biggest contribution of the OSI model is to make the distinction between these three concepts explicit. Each layer performs some services for the layer above it. The service definition tells what the layer does, not how entities above it access it or how the layer works. It defines the layer's semantics.

A layer's interface tells the processes above it how to access it. It specifies what the parameters are and what results to expect. It, too says nothing about how the layer works inside.

Finally, the peer protocols tells the layer's own business. It can use any protocols it wants t, as long as it gets the job done (i.e., provides the offered services). It can also change them at will without affecting software in higher layers.

These ideas fit very nicely with modern ideas about object-oriented programming. An object, like a layer, has a set of methods (operations) that processes outside the object can invoke. The semantic of these methods define the set of services that the object offers. The methods parameters and results from the object's interface. The code internal to the objects is its protocol and is not visible or of any concern outside the object.

The TCP/IP model did not originally clearly distinguish between service, interface, and protocol, although people have tried to retrofit it after the fact to make it more OSI-like. For example, the only real services offers by the internet by the internet layer are SEND IP PACKET and RECEVICE IP PACKET.

As a consequence, the protocol in the OSI model are better than in the TCP/IP model and can be replaced relatively easily as the technology changes. Being able to make such is one of the main purposes of having layered protocols in the first place.

The OSI reference model was devised before the corresponding protocols were invented. This ordering means that the model was not.

Introduction :

- A major concern of the physical layer is moving information in the form of electromagnetic Signals across a transmission medium.

- Encoder create a stream of 1's and 0s that tells the receiving device how to reconstruct data to its original form.

- even 1s and 0s cannot be sent as such across network links. They must be further converted into a form that transmission media can accept.

- Data stream of 1s and 0s must be turned into energy in the form of electromagnetic signal.

Analog And Digital :-

- Both Data and the signals that represent them can take either analog or digital form.

- Analogue refers to something that is continuous [a set of specific point of data and all possible points between them.

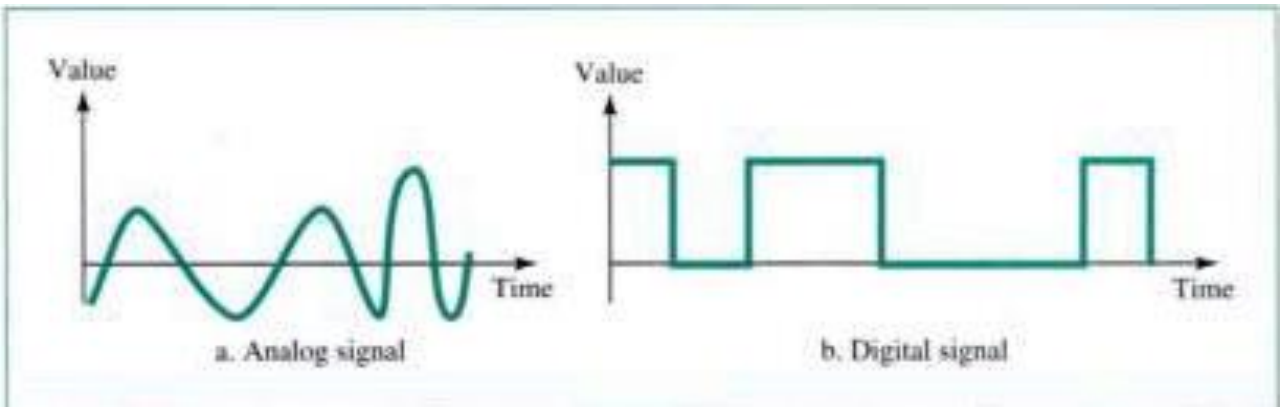
- Digital refers to something that is discrete.

- An analog signal is continuous wave from that changes smoothly over time [include infinite number of values].

- A digital signal, is discrete [it can have only a limited number of defined values, often as 1,0].

- Signals represented by plotting them on a pair of perpendicular axes. The vertical axis represents the value or strength of a signal. The horizontal axis represents the time.

- Figure illustrates comparison between analog and digital.



- Analog signal changes continuously with respect to time, while the digital signal changes instantaneously.

Signals can be analog or digital. Analog signals can have an infinite number of Values in a range; digital signals can have only a limited number of values.

- Aperiodic And Periodic Signal :-

Both analog and digital signal can be of two forms periodic and aperiodic (nonperiodic) . A periodic signal completes a pattern within a measurable time frame, called a period (T), 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.

Both analog and digital signals can be periodic or nonperiodic. In data communications, we commonly use periodic analog signals (because they need less bandwidth, and nonperiodic digital signals (because they can represent variation in data.

An aperiodic signals can be decomposed into an infinite number of periodic signals.

