(Prepared By Deepika Panda, Lecturer in Electronics, Dept of ETC, U.C.P Engineering School, Berhampur-760010)

TH-2 DATA COMMUNICATION & COMPUTER NETWORK

(Common to CSE/IT/ETC/AEI**)**

Theory : 4 Periods per week
I.A. : 20 Marks
Total Periods : 60 Periods
Term End Exam : 80 Marks
Examination : 3 Hours
TOTAL MARKS : 100 Marks

Chapter wise Distribution of periods with Total periods

Sl. No	Topics	Total periods
1	NETWORK& PROTOCOL	08
2	DATA TRANSMISSION & MEDIA	08
3	DATA ENCODING	08
4	DATA COMMUNICATION & DATA LINK CONTROL	08
5	SWITCHING & ROUTING	10
6	LAN TECHNOLOGY	10
7	TCP/IP	08
	Total	60

RATIONALE:

Now a days the growth of data communication technology has become very fast in development of various application areas. This subject will expose the learner to have an idea about the architecture computer network and different protocols to be followed to communicate. Further they will have an idea about different mode of communication.

Objective:

After completion of this course the student will be able to:

- Know the concepts of Data Communication, networking, protocols, and networking models
- Know the various transmission Medias
- Understand the concepts of switching
- Understand various Error detection and correction methods
- Know about data flow and error control
- Know about data link control
- Understand multiple access
- Learn the concepts of wired LANs and Ethernet
- Compare various connecting devices
- Know the concepts of network layer, logical addressing, IP, Forwarding and routing
- Understand brief concept on TCP/IP

Detailed Contents:

Sl.No	Topic
Unit-1	**Network& Protocol**
1.1	Data Communication

1.2	Networks
1.3	Protocol & Architecture, Standards, OSI, TCP/IP
Unit-2	**Data Transmission & Media**
2.1	Data transmission Concepts and Terminology
2.2	Analog and Digital Data transmission
2.3	Transmission impairments, Channel capacity
2.4	Transmission media, Guided Transmission, Wireless Transmission
Unit-3	**Data Encoding**
3.1	Data encoding
3.2	Digital data digital signals
3.3	Digital data analog signals
3.4	Analog data digital signals
3.5	Analog data analog signals
Unit-4	**Data Communication & Data link Control**
4.1	Asynchronous and Synchronous Transmission
4.2	Error Detection
4.3	Line configuration
4.4	Flow Control,
4.5	Error Control
4.6	Multiplexing
4.7	FDM synchronous TDM
4.8	Statistical TDM
Unit -5	**Switching & Routing**
5.1	Circuit Switching networks
5.2	Packet Switching Principles
5.3	X.25
5.4	Routing in Packet switching
5.5	Congestion
5.6	Effects of congestion, congestion control
5.7	Traffic Management
5.8	Congestion Control in Packet Switching Network
Unit-6	**LAN Technology**
6.1	Topology and Transmission Media
6.2	LAN protocol architecture
6.3	Medium Access control
6.4	Bridges, Hub, Switch
6.5	Ethernet (CSMA/CD), Fiber Channel
6.6	Wireless LAN Technology..
Unit-7	**TCP/IP**
7.1	TCP/IP Protocol Suite
7.2	Basic Protocol functions
7.3	Principles of Internetworking
7.4	Internet Protocol operations
7.5	Internet Protocol

Unit – 1. Network & Protocol

1.1 Data Communication

Data communication refers to the exchange of data between a source and a receiver via form of transmission media such as a wire cable or wireless media. Data communication is said to be local if communicating devices are in the same building or a similarly restricted geographical area.

Components:

A data communications system has five components.

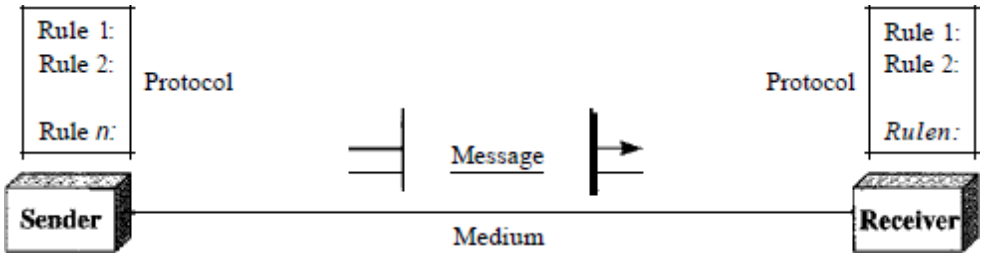


Fig 1.1(a)

1. Message: The message is the information (data) to be communicated. Popular forms of information include text, numbers, pictures, audio, and video.
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. Transmission medium: The transmission medium is the physical path by which a message travels from sender to receiver. Some examples of transmission media include twisted-pair wire, coaxial cable, fiber-optic cable, and radio waves
5. Protocol: A protocol is a set of rules that govern data communications. 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.

Data Representation:

Information today comes in different forms such as:

- Text
- Numbers
- Images,
- Audio
- Video.

Text: In data communications, text is represented as a bit pattern, a sequence of bits (O's or 1's). Different sets of bit patterns have been designed to represent text symbols. Each set is called a code, and the process of representing symbols is called coding. Today, the prevalent coding system is called Unicode, which uses 32 bits to represent a symbol or character used in any language in the world. The American Standard Code for Information Interchange (ASCII) , developed some decades ago in the United States, now constitutes the first 127 characters in Unicode and is also referred to as Basic Latin.

Numbers: Numbers are also represented by bit patterns. However, a code such as ASCII is not used to represent numbers; the number is directly converted to a binary number to simplify mathematical operations.
Images: Images are also represented by bit patterns. In its simplest form, an image is composed of a matrix of pixels (picture elements), where each pixel is a small dot. The size of the pixel depends on the *resolution.* For example, an image can be divided into 1000 pixels or 10,000 pixels A black pixel can be represented by 00, a dark gray pixel by 01, a light gray pixel by 10, and a white pixel by 11. There are several methods to represent color images. One method is called RGB, so called because each color is made of a combination of three primary colors: *red,* green, and blue. The intensity of each color is measured, and a bit pattern is assigned to it. Another method is called YCM, in which a color is made of a combination of three other primary colors: yellow, cyan, and magenta.
Audio: Audio refers to the recording or broadcasting of sound or music. Audio is by nature different from text, numbers, or images. It is continuous, not discrete. Even when we use a microphone to change voice or music to an electric signal, we create a continuous signal.
Video: Video refers to the recording or broadcasting of a picture or movie. Video can either be produced as a continuous entity (e.g., by a TV camera), or it can be a combination of images, each a discrete entity, arranged to convey the idea of motion. Again we can change video to a digital or an analog signal.

1.2 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.

Distributed Processing:

Most networks use distributed processing, in which a task is divided among multiple computers. Instead of one single large machine being responsible for all aspects of a process, separate computers (usually a personal computer or workstation) handle a subset.

Network Criteria:

A network must be able to meet a certain number of criteria. The most important of these are :

- Performance
- Reliability
- 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. Performance is often evaluated by two networking metrics: throughput and delay. We often need more throughput and less delay. However, these two criteria are often contradictory. If we try to send more data to the network, we may increase throughput but we increase the delay because of traffic congestion in the network.
Reliability: In addition to accuracy of delivery, 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 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. Detail explanation of topology will be given in Chapter 6.

Categories of Networks:

Local Area Networks: Local area networks, generally called LANs, are privately-owned networks within a single building or campus of up to a few kilometers in size. They are widely used to connect personal computers and workstations in company offices and factories to share resources (e.g., printers) and exchange information. LANs are distinguished from other kinds of networks by three characteristics:

(1) Their size,

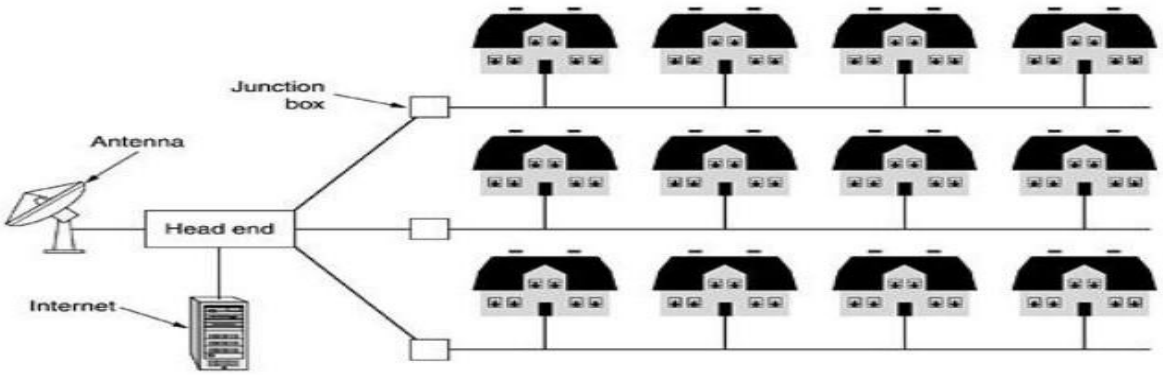
(2) Their transmission technology, and

(3) Their topology.

LANs are restricted in size, which means that the worst-case transmission time is bounded and known in advance. Knowing this bound makes it possible to use certain kinds of designs that would not otherwise be possible. It also simplifies network management. LANs may use a transmission technology consisting of a cable to which all the machines are attached, like the telephone company party lines once used in rural areas. Traditional LANs run at speeds of 10 Mbps to 100 Mbps, have low delay (microseconds or nanoseconds), and make very few errors. Newer LANs operate at up to 10 Gbps.