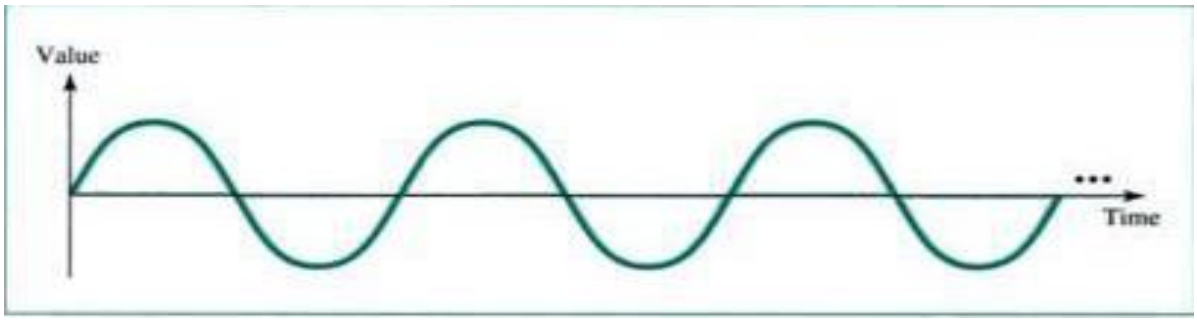
- Periodic Analog Signals :-

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 waves are the most fundamental form of a periodic analog signal.

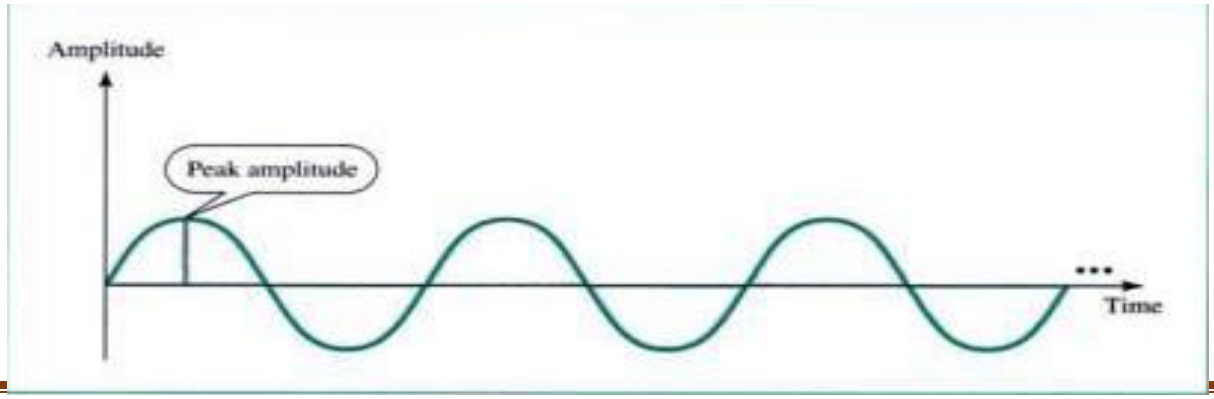
Sine wave can be fully describes by three characteristics.
Peak amplitude, **period or frequency** and **phase**.



Sine Wave

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.



Peak amplitude

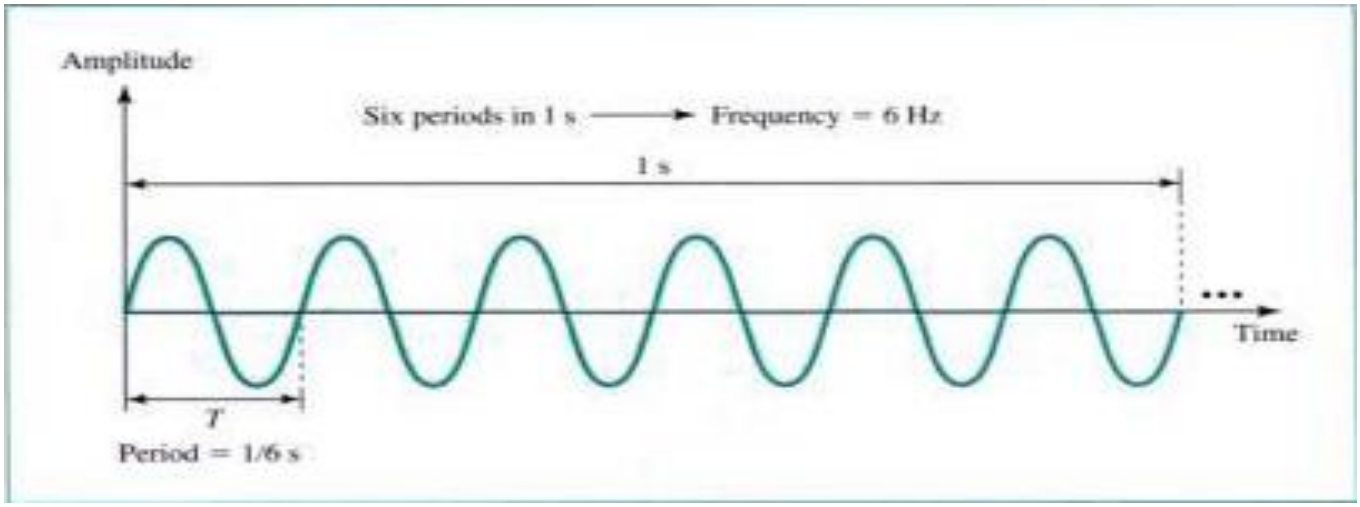
Period and Frequency :-

Period : is the amount of time (in seconds) it takes a signal to complete one cycle.

Frequency : is the number of cycle per second.

- ✓ Frequency and period are inverse of each other :

  f = 1 / T and T= 1 / f



Frequency and Period

- ✓ Frequency is expressed in Hertz (Hz).
- ✓ Period is expressed in seconds.

Unit	*Equivalent*	*Unit*	*Equivalent*
Seconds (s)	1 s	hertz (Hz)	1 Hz
Milliseconds (ms)	10^{-3} s	kilohertz (KHz)	10^{3} Hz
Microseconds (µs)	10^{-6} s	megahertz (MHz)	10^{6} Hz
Nanoseconds (ns)	10^{-9} s	gigahertz (GHz)	10^{9} Hz
Picoseconds (ps)	10^{-12} s	terahertz (THz)	10^{12} Hz

Example :
The power we use at home has a frequency of 60 Hz (50 Hz in Europe). The period of this sine wave can be determined as follows:

$T = 1 / f \rightarrow 1 / 60 = 0.0166\ S \quad = 0.0166 \times 103\ ms = 16.6\ ms$

This means that the period of the power for our lights at home is 0.0116 s, or 16.6 ms.
Our eyes are not sensitive enough to distinguish these rapid changes in amplitude.

Ex : A sine wave has a frequency of 8 KHz. What its period?

$T = 1 / f = 1 / 8000 = 0.000125\ S = 125\ \mu s$

Ex : A sine wave complete one cycle in 25 μs what is its frequency ?

$f = 1 / T \rightarrow 1 (25 * 10^{-6}) = 40{,}000 = 40\ KHz$

Express a period of 100 ms in microseconds.

If a signal does not change at all, its frequency is zero.

If a signal changes instantaneously, its frequency is infinite.

Phase:-

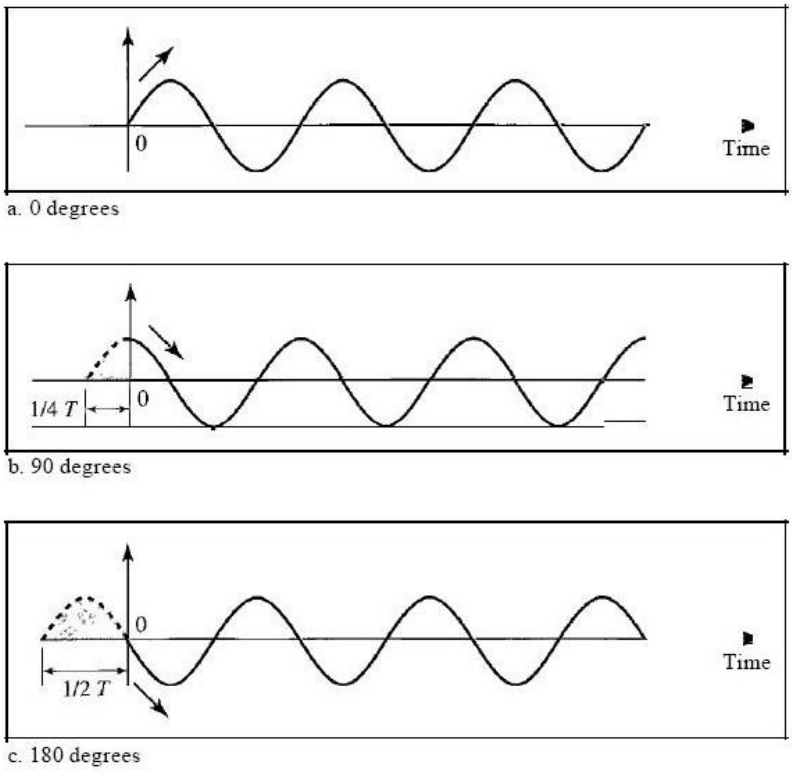
The term phase describes the position of the wave from relative to time zero.

If we think of the waves as something that can be shifted backward along the time axis, phase describe the amount of that shift. It indicates the statues of the first cycle.

- Phase is measured in degree.

Notes on Phase

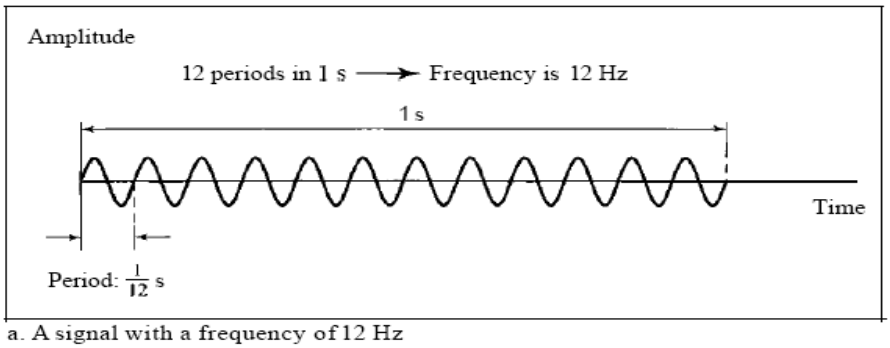
Figure 3.5 *Three sine waves with the same amplitude andfrequency, but different phases*