Metropolitan Area Network (MAN): A metropolitan area network, or MAN, covers a city. The best-known example of a MAN is the cable television network available in many cities. This system grew from earlier community antenna systems used in areas with poor over-the-air television reception. In these early systems, a large antenna was placed on top of a nearby hill and signal was then piped to the subscribers' houses. At first, these were locally-designed, ad hoc systems. Then companies began jumping into the business, getting contracts from city governments to wire up an entire city. The next step was television programming and even entire channels designed for cable only. Often these channels were highly specialized, such as all news, all sports, all cooking, all gardening, and so on. But from their inception until the late 1990s, they were intended for television reception only. To a first approximation, a MAN might look something like the system shown in Fig. In this figure both television signals and Internet are fed into the centralized head end for subsequent distribution to people's homes. Cable television is not the only MAN. Recent developments in high-speed wireless Internet access resulted in another MAN, which has been standardized as IEEE 802.16

Fig 1.2(b)



MAN is implemented by a standard called DQDB (Distributed Queue Dual Bus) or IEEE 802.16. DQDB has two unidirectional buses (or cables) to which all the computers are attached.

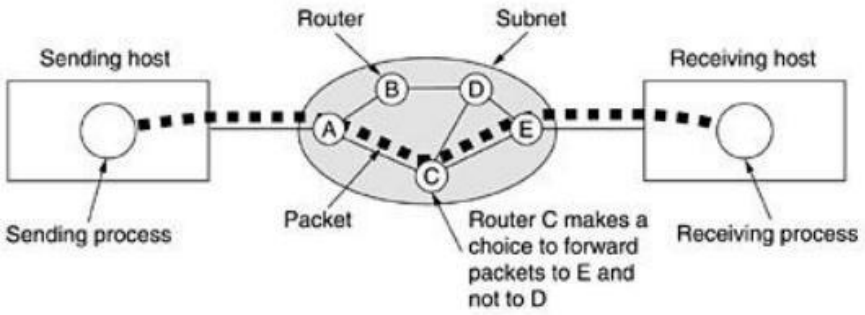
Wide Area Network (WAN): A wide area network, or WAN, spans a large geographical area, often a country or continent. It contains a collection of machines intended for running user (i.e., application) programs. These machines are called as hosts. The hosts are connected by a communication subnet, or just subnet for short. The hosts are owned by the customers (e.g., people's personal computers), 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, just as the telephone system carries words from speaker to listener.

Separation of the pure communication aspects of the network (the subnet) from the application aspects (the hosts), greatly simplifies the complete network design. In most wide area networks, the subnet consists of two distinct components: transmission lines and switching elements. Transmission lines move bits between machines. They can be made of copper wire, optical fiber, or even radio links. In most WANs, the network contains numerous transmission

lines, each one connecting a pair of routers. If two routers that do not share a transmission line wish to communicate, they must do this indirectly, via other routers. When a packet is sent from one router to another via one or more intermediate routers, the packet is received at each intermediate router in its entirety, stored there until the required output line is free, and then forwarded. A subnet organized according to this principle is called a store-and-forward or packet-switched subnet. Nearly all wide area networks (except those using satellites) have store-and-forward subnets. When the packets are small and all the same size, they are often called cells.
The principle of a packet-switched WAN is so important. Generally, 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 Fig.1.2(c)

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 packed 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.

Fig.1.2(c)



THE INTERNET

The Internet has revolutionized many aspects of our daily lives. It has affected the way we do business as well as the way we spend our leisure time. Count the ways you've used the Internet recently. Perhaps you've sent electronic mail (e-mail) to a business associate, paid a utility bill, read a newspaper from a distant city, or looked up a local movie schedule-all by using the Internet. Or maybe you researched a medical topic, booked a hotel reservation, chatted with a fellow Techie, or comparison-shopped for a car. The Internet is a communication system that has brought a wealth of information to our fingertips and organized it for our use.

1.3 Protocol & Architecture, Standards, OSI, TCP/IP

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.

Data communication standards fall into two categories: *de facto* (meaning "by fact" or "by convention") and *de jure* (meaning "by law" or "by regulation").

- **De facto:** Standards that have not been approved by an organized body but have been adopted as standards through widespread use are de facto standards. De facto standards are often established originally by manufacturers who seek to define the functionality of a new product or technology.
- **De jure:** Those standards that have been legislated by an officially recognized body are de jure standards.

LAYERED TASKS

We use the concept of layers in our daily life. As an example, let us consider two friends who communicate through postal mail. The process of sending a letter to a friend would be complex if there were no services available from the post office. Below Figure shows the steps in this task.

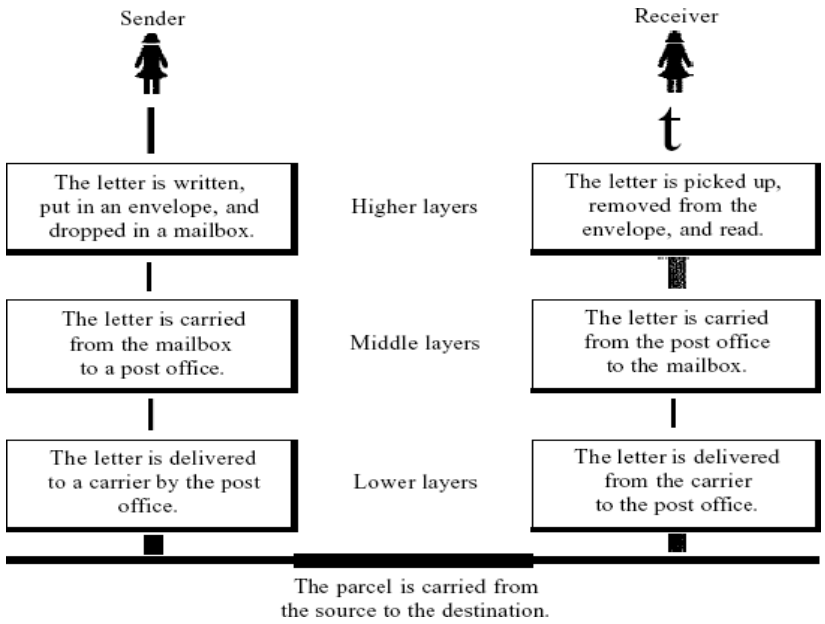


Fig 1.3(a)

In Figure we have a sender, a receiver, and a carrier that transports the letter. There is a hierarchy of tasks.

At the Sender Site: Let us first describe, in order, the activities that take place at the sender site.

- Higher layer: The sender writes the letter, inserts the letter in an envelope, writes the sender and receiver addresses, and drops the letter in a mailbox.
- Middle layer: The letter is picked up by a letter carrier and delivered to the post office.

- Lower layer: The letter is sorted at the post office; a carrier transports the letter.

On the Way: The letter is then on its way to the recipient. On the way to the recipient's local post office, the letter may actually go through a central office. In addition, it may be transported by truck, train, airplane, boat, or a combination of these.

At the Receiver Site:

- Lower layer: The carrier transports the letter to the post office.
- Middle layer: The letter is sorted and delivered to the recipient's mailbox.
- Higher layer: The receiver picks up the letter, opens the envelope, and reads it.

The OSI Reference Model:

The OSI model (minus the physical medium) is shown in Fig.1.3(b) This 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.

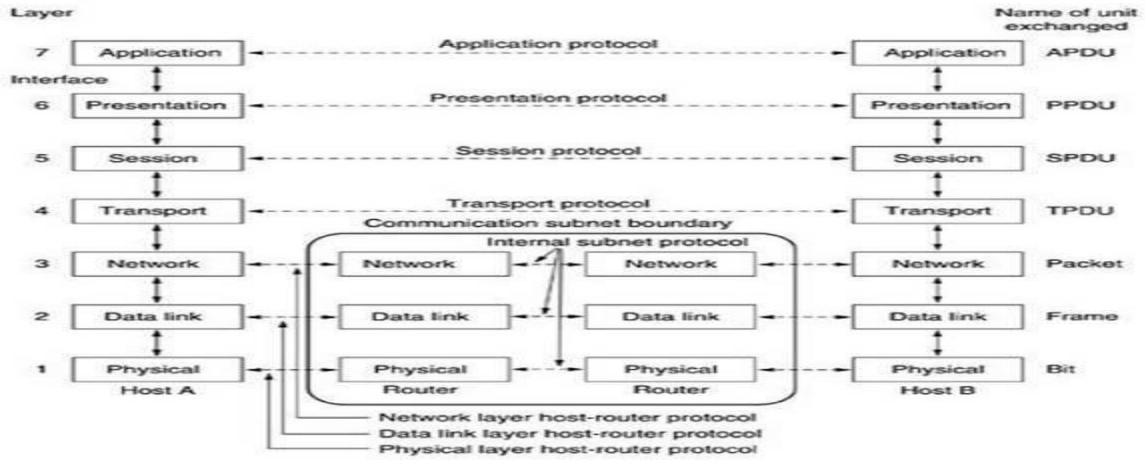


Fig 1.3(b) OSI Reference Mode

The OSI model has seven layers as shown in fig 1.3(b) , The principles that were applied to arrive at the seven layers can be briefly summarized as follows:

1. A layer should be created where a different abstraction is needed.
2. Each layer should perform a well-defined function.
3. The function of each layer should be chosen with an eye toward defining internationally standardized protocols.
4. The layer boundaries should be chosen to minimize the information flow across the interfaces.
5. The number of layers should be large enough that distinct functions need not be thrown together in the same layer out of necessity and small enough that the architecture does not become unwieldy.

The Physical Layer:

The physical layer is concerned with transmitting raw bits over a communication channel. The design issues have to do with making sure that when one side sends a 1 bit, it is received by the other side as a 1 bit, not as a 0 bit.

The Data Link Layer:

The main task of the data link layer is to transform a raw 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 input data into data frames (typically a few hundred or a few thousand bytes) and transmits the frames sequentially. If the service is reliable, the receiver confirms correct receipt of each frame by sending back an acknowledgement frame.

Another issue that arises in the data link layer (and most of the higher layers as well) is how to keep a fast transmitter from drowning a slow receiver in data. Some traffic regulation mechanism is often needed to let the transmitter know how much buffer space the receiver has at the moment. Frequently, this flow regulation and the error handling are integrated.