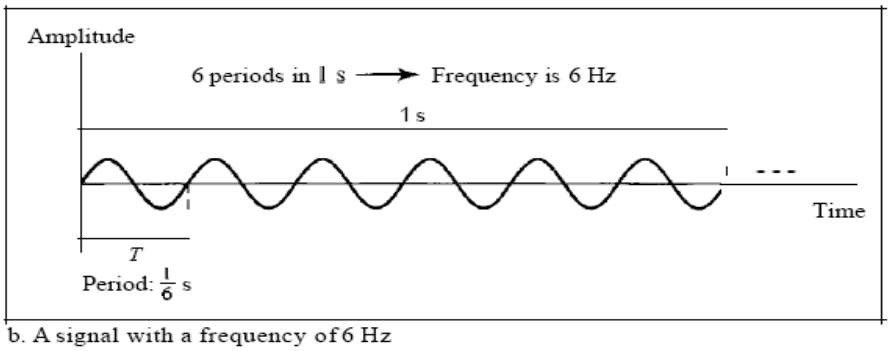
Looking at Figure 3.5, we can say that

1. A sine wave with a phase of 0° starts at time 0 with a zero amplitude. The amplitude is increasing.
2. A sine wave with a phase of 90° starts at time 0 with a peak amplitude. The amplitude is decreasing.
3. A sine wave with a phase of 180° starts at time 0 with a zero amplitude. The amplitude is decreasing.

Figure 3.4 *Two signals with the same amplitude and phase, but different frequencies*



a. A signal with a frequency of 12 Hz



b. A signal with a frequency of 6 Hz

Time and Frequency Domains

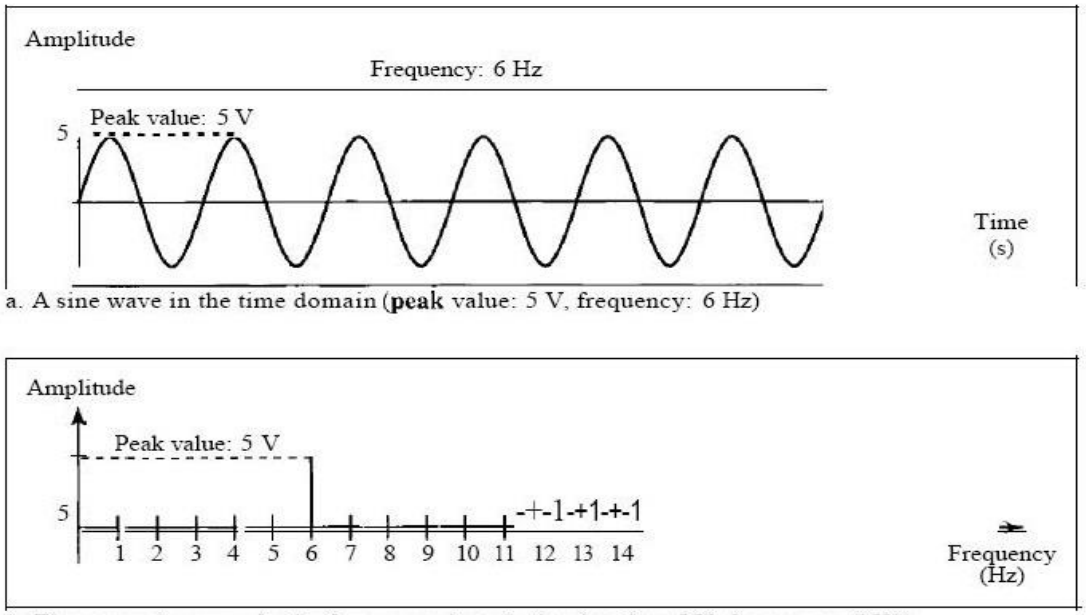
A sine wave is comprehensively defined by its amplitude, frequency, and phase. We have been showing a sine wave by using what is called a time-domain plot. The time-domain plot shows changes in signal amplitude with respect to time (it is an amplitude-versus-time plot). Phase is not explicitly shown on a time-domain plot.

To show the relationship between amplitude and frequency, we can use what is

called a frequency-domain plot. A frequency-domain plot is concerned with only the peak value and the frequency. Changes of amplitude during one period are not shown.

Figure below shows a signal in both the time and frequency domains.

Figure 3.7 *The time-domain andfrequency-domain plots ofa sine wave*



a. A sine wave in the time domain (**peak** value: 5 V, frequency: 6 Hz)

b. The same sine wave in the frequency domain (peak value: 5 V, frequency: 6 Hz)

Composite Signals

So far, we have focused on simple sine waves. Simple sine waves have many applications in daily life. We can send a single sine wave to carry electric energy from one place to

another. For example, the power company sends a single sine wave with a frequency of 60 Hz to distribute electric energy to houses and businesses.

If we had only one single sine wave to convey a conversation over the phone, it

would make no sense and carry no information. We would just hear a buzz.

A composite signal is made of many simple sine waves.

A single frequency sine wave is not useful in data communications;

we need to send a composite signal, a signal made of many simple sine waves.

In the early 1900s, the French mathematician Jean-Baptiste Fourier showed that

any composite signal is actually a combination of simple sine waves with different frequencies, According to Fourier analysis, any composite signal is a combination of simple sine waves with different frequencies, amplitudes, and phases.

A composite signal can be periodic or nonperiodic. A periodic composite signal

can be decomposed into a series of simple sine waves with discrete frequencies that have integer values (1, 2, 3, and so on). A nonperiodic composite signal can be decomposed into a combination of an infinite number of simple sine waves with continuous frequencies, frequencies that have real values.

If the composite signal is periodic, the decomposition gives a series of signals with discrete frequencies; if the composite signal is nonperiodic, the decomposition gives a combination of sine waves with continuous frequencies.

Example 3.8

Figure 3.9 shows a periodic composite signal with frequency *f, this* type of signal is not typical of those found in data communications. The analysis of this signal can give us a good understanding of how to decompose signals.

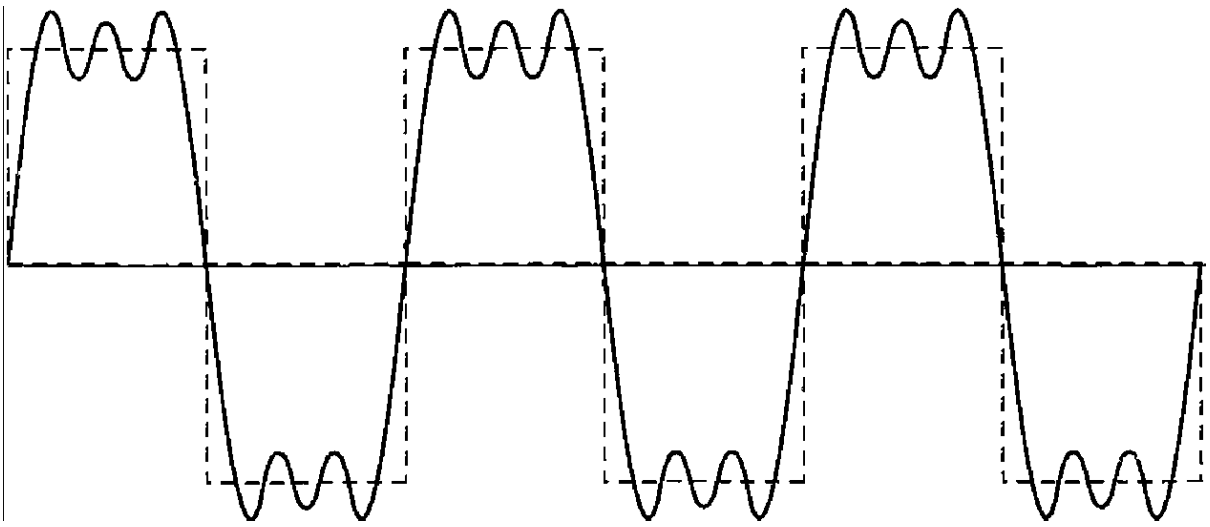
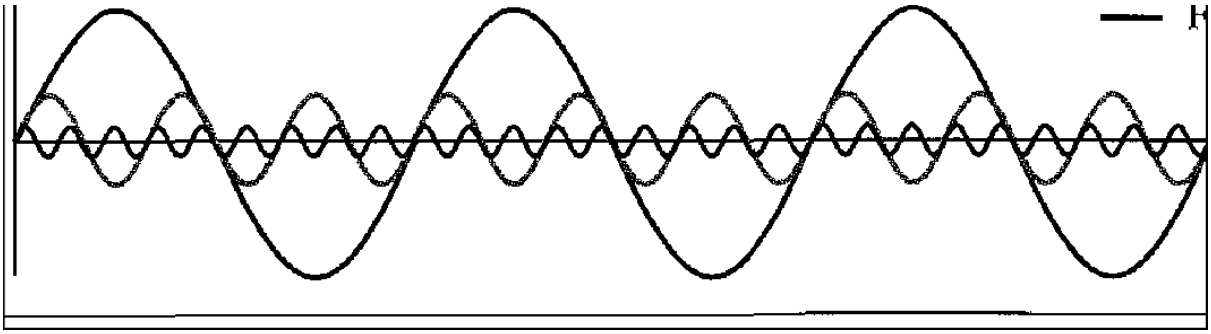
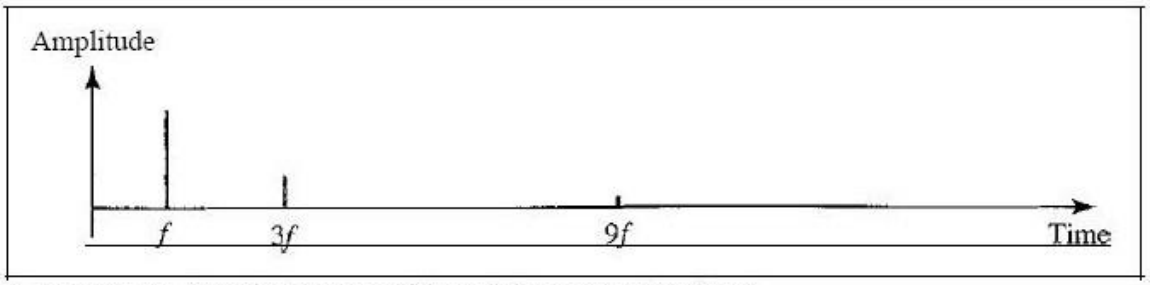


Figure 3.10 shows the result of

decomposing the above signal in both the time and frequency domains.