The Network Layer:

The network layer controls the operation of the subnet. A key design issue is determining how packets are routed from source to destination. Routes can be based on static tables that are ''wired into'' the network and rarely changed. They can also be determined at the start of each conversation, for example, a terminal session (e.g., a login to a remote machine). Finally, they can be highly dynamic, being determined anew for each packet, to reflect the current network load.

If too many packets are present in the subnet at the same time, they will get in one another's way, forming bottlenecks. The control of such congestion also belongs to the network layer. More generally, the quality of service provided (delay, transit time, jitter, etc.) is also a network layer issue.

When a packet has to travel from one network to another to get to its destination, many problems can arise. The addressing used by the second network may be different from the first one. The second one may not accept the packet at all because it is too large. The protocols may differ, and so on. It is up to the network layer to overcome all these problems to allow heterogeneous networks to be interconnected. In broadcast networks, the routing problem is simple, so the network layer is often thin or even nonexistent.

The Transport Layer:

The basic function of the transport layer is to accept data from above, split it up into smaller units if need be, pass these to the network layer, and ensure that the pieces all arrive correctly at the other end. Furthermore, all this must be done efficiently and in a way that isolates the upper layers from the inevitable changes in the hardware technology. The transport layer also determines what type of service to provide to the session layer, and, ultimately, to the users of the network. The most popular type of transport connection is an error-free point-to-point channel that delivers messages or bytes in the order in which they were sent. However, other possible kinds of transport service are the transporting of isolated messages, with no guarantee about the order of delivery, and the broadcasting of messages to multiple destinations. The type of service is determined when the connection is established.

The transport layer is a true end-to-end layer, all the way from the source to the destination. In other words, a program on the source machine carries on a conversation with a similar program on the destination machine, using the message headers and control messages. In the lower layers, the protocols are between each machine and its immediate neighbors, and not between the ultimate source and destination machines, which may be separated by many routers.

The 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), and synchronization (check pointing long transmissions to allow them to continue from where they were after a crash).

The Presentation Layer:

The presentation layer is concerned with the syntax and semantics of the information transmitted. In order to make it possible for computers with different data representations to communicate, the data structures to be exchanged can be defined in an abstract way, along with a standard encoding to be used ''on the wire.'' The presentation layer manages these abstract data structures and allows higher-level data structures (e.g., banking records), to be defined and exchanged.

The Application Layer:

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:

The TCP/IP reference model was developed prior to OSI model. The major design goals of this model were,

1. To connect multiple networks together so that they appear as a single network.
2. To survive after partial subnet hardware failures.
3. To provide a flexible architecture.

Unlike OSI reference model, TCP/IP reference model has only 4 layers. They are,

1. Network Access Layer
2. Internet Layer
3. Transport Layer
4. Application Layer

Network Access Layer:

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 to host and network to network.

Internet Layer:

This layer, called the 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 they 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 present in the Internet.

The internet layer defines an official packet format and protocol called IP (Internet Protocol). 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. Fig. shows this correspondence.

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 peer entities on the source and destination hosts to carry on a conversation, just as in the OSI transport layer. Two end-to-end transport protocols have been defined here. The first one, 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 stream into discrete messages and passes each one on to the internet layer. At the destination, the receiving TCP process reassembles the received messages into the output stream. TCP also handles flow control to make sure a fast sender cannot swamp a slow receiver with more messages than it can handle.

The second protocol in this layer, 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 is also widely used for one-shot, client-server-type request-reply queries and applications in which prompt delivery is more important than accurate delivery, such as transmitting speech or video. The relation of IP, TCP, and UDP is shown in Fig.2. Since the model was developed, IP has been implemented on many other networks.

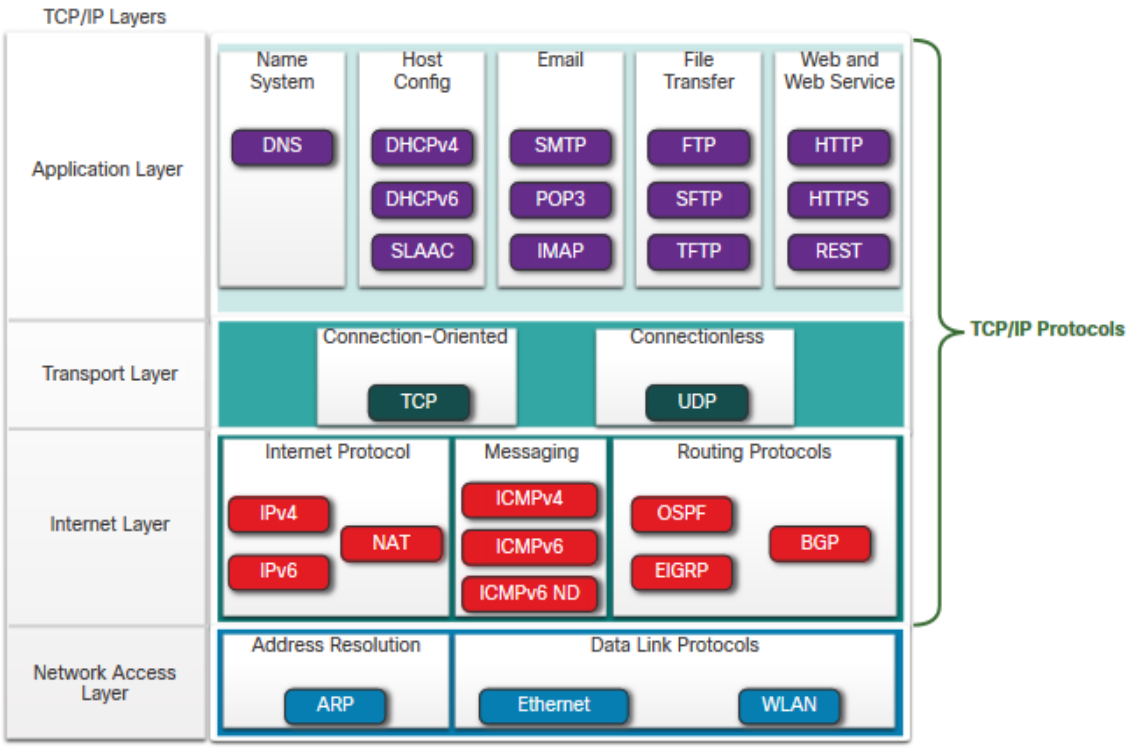
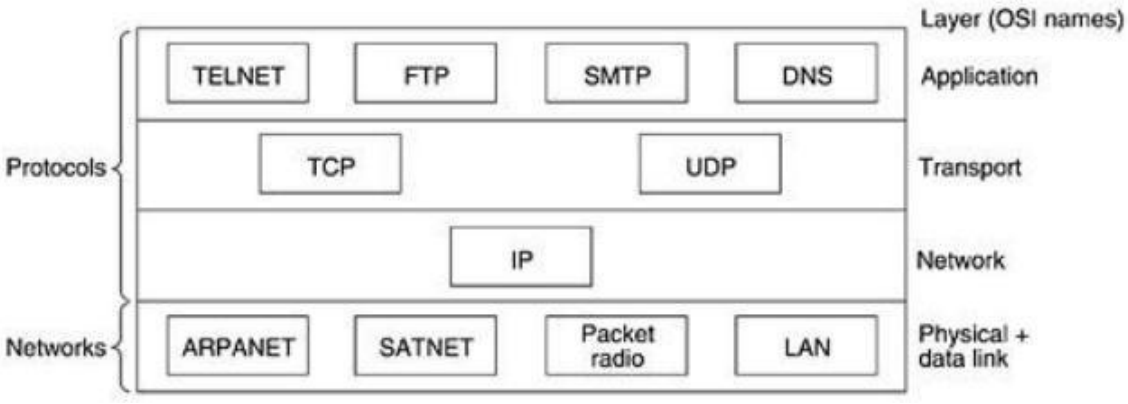


Fig. 1.3(c) The TCP/IP reference model.

Fig. 1.3(d) Important Protocols and networks in the TCP/IP model initially.



The Application Layer:

The TCP/IP model does not have session or presentation layers. On top of the transport layer is the application layer. It contains all the higher-level protocols. The early ones included virtual terminal (TELNET), file transfer (FTP), and electronic mail (SMTP), as shown in Fig.6.2. 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 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.

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 form 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 Three 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 used in a layer are the layer's own business. It can use any protocols it wants to, 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.

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 offered by the internet layer are SEND IP PACKET and RECEIVE IP PACKET.

As a consequence, the protocols in the OSI model are better hidden than in the TCP/IP model and can be replaced relatively easily as the technology changes. Being able to make such changes 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 biased toward one particular set of protocols, a fact that made it quite general. The downside of this ordering is that the designers did not have much experience with the subject and did not have a good idea of which functionality to put in which layer.

Another difference is in the area of connectionless versus connection-oriented communication. The OSI model supports both connectionless and connection-oriented communication in the network layer, but only connection-oriented communication in the transport layer, where it counts (because the transport service is visible to the users). The TCP/IP model has only one mode in the network layer (connectionless) but supports both modes in the transport layer, giving the users a choice. This choice is especially important for simple request-response protocols.

Unit-2. Data Transmission & Media

2.1 Data transmission Concepts and Terminology

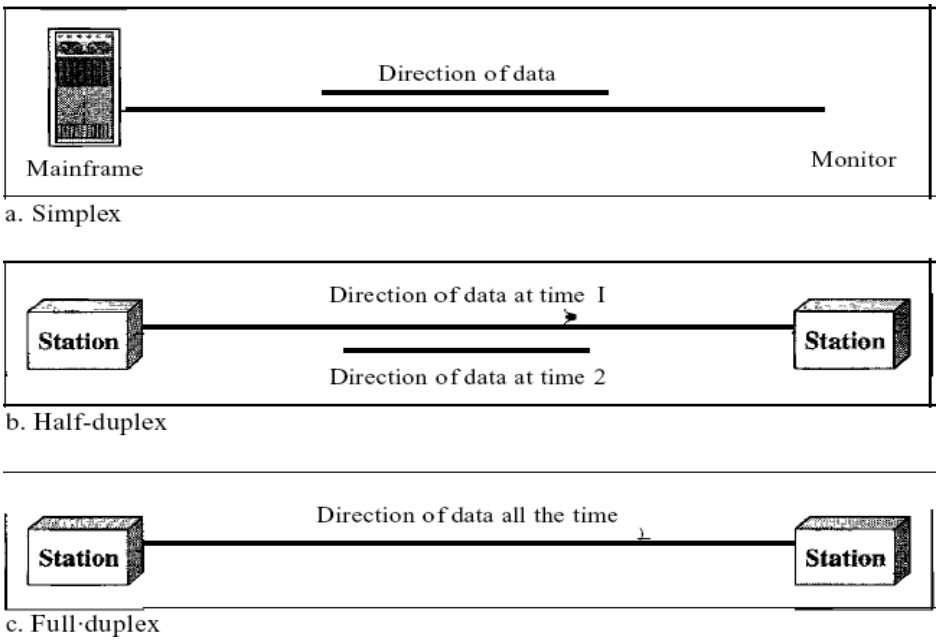
Data Flow:

Communication between two devices is divided into 3 categories such as **simplex, half-duplex and full-duplex** as shown in Figure 2.1(a)

Simplex: In simplex mode, the communication is unidirectional, as on a one-way street. Only one of the two devices on a link can transmit; the other can only receive (see Figure a of Fig 2.1(a)). Keyboards and traditional monitors are examples of simplex devices. The keyboard can only introduce input; the monitor can only accept output. The simplex mode can use the entire capacity of the channel to send data in one direction.

Half-Duplex: In half-duplex mode, each station can both transmit and receive, but not at the same time. When one device is sending, the other can only receive, and vice versa. The half-duplex mode is like a one-lane road with traffic allowed in both directions.
When cars are traveling in one direction, cars going the other way must wait. In a half-duplex transmission, the entire capacity of a channel is taken over by whichever of the two devices is transmitting at the time. Walkie-talkies and CB (citizens band) radios are both half-duplex systems. The half-duplex mode is used in cases where there is no need for communication in

Fig 2.1(a)



both directions at the same time; the entire capacity of the channel can be utilized for each direction.

Full-Duplex: In full-duplex both stations can transmit and receive simultaneously (see Figure c of Fig 2.1(a))). The full-duplex mode is like street with traffic flowing in both directions at the same time. In full-duplex mode, signals going in one direction share the capacity of the link: with signals going in the other direction. This sharing can occur in two ways: Either the link must contain two physically separate transmission ID paths, one for sending and the 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. The full-duplex mode is used when communication in both directions is required all the time. The capacity of the channel, however, must be divided between the two directions.

Physical Structures:

Type of Connection: A network is two or more devices connected through links. A link is a communications pathway that transfers data from one device to another. For visualization purposes, it is simplest to imagine any link as a line drawn between two points. For communication to occur, two devices must be connected in some way to the same link at the same time. There are two possible types of connections:

- Point-to-point
- 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) 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.

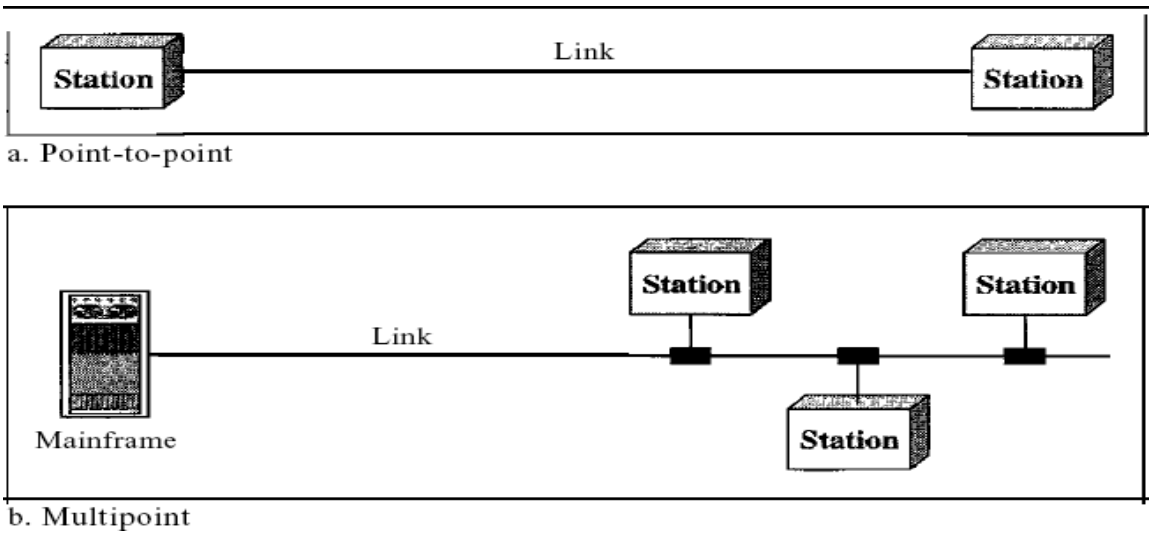


Fig 2.1(a)

Channel:

Physical medium like cables over which information is exchanged is called **channel**. Transmission channel may be **analog** or **digital**. As the name suggests, analog channels transmit data using **analog signals** while digital channels transmit data using **digital signals**.

In popular network terminology, path over which data is sent or received is called **data channel**. This data channel may be a tangible medium like copper wire cables or broadcast medium like **radio waves**.

Data Transfer Rate:

The speed of data transferred or received over transmission channel, measured per unit time, is called data transfer rate. The smallest unit of measurement is bits per second (bps). 1 bps means 1 bit (0 or 1) of data is transferred in 1 second.

Here are some commonly used data transfer rates −

- 1 Bps = 1 Byte per second = 8 bits per second
- 1 kbps = 1 kilobit per second = 1024 bits per second
- 1 Mbps = 1 Megabit per second = 1024 Kbps
- 1 Gbps = 1 Gigabit per second = 1024 Mbps

Bandwidth:

Data transfer rates that can be supported by a network is called its bandwidth. It is measured in bits per second (bps). Modern day networks provide bandwidth in Kbps, Mbps and Gbps. Some of the factors affecting a network's bandwidth include −

- Network devices used
- Protocols used
- Number of users connected
- Network overheads like collision, errors, etc.

Throughput:

Throughput is the actual speed with which data gets transferred over the network. Besides transmitting the actual data, network bandwidth is used for transmitting error messages, acknowledgement frames, etc.

Throughput is a better measurement of network speed, efficiency and capacity utilization rather than bandwidth.

Protocol:

As we have discussed earlier in Unit 1, Protocol is a set of rules and regulations used by devices to communicate over the network. Just like humans, computers also need rules to ensure successful communication. If two people start speaking at the same time or in different languages when no interpreter is present, no meaningful exchange of information can occur.

Similarly, devices connected on the network need to follow rules defining situations like when and how to transmit data, when to receive data, how to give error-free message, etc.

2.2 Analog and Digital Data transmission

Analog and Digital Signals:

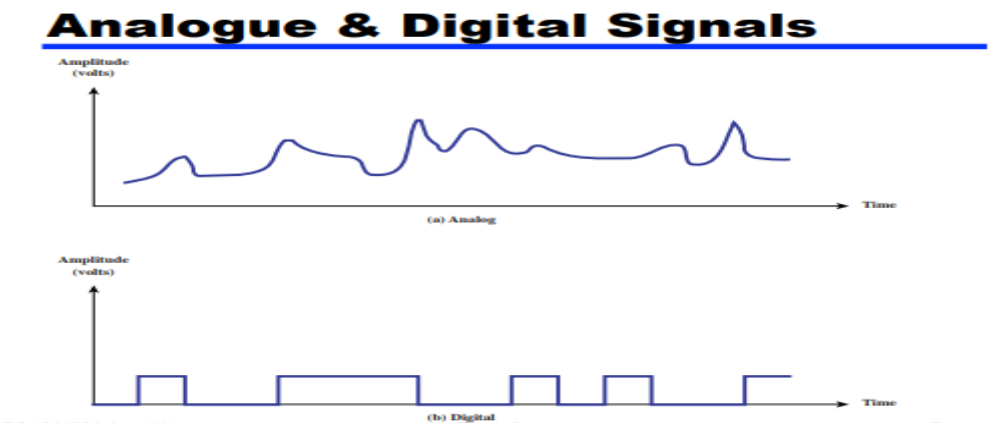


Fig 2.2(a)

Analog Signal: An analog signal is any continuous signal whose amplitude varies with time. It is continuous in amplitude and in time. Fig 2.2(a).a represents a practical analog signal which is aperiotic. Sinusoidal wave is an example of periodic Analog Signal.

Digital Signal: A digital signal refers to an electrical signal that is converted into a pattern of bits. Unlike an analog signal, which is a continuous signal that contains time-varying quantities, a digital signal has a discrete value at each sampling point.

There are a number of differences between analog and digital transmission, and it is important to understand how conversions between analog and digital occur. Let's look first at the older form of transmission, analog.

Analog Transmission

An analog wave form (or signal) is characterized by being continuously variable along amplitude and frequency. In the case of telephony, for instance, when you speak into a handset, there are changes in the air pressure around your mouth. Those changes in air pressure fall onto the handset, where they are amplified and then converted into current, or voltage fluctuations. Those fluctuations in current are an analog of the actual voice pattern—hence the use of the term *analog* to describe these

When it comes to an analog circuit—what we also refer to as a voice-grade line—we need to also define the frequency band in which it operates. The human voice, for example, can typically generate frequencies from 100Hz to 10,000Hz, for a bandwidth of 9,900Hz. But the ear does not require a vast range of frequencies to elicit meaning from ordinary speech; the vast majority of sounds we make that constitute intelligible speech fall between 250Hz and 3,400Hz. So, the phone company typically allotted a total bandwidth of 4,000Hz for voice transmission. Total frequency spectrum of twisted-pair is 1MHz. To provision a voice-grade analog circuit, bandwidth-limiting filters are put on that circuit to filter out all frequencies above 4,000Hz. That's why analog circuits can conduct only fairly low-speed data communications. The maximum data rate over an analog facility is 33.6Kbps when there are analog loops at either end.