Time domain



b. Frequency-domain decomposition of the composite signal

figure (3.10)

The amplitude of the sine wave with frequency **f** is almost the same as the peak amplitude of the composite signal. The amplitude of the sine wave with frequency **3f** is one-third of that of the first, and the amplitude of the sine wave with frequency **9f** is one-ninth of the first. The frequency of the sine wave with frequency **f** is the same as the frequency of the composite signal; it is called

the fundamental frequency, or first harmonic. The sine wave with frequency **3f** has a frequency of 3 times the fundamental frequency; it is called the third harmonic. The third sine wave with frequency **9f** has a frequency of 9 times the fundamental frequency; it is called the ninth harmonic.

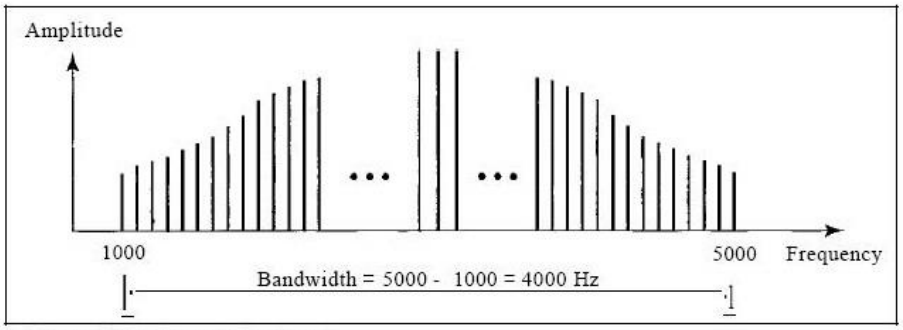
Bandwidth

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.

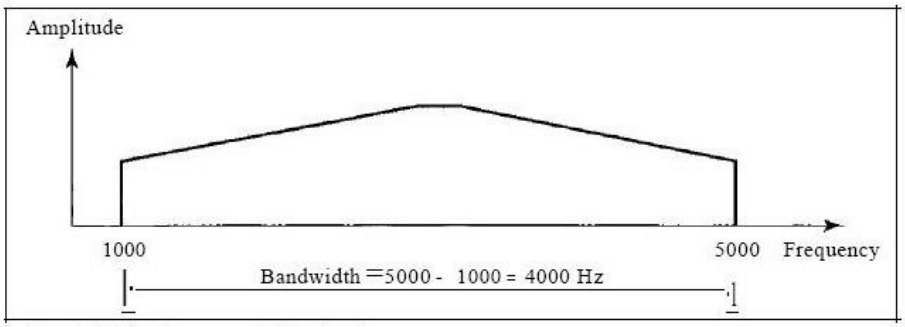
The bandwidth of a composite signal is the difference between the highest and the lowest frequencies contained in that signal.

Figure 3.12 shows the concept of bandwidth.

Figure 3.12 *The bandwidth ofperiodic and nonperiodic composite signals*



a. Bandwidth of a periodic signal



b. Bandwidth of a nonperiodic signal

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, 1001, 1002, ...). The bandwidth

of the nonperiodic signals has the same range, but the frequencies are continuous.

Example 3.10

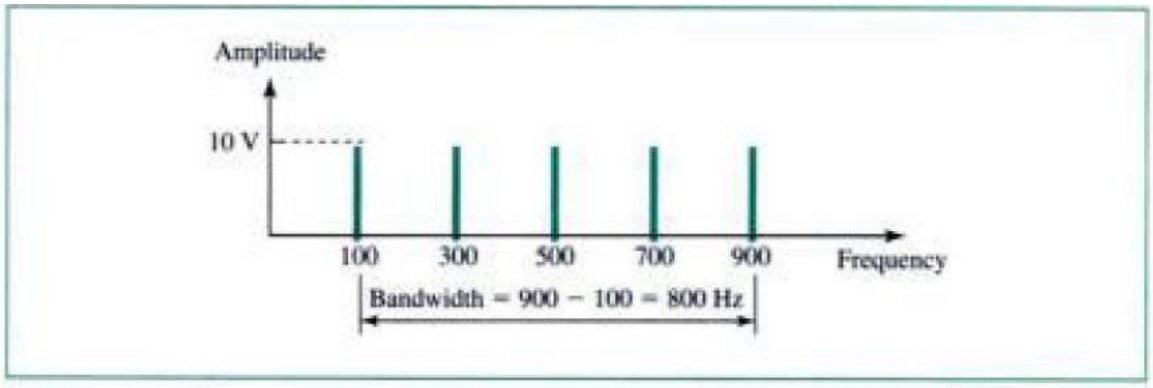
If a periodic signal is decomposed into five sine waves with frequencies of 100, 300, 500, 700, and 900 Hz, what is its bandwidth? Draw the spectrum, assuming all components have a maximum amplitude of 10 V.

Solution:

Let fh be the highest frequency, *fl* the lowest frequency, and *B* the bandwidth. Then

B =*fh* - fl = 900 - 100 =800 Hz

The spectrum has only five spikes, at 100, 300, 500, 700, and 900 Hz (see Figure 3.13).