Analog facilities have limited bandwidth, which means they cannot support high-speed data. Another characteristic of analog is that noise is accumulated as the signal traverses the network. As the signal moves across the distance, it loses power and becomes impaired by factors such as moisture in the cable, dirt on a contact, and critters chewing on the cable somewhere in the network. By the time the signal arrives at the amplifier, it is not only attenuated, it is also impaired and noisy. One of the problems with a basic amplifier is that it is a dumb device. All it knows how to do is to add power, so it takes a weak and impaired signal, adds power to it, and brings it back up to its original power level. But along with an increased signal, the amplifier passes along an increased noise level. So in an analog network, each time a signal goes through an amplifier, it accumulates noise. After you mix together coffee and cream, you can no longer separate them. The same concept applies in analog networks: After you mix the signal and the noise, you can no longer separate the two, and, as a result, you end up with very high error rates.

Digital Transmission

Digital transmission is quite different from analog transmission. For one thing, the signal is much simpler. Rather than being a continuously variable wave form, it is a series of discrete pulses, representing one bits and zero bits computer uses a coding scheme that defines what combinations of ones and zeros constitute all the characters in a character set (that is, lowercase letters, uppercase letters, punctuation marks, digits, keyboard control functions).

How the ones and zeros are physically carried through the network depends on whether the network is electrical or optical. In electrical networks, one bits are represented as high voltage, and zero bits are represented as null, or low voltage. In optical networks, one bits are represented by the presence of light, and zero bits are represented by the absence of light. The ones and zeros—the on/off conditions—are carried through the network, and the receiving device repackages the ones and zeros to determine what character is being represented. Because a digital signal is easier to reproduce than an analog signal, we can treat it with a little less care in the network. Rather than use dumb amplifiers, digital networks use *regenerative repeaters*, also referred to as *signal regenerators*. As a strong, clean, digital pulse travels over a distance, it loses power, similar to an analog signal. The digital pulse, like an analog signal, is eroded by impairments in the network. But the weakened and impaired signal enters the regenerative repeater, where the repeater examines the signal to determine what was supposed to be a one and what was supposed to be a zero. The repeater regenerates a new signal to pass on to the next point in the network, in essence eliminating noise and thus vastly improving the error rate.

Analog Versus Digital Transmission

Table 2.1 summarizes the characteristics of analog and digital networks.

Table 2.1 Characteristics of Analog and Digital Networks

Feature	**Analog Characteristics**	**Digital Characteristics**
Signal	Continuously variable, in both amplitude and frequency	Discrete signal, represented as either changes in voltage or changes in light levels
Traffic measurement	Hz (for example, a telephone channel is 4KHz)	Bits per second (for example, a T-1 line carries 1.544Mbps, and an E-1 line transports 2.048Mbps)
Bandwidth	Low bandwidth (4KHz), which means low data transmission rates (up to 33.6Kbps) because of limited channel bandwidth	High bandwidth that can support high-speed data and emerging applications that involve video and multimedia
Network capacity	Low; one conversation per telephone channel	High; multiplexers enable multiple conversations to share a communications channel and hence to achieve greater transmission efficiencies
Network manageability	Poor; a lot of labor is needed for network maintenance and control because dumb analog devices do not provide management information streams that allow the device to be remotely	Good; smart devices produce alerts, alarms, traffic statistics, and performance measurements, and technicians at a network control center (NCC) or network operations center (NOC) can remotely monitor

Feature	Analog Characteristics	Digital Characteristics
	managed	and manage the various network elements
Power requirement	High because the signal contains a wide range of frequencies and amplitudes	Low because only two discrete signals—the one and the zero—need to be transmitted
Security	Poor; when you tap into an analog circuit, you hear the voice stream in its native form, and it is difficult to detect an intrusion	Good; encryption can be used
Error rates	High; 10^{-5} bits (that is, 1 in 100,000 bits) is guaranteed to have an error	Low; with twisted-pair, 10^{-7} (that, is 1 in 10 million bits per second) will have an error, with satellite, 10^{-9} (that is, 1 in 1 billion per second) will have an error, and with fiber, 10^{-11} (that is only 1 in 10 trillion bits per second) will have an error

Conversion: Codecs and Modems

(Additional Information:

The fact is that today we don't have all-digital or all-analog networks; we have a mix of the two. Therefore, at various points in a network, it is necessary to convert between the two signal types. The devices that handle these conversions are codecs and modems .

A *codec* (which is a contraction of *coder-decoder*) converts analog signals into digital signals. There are different codecs for different purposes. For the PSTN, for example, there are codecs that minimize the number of bits per second required to carry voice digitally through the PSTN. In cellular networks, because of the constraints and available spectrum, a codec needs to compress the voice further, to get the most efficient use of the spectrum. Codecs applied to video communication also require very specific compression techniques to be able to move those high-bandwidth signals over what may be somewhat limited channels today.

A *modem* (which is a contraction of *modulator-demodulator*) is used to infuse digital data onto transmission facilities. Some modems are designed specifically to work with analog voice-grade lines. There are also modems that are designed to work specifically with digital facilities (for example, ISDN modems, ADSL modems). A modem manipulates the variables of the electromagnetic wave to differentiate between the ones and zeros.

Although it is possible to convert between analog and digital networks, in general, conversions are a weak link in a network. A conversion is a point at which network troubles can occur, an opportunity for errors and distortions to be introduced. Therefore, ideally, we want to move toward an end-to-end digital and end-to-end optical environment. This means that nowhere between the transmitter and the receiver do signal conversions need to be done.)

2.3 Transmission impairments, Channel capacity

Transmission Impairments

Attenuation: The Reduction in the amplitude of an electrical signal with little or no distortion is known as attenuation. It is Logarithmic in nature for guided media; expressed as a constant number of decibels per unit distance. For unguided media, complex function of distance and atmospheric conditions. Three considerations for transmission engineer:

1. Received signal must have sufficient strength to enable detection
2. Signal must maintain a level sufficiently higher than noise to be received without error
3. Attenuation is an increasing function of frequency

Signal strength must be strong but not too strong to overload the circuitry of transmitter or receiver, which will cause distortion.

Data Transmission: Beyond a certain distance, attenuation becomes large to require the use of repeaters or amplifiers to boost the signal. Attenuation distorts the received signal, reducing intelligibility.

- Attenuation can be equalized over a band of frequencies .
- Use amplifiers than can amplify higher frequencies more than low frequencies.

Delay distortion: It is Peculiar to guided transmission media and caused by the fact that the velocity of signal propagation through a guided medium varies with frequency. In band limited signal, velocity tends to be highest near the center frequency and falls off towards the two edges of band varying frequency components arrive at the receiver at different times, resulting in phase shifts between different frequencies. In digital data transmission, some signal components of one bit position will spill over into other bit positions, causing inter symbol interference. It may be reduced by using equalization techniques

Noise: Undesired signals that are inserted into the real signal during transmission is called Noise. Four types of noise is describe below.

1. Thermal noise :

- Also called white noise
- Occurs due to thermal agitation of electrons
- Function of temperature and present in all electronic devices
- Uniformly distributed across frequency spectrum
- Cannot be eliminated and places an upper bound on system performance
- Thermal noise in a bandwidth of 1 Hz in any device or conductor is

$$\mathbf{N_0 = kT}\ \text{W/Hz}$$

where N_0 = Noise power density in watts per 1 Hz of bandwidth

k = Boltzmann's constant = 1.3803×10^{-23} J/◦K

T = temperature in degree Kelvin

- At room temperature, T = 17◦C, or 290◦K, and the thermal noise power density is:

$$N_0 = 1.3803 \times 10^{-23} \times 290$$
$$= 4 \times 10{-}21\ \text{W/Hz}$$
$$= -204\ \text{dBW/Hz}$$

- Noise is assumed to be independent of frequency · Thermal noise in a bandwidth of B Hz can be expressed as:

$$\mathbf{N = kT\,B}\ \text{or,}$$

in decibel-watts $\mathbf{N = 10 \log k + 10 \log T + 10 \log B}$

$$= -228.6 + 10 \log T + 10 \log B\,\text{dBW}$$

- Eg: Given a receiver with an effective noise temperature of 100◦ K and a 10 Mhz bandwidth, thermal noise

level at the output is:

$$N = -228.6 + 10 \log 102 + 10 \log 107$$
$$= -228.6 + 20 + 70$$
$$= -138.6\text{dBW}$$

2. Intermodulation noise:

- Signals at different frequencies share the same transmission medium
- May result in signals that are sum or difference or multiples of original frequencies
- Occurs when there is some nonlinearity in the transmitter, receiver, or intervening transmission system · Nonlinearity may be caused by component malfunction or excessive signal strength

3. Crosstalk:

- Unwanted coupling between signal paths
- Occurs due to electric coupling between nearby twisted pairs, multiple signals on a coaxial cable, or unwanted signals picked up by microwave antennas
- Typically same order of magnitude or less than thermal noise

4. Impulse noise:

- Noncontinuous noise, consisting of irregular pulses or noise spikes of short duration and high amplitudes
- May be caused by lightning, or flaws in communications system
- Not a major problem for analog data but can be significant for digital data ·
- Eg:A spike of 0.01 s will not destroy any voice data but will destroy 560 bits being transmitted at 56 kbps

Channel capacity:

- Maximum rate at which data can be transmitted over a communication path or channel
- Depends on four factors:
  1. Data rate – in bps
  2. Bandwidth – constrained by transmitter and nature of transmission medium, expressed in cycles per second, or Hz
  3. Noise – Average noise level over channel
  4. Error rate – Percentage of time when bits are flipped – Bandwidth is proportional to cost
- For digital data, we'll like to get as high a data rate as possible within a limit of error rate for a given bandwidth.
- Limitation on data rate for a noise free channel; equals that of channel bandwidth
- If the rate of signal transmission is 2B, then a signal with frequencies no greater than B is sufficient to carry the signal rate
- Given a bandwidth B, the highest possible signal rate is 2B.
- The above is true for signals with two voltage levels ∗ With multilevel signaling, Nyquist formulation is:

$$\textbf{Channel Capcity= C = 2B log2 M}$$

- For a given bandwidth, data rate can be increased by increasing the number of different signal elements.
- Value of M is practically limited by noise and other impairments on transmission line.

Shannon capacity formula:

- Nyquist formula gives the relationship between bandwidth and data rate.
- Noise can corrupt bits of data. Higher data rate means higher error rate.
- Higher signal strength can lead to better discrimination of signal in the presence of noise.
- Signal-to-noise (SNR) ratio: Ratio of power in signal to the power in noise present at a particular point in the noise. It is typically measured at the receiver to process the signal and eliminate unwanted noise · Often measured in decibels.

$$(\text{SNR})_{\text{dB}} = 10 \log 10 \text{ signal power noise power}$$

SNR expresses the amount by which the intended signal exceeds the noise level .High SNR implies a high quality signal while low SNR indicates the need for repeaters.

- snr sets the upper bound on achievable data rate ·
- Maximum channel capacity C, in bps, is given by

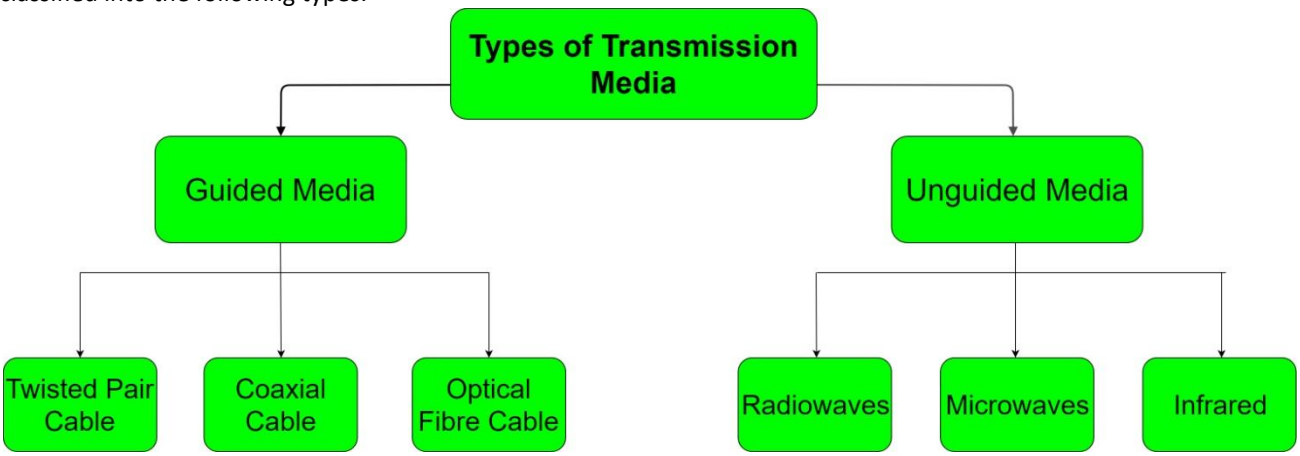
$$C = B \log_2 (1 + SNR)$$

where B is the bandwidth of the channel in Hz. Shannon formula gives the maximum possible capacity assuming only white noise; it does not take into account the impulse noise, delay distortion, and attenuation.

2.4 Transmission media, Guided Transmission, Wireless Transmission

Types of Transmission Media:

In data communication terminology, a transmission medium is a physical path between the transmitter and the receiver i.e it is the channel through which data is sent from one place to another. Transmission Media is broadly classified into the following types:



Guided Media is also called as Wired Transmission media and Unguided Media is known as Wireless Transmission Media.

1. Guided Media: It is also referred to as Wired or Bounded transmission media. Signals being transmitted are directed and confined in a narrow pathway by using physical links.

Features:

- High Speed
- Secure
- Used for comparatively shorter distances

There are 3 major types of Guided Media: **Twisted Pair, Coaxial Cable, Optical Fiber Cable**

(i) Twisted Pair Cable: It consists of 2 separately insulated conductor wires wound about each other. Generally, several such pairs are bundled together in a protective sheath. They are the most widely used Transmission Media. Twisted Pair is of two types:

1. **Unshielded Twisted Pair (UTP):** This type of cable has the ability to block interference and does not depend on a physical shield for this purpose. It is used for telephonic applications.
   Advantages:
   - Least expensive
   - Easy to install
   - High speed capacity

Disadvantages:

- Susceptible to external interference
- Lower capacity and performance in comparison to STP
- Short distance transmission due to attenuation

2. **Shielded Twisted Pair (STP):** This type of cable consists of a special jacket to block external interference. It is used in fast-data-rate Ethernet and in voice and data channels of telephone lines.

Advantages:

- Better performance at a higher data rate in comparison to UTP
- Eliminates crosstalk
- Comparitively faster

Disadvantages:

- Comparitively difficult to install and manufacture
- More expensive
- Bulky

(ii) Coaxial Cable : It has an outer plastic covering containing 2 parallel conductors each having a separate insulated protection cover. Coaxial cable transmits information in two modes: Baseband mode(dedicated cable bandwidth) and Broadband mode(cable bandwidth is split into separate ranges). Cable TVs and analog television networks widely use Coaxial cables.

Advantages:

- High Bandwidth
- Better noise Immunity
- Easy to install and expand
- Inexpensive

Disadvantages:

- Single cable failure can disrupt the entire network

(iii) Optical Fiber Cable : It uses the concept of reflection of light through a core made up of glass or plastic. The core is surrounded by a less dense glass or plastic covering called the cladding. It is used for transmission of large volumes of data.

Advantages:

- Increased capacity and bandwidth
- Light weight
- Less signal attenuation
- Immunity to electromagnetic interference
- Resistance to corrosive materials

Disadvantages:

- Difficult to install and maintain
- High cost
- Fragile
- unidirectional, ie, will need another fiber, if we need bidirectional communication

2. Unguided Media: It is also referred to as Wireless or Unbounded transmission media.No physical medium is required for the transmission of electromagnetic signals.

Features:

- Signal is broadcasted through air
- Less Secure
- Used for larger distances

There are **3 major types** of Unguided Media:

(i) Radio waves : These are easy to generate and can penetrate through buildings. The sending and receiving antennas need not be aligned. Frequency Range:3KHz – 1GHz. AM and FM radios and cordless phones use Radiowaves for transmission.Further Categorized as (i) Terrestrial and (ii) Satellite.

(ii) Microwaves : It is a line of sight transmission i.e. the sending and receiving antennas need to be properly aligned with each other. The distance covered by the signal is directly proportional to the height of the antenna. Frequency Range:1GHz – 300GHz. These are majorly used for mobile phone communication and television distribution.
(iii) Infrared : Infrared waves are used for very short distance communication. They cannot penetrate through obstacles. This prevents interference between systems. Frequency Range:300GHz – 400THz. It is used in TV remotes, wireless mouse, keyboard, printer, etc.

Unit-3. Data Encoding

3.1 Data Encoding

Encoding is the process of converting the data or a given sequence of characters, symbols, alphabets etc., into a specified format, for the secured transmission of data. **Decoding** is the reverse process of encoding which is to extract the information from the converted format.

Data Encoding: Encoding is the process of using various patterns of voltage or current levels to represent 1s and 0s of the digital signals on the transmission link. The common types of line encoding are Unipolar, Polar, Bipolar, and Manchester.
Encoding Techniques: The data encoding technique is divided into the following types, depending upon the type of data conversion.

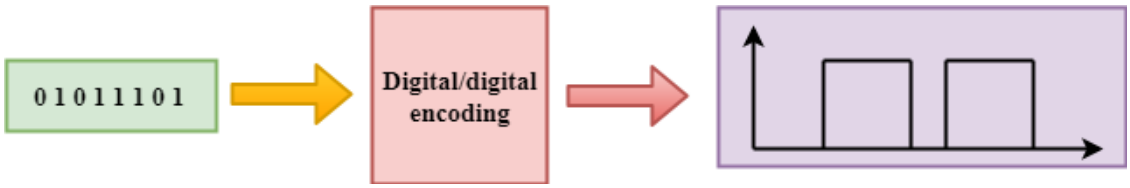
- **Analog data to Analog signals** : The modulation techniques such as Amplitude Modulation, Frequency Modulation and Phase Modulation of analog signals, fall under this category.
- **Analog data to Digital signals**: This process can be termed as digitization, which is done by Pulse Code Modulation PCMPCM. Hence, it is nothing but digital modulation. As we have already discussed, sampling and quantization are the important factors in this. Delta Modulation gives a better output than PCM.
- **Digital data to Analog signals**: The modulation techniques such as Amplitude Shift Keying ASKASK, Frequency Shift Keying FSKFSK, Phase Shift Keying PSKPSK, etc., fall under this category. These will be discussed in subsequent chapters.
- **Digital data to Digital signals**: Digital-to-digital encoding is the representation of digital information by a digital signal. When binary 1s and 0s generated by the computer are translated into a sequence of voltage pulses that can be propagated over a wire, this process is known as digital-to-digital encoding. Unipolar,Polar and Bipolar Schemes are used for data encoding.