Example 3.11

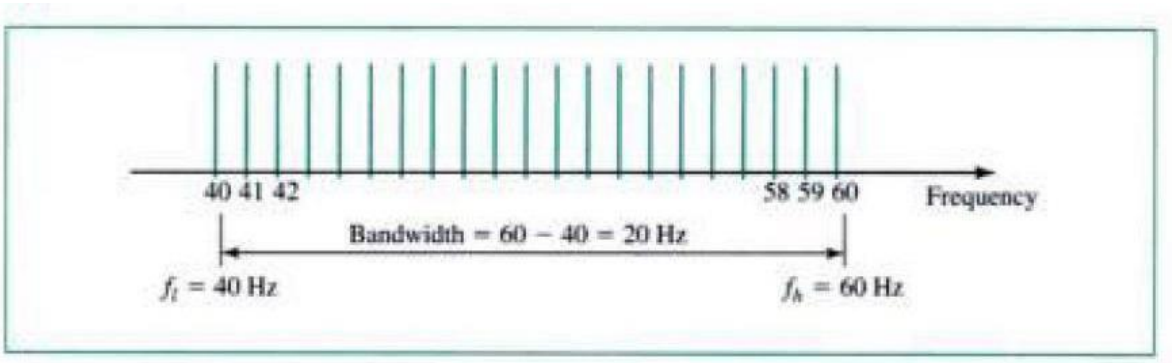
A periodic signal has a bandwidth of 20 Hz. The highest frequency is 60 Hz. What is the lowest frequency? Draw the spectrum if the signal contains all frequencies of the same amplitude.

Solution:

Let fh be the highest frequency, fz the lowest frequency, and B the bandwidth. Then

$B = fh - fz \rightarrow 20 = 60 - fz \rightarrow fz = 60 - 20 = 40$ Hz

The spectrum contains all integer frequencies. We show this by a series of spikes (see Figure 3.14).

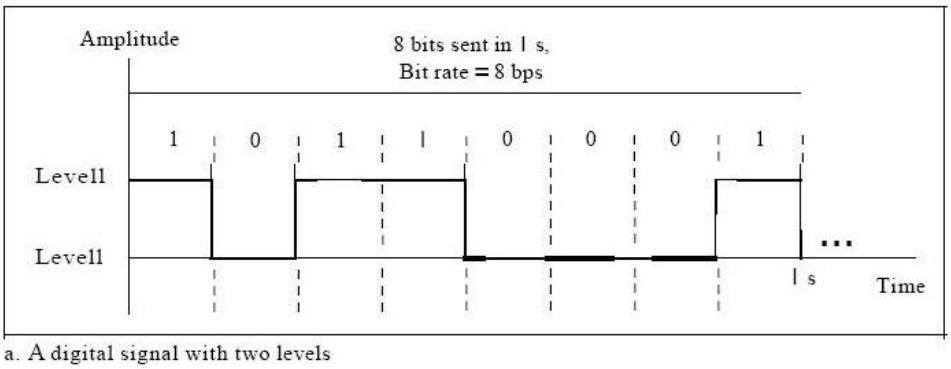


3.3 DIGITAL SIGNALS

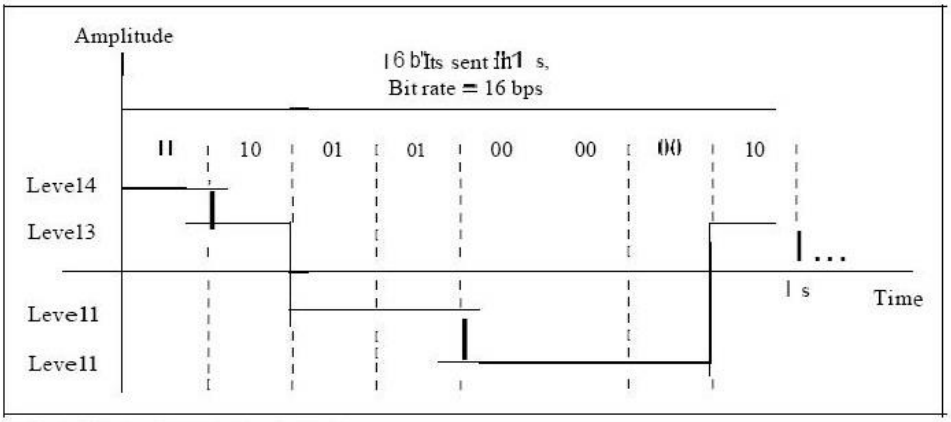
In addition to being represented by an analog signal, information can also be represented by a digital signal. For example, a I can be encoded as a positive voltage and a 0 as zero voltage. A digital signal can have more than two levels. In this case, we can send more than 1 bit for each level. Figure 3.16 shows two signals, one with two levels and the other with four.

Figure 3.16 *Two digital signals: one with two signal levels and the other with four signal levels*

Figure 3.16 *Two digital signals: one with two signal levels and the other with four signal levels*



a. A digital signal with two levels



b. A digital signal with four levels

We send 1 bit per level in part a of the figure and 2 bits per level in part b of the figure. In general, if a signal has *L* levels, each level needs ***log L*** bits.

Example 3.16

A digital signal has eight levels. How many bits are needed per level? We calculate the number of bits from the formula

Number of bits per level = log 8 = 3

Each signal level is represented by 3 bits.