3.2 Digital data digital signals

Digital Transmission

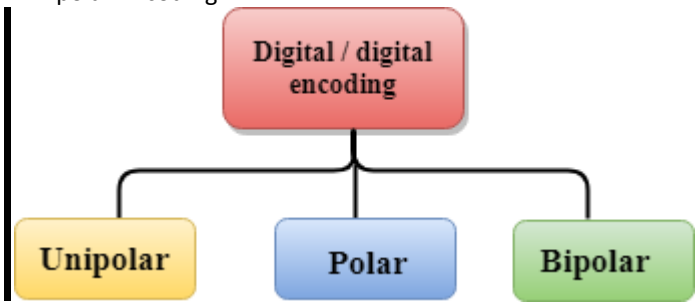
Data can be represented either in analog or digital form. The computers used the digital form to store the information. Therefore, the data needs to be converted in digital form so that it can be used by a computer.

DIGITAL-TO-DIGITAL CONVERSION: Digital-to-digital encoding is the representation of digital information by a digital signal. When binary 1s and 0s generated by the computer are translated into a sequence of voltage pulses that can be propagated over a wire, this process is known as digital-to-digital encoding.



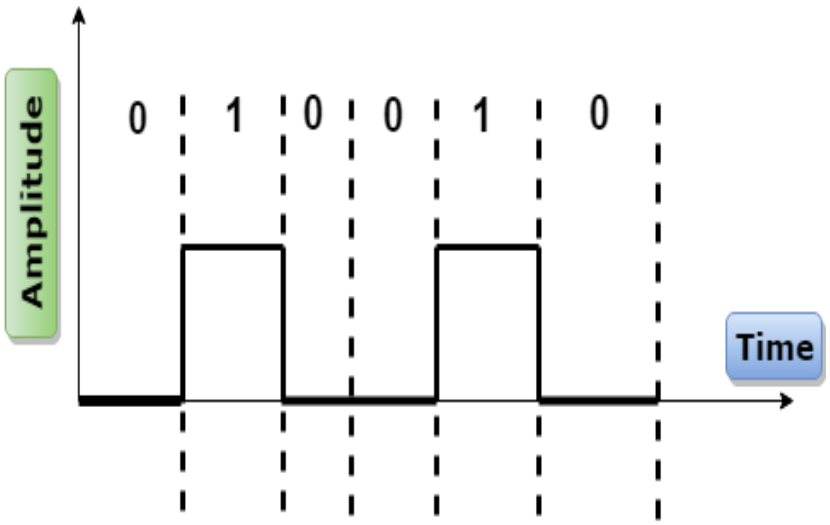
Digital-to-digital encoding is divided into three categories:

- Unipolar Encoding
- Polar Encoding
- Bipolar Encoding



Unipolar:

- Digital transmission system sends the voltage pulses over the medium link such as wire or cable.
- In most types of encoding, one voltage level represents 0, and another voltage level represents 1.
- The polarity of each pulse determines whether it is positive or negative.
- This type of encoding is known as Unipolar encoding as it uses only one polarity.
- In Unipolar encoding, the polarity is assigned to the 1 binary state.
- In this, 1s are represented as a positive value and 0s are represented as a zero value.
- In Unipolar Encoding, '1' is considered as a high voltage and '0' is considered as a zero voltage.
- Unipolar encoding is simpler and inexpensive to implement.

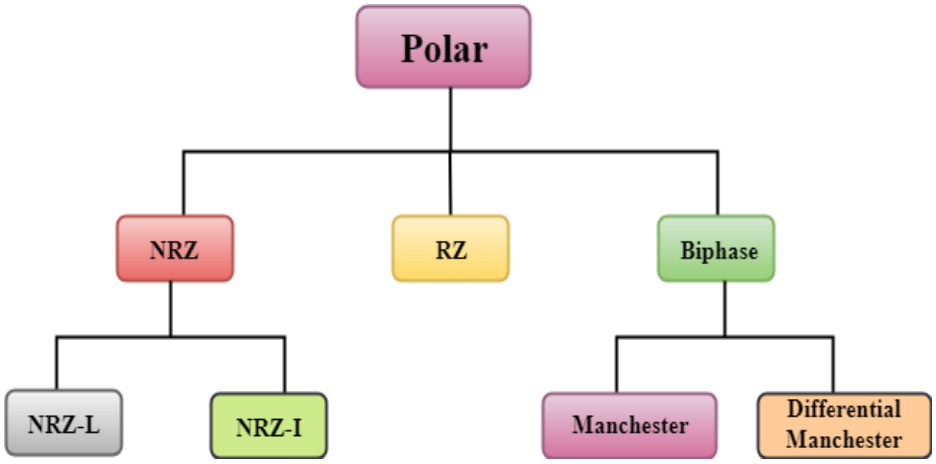


Unipolar encoding has two problems that make this scheme less desirable:

- DC Component
- Synchronization

Polar

- Polar encoding is an encoding scheme that uses two voltage levels: one is positive, and another is negative.
- By using two voltage levels, an average voltage level is reduced, and the DC component problem of unipolar encoding scheme is alleviated.



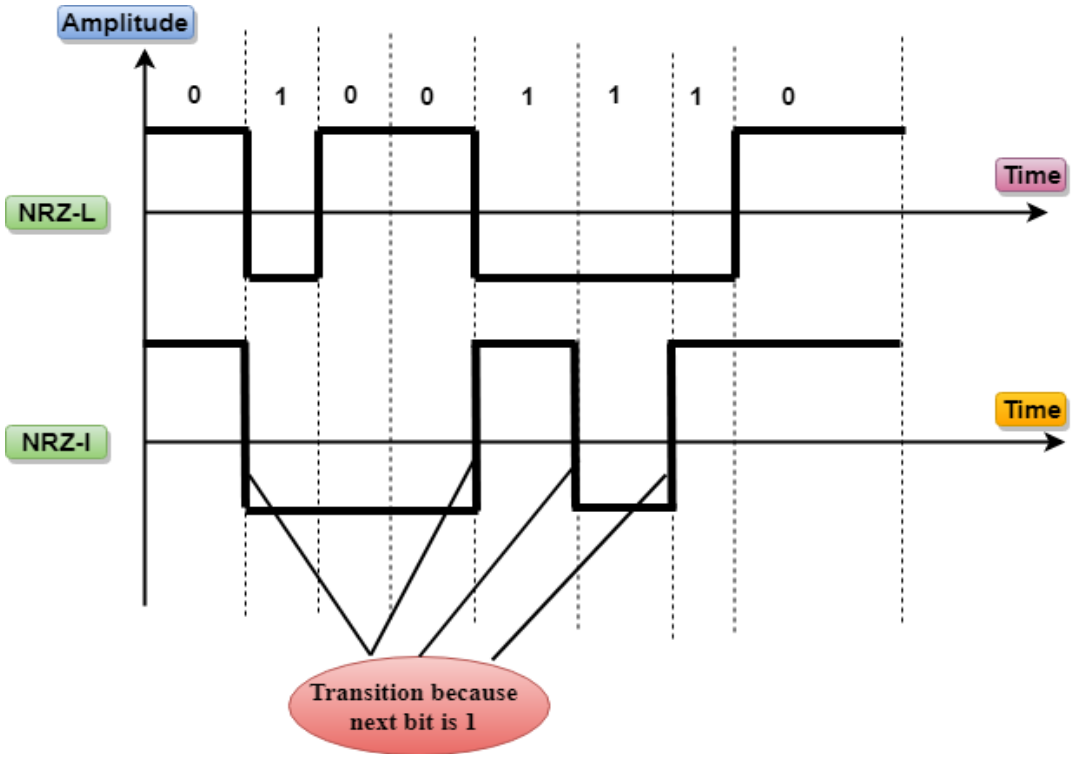
NRZ

- NRZ stands for Non-return zero.
- In NRZ encoding, the level of the signal can be represented either positive or negative.

The two most common methods used in NRZ are:

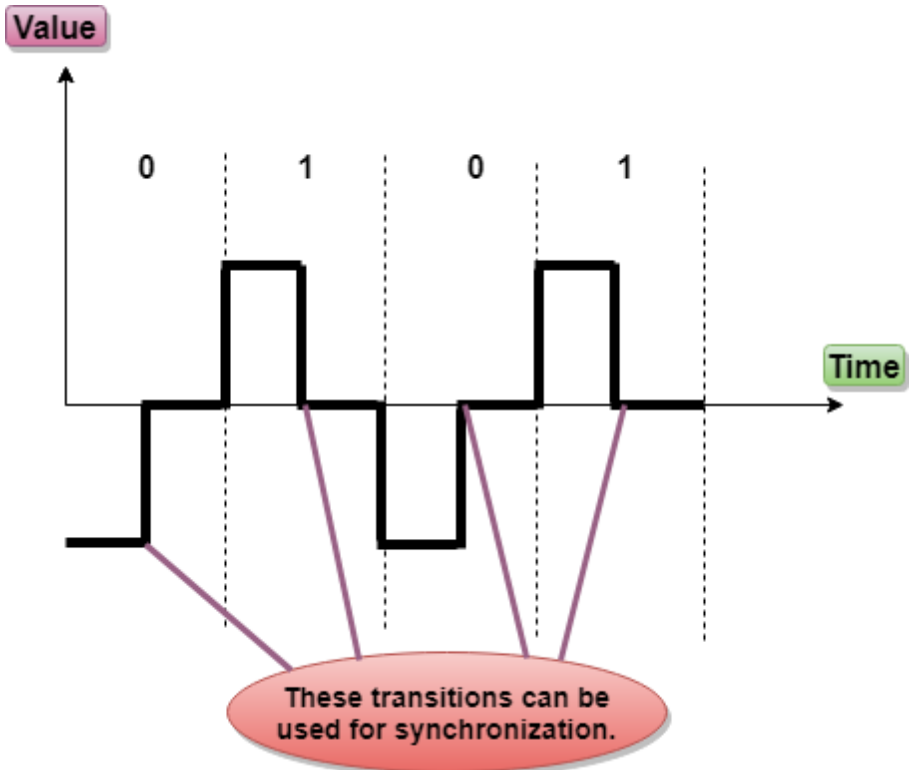
NRZ-L: In NRZ-L encoding, the level of the signal depends on the type of the bit that it represents. If a bit is 0 or 1, then their voltages will be positive and negative respectively. Therefore, we can say that the level of the signal is dependent on the state of the bit.