Bit Rate

Most digital signals are nonperiodic, and thus period and frequency are not appropriate characteristics. Another *term-bit rate* (instead *of frequency)-is* used to describe digital signals. The bit rate is the number of bits sent in Is, expressed in bits per second (bps). Figure 3.16 shows the bit rate for two signals.

Example 3.18

Assume we need to download text documents that have of 100 pages . What is the required bit rate of the channel?

Solution:

A page is an average of 24 lines with 80 characters in each line. If we assume that one character requires 8 bits, the bit rate is

100 x 24 x 80 x 8 =1,636,000 bps =1.636 Mbps

Bit Length

The bit length is the distance one bit occupies on the transmission medium.

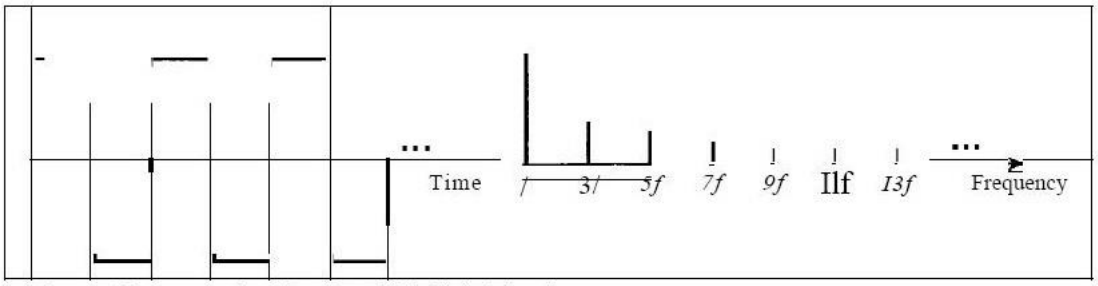
Bit length = propagation speed x bit duration

Digital Signal as a Composite Analog Signal

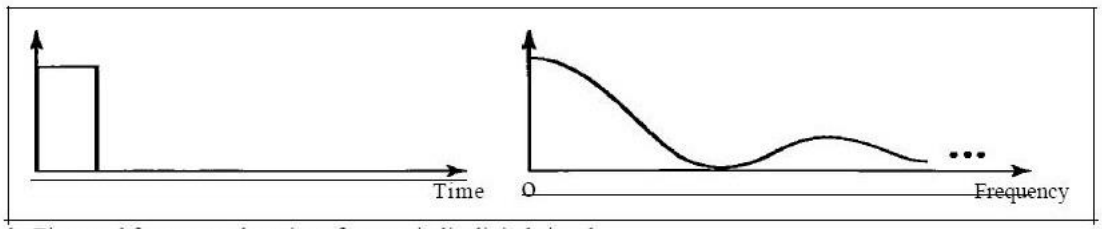
Based on Fourier analysis, a digital signal is a composite analog signal. The bandwidth is infinite, as you may have guessed. We can intuitively corne up with this concept when we consider a digital signal. A digital signal, in the time domain, comprises connected vertical and horizontal line segments. A vertical line in the time domain means a frequency of infinity (sudden change in time); a horizontal line in the time domain means a frequency of zero (no change in time). Going from a frequency of zero to a frequency of infinity (and vice versa) implies all frequencies in between are part of the domain.

Fourier analysis can be used to decompose a digital signal. If the digital signal is periodic, which is rare in data communications, the decomposed signal has a frequency domain representation with an infinite bandwidth and discrete frequencies. If the digital signal is nonperiodic, the decomposed signal still has an infinite bandwidth, but the frequencies are continuous. Figure 3.17 shows a periodic and a nonperiodic digital signal and their bandwidths.

Figure 3.17 *The time andfrequency domains ofperiodic and nonperiodic digital signals*



a. Time and frequency domains of **periodic** digital signal



b. Time and frequency domains of nonperiodic digital signal

Transmission of Digital Signals

The previous discussion asserts that a digital signal, periodic or nonperiodic, is a composite analog signal with frequencies between zero and infinity. For the remainder of the discussion, let us consider the case of a nonperiodic digital signal, similar to the ones we encounter in data communications. The fundamental question is, How can we send a digital signal from point *A* to point *B?* We can transmit a digital signal by using one of two different approaches: baseband transmission or broadband transmission (using modulation).

1- Baseband Transmission

Baseband transmission means sending a digital signal over a channel without changing the digital signal to an analog signal.

2- Broadband Transmission (Using Modulation)

Broadband transmission or modulation means changing the digital signal to an analog signal for transmission.

Medium Bandwidth and Significant Bandwidth:

A transmission medium has a limited bandwidth which mean that it can transfer only a some range of frequencies. A transmission medium with a particular bandwidth is capable of transmitting only digital signals whose significant bandwidth is less than the bandwidth of the medium. If a signal is sent on a transmission medium whose bandwidth is less than the required significant bandwidth, the signal may be so distorted that it is not recognizable at the receiver.

Medium Bandwidth and Data Rate (Channel Capacity)

The significant bandwidth of a signal increases with bit rate. This means when the bit rate is increased, we have wider significant bandwidth, and consequently we need medium with wider bandwidth to transfer that signal. The maximum bit rate a transmission medium can transfer is called channel capacity of the medium.for example , a normal telephone line with a bandwidth of 3000 Hz is capable of transferring up to 20000 bps, but other factors can decrease this rate.