NRZ-I: NRZ-I is an inversion of the voltage level that represents 1 bit. In the NRZ-I encoding scheme, a transition occurs between the positive and negative voltage that represents 1 bit. In this scheme, 0 bit represents no change and 1 bit represents a change in voltage level.



RZ

- RZ stands for Return to zero.
- There must be a signal change for each bit to achieve synchronization. However, to change with every bit, we need to have three values: positive, negative and zero.
- RZ is an encoding scheme that provides three values, positive voltage represents 1, the negative voltage represents 0, and zero voltage represents none.
- In the RZ scheme, halfway through each interval, the signal returns to zero.
- In RZ scheme, 1 bit is represented by positive-to-zero and 0 bit is represented by negative-to-zero.



Disadvantage of RZ:

It performs two signal changes to encode one bit that acquires more bandwidth.

Biphase

- Biphase is an encoding scheme in which signal changes at the middle of the bit interval but does not return to zero.

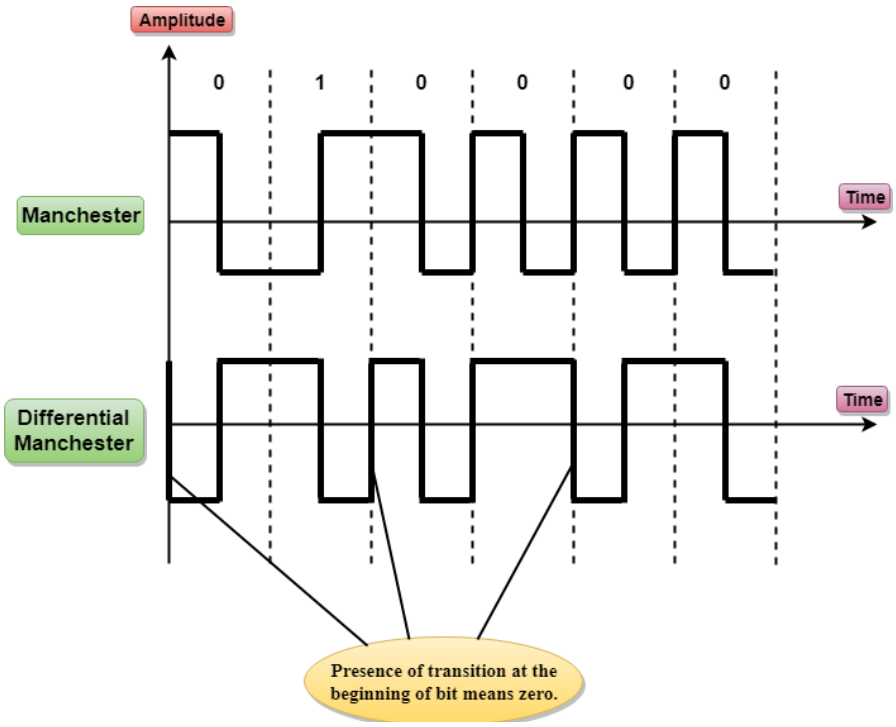
Biphase encoding is implemented in two different ways:

Manchester

- It changes the signal at the middle of the bit interval but does not return to zero for synchronization.
- In Manchester encoding, a negative-to-positive transition represents binary 1, and positive-to-negative transition represents 0.
- Manchester has the same level of synchronization as RZ scheme except that it has two levels of amplitude.

Differential Manchester

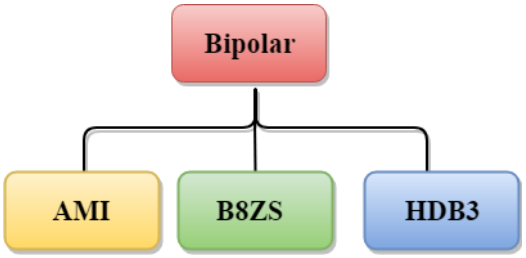
- It changes the signal at the middle of the bit interval for synchronization, but the presence or absence of the transition at the beginning of the interval determines the bit. A transition means binary 0 and no transition means binary 1.
- In Manchester Encoding scheme, two signal changes represent 0 and one signal change represent 1.



Bipolar

- Bipolar encoding scheme represents three voltage levels: positive, negative, and zero.
- In Bipolar encoding scheme, zero level represents binary 0, and binary 1 is represented by alternating positive and negative voltages.
- If the first 1 bit is represented by positive amplitude, then the second 1 bit is represented by negative voltage, third 1 bit is represented by the positive amplitude and so on. This alternation can also occur even when the 1bits are not consecutive.

Bipolar can be classified as:



AMI

- AMI stands for **alternate mark inversion** where mark work comes from telegraphy which means 1. So, it can be redefined as **alternate 1 inversion**.
- In Bipolar AMI encoding scheme, 0 bit is represented by zero level and 1 bit is represented by alternating positive and negative voltages.

Advantage:

- DC component is zero.
- Sequence of 1s bits are synchronized.

Disadvantage:

- This encoding scheme does not ensure the synchronization of a long string of 0s bits.

B8ZS

- B8ZS stands for **Bipolar 8-Zero Substitution**.
- This technique is adopted in North America to provide synchronization of a long sequence of 0s bits.

- In most of the cases, the functionality of B8ZS is similar to the bipolar AMI, but the only difference is that it provides the synchronization when a long sequence of 0s bits occur.
- B8ZS ensures synchronization of a long string of 0s by providing force artificial signal changes called violations, within 0 string pattern.
- When eight 0 occurs, then B8ZS implements some changes in 0s string pattern based on the polarity of the previous 1 bit.
- If the polarity of the previous 1 bit is positive, the eight 0s will be encoded as zero, zero, zero, positive, negative, zero, negative, positive.

HDB3

- HDB3 stands for **High-Density Bipolar 3**.
- HDB3 technique was first adopted in Europe and Japan.
- HDB3 technique is designed to provide the synchronization of a long sequence of 0s bits.
- In the HDB3 technique, the pattern of violation is based on the polarity of the previous bit.
- When four 0s occur, HDB3 looks at the number of 1s bits occurred since the last substitution.
- If the number of 1s bits is odd, then the violation is made on the fourth consecutive of 0. If the polarity of the previous bit is positive, then the violation is positive. If the polarity of the previous bit is negative, then the violation is negative.

3.3 Digital data Analog Signals

To send the digital data over an analog media, it needs to be converted into analog signal. There can be two cases according to data formatting.

Bandpass: The filters are used to filter and pass frequencies of interest. A bandpass is a band of frequencies which can pass the filter.

Low-pass: Low-pass is a filter that passes low frequencies signals.

When digital data is converted into a bandpass analog signal, it is called digital-to-analog conversion. When low-pass analog signal is converted into bandpass analog signal, it is called analog-to-analog conversion.

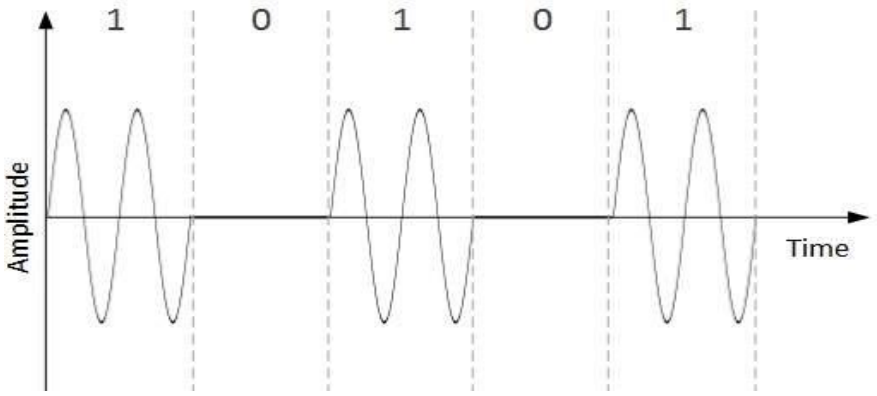
Digital-to-Analog Conversion

When data from one computer is sent to another via some analog carrier, it is first converted into analog signals. Analog signals are modified to reflect digital data.

An analog signal is characterized by its amplitude, frequency, and phase. There are three kinds of digital-to-analog conversions:

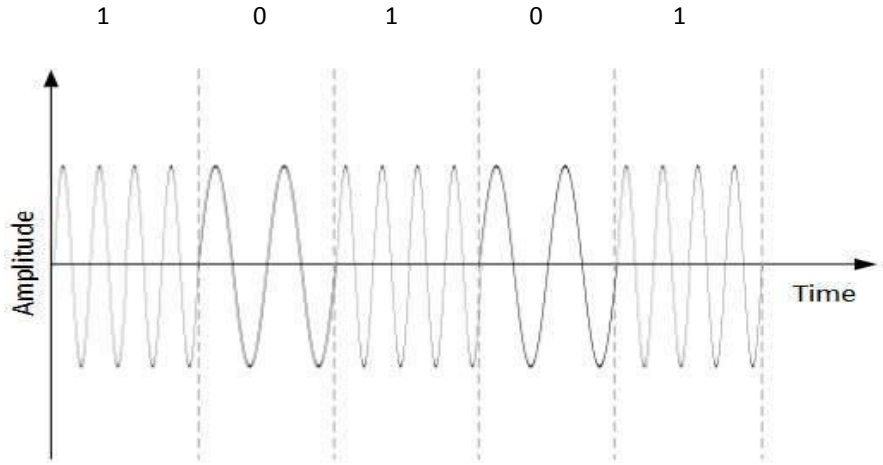
- **Amplitude Shift Keying(ASK)**

  In this conversion technique, the amplitude of analog carrier signal is modified to reflect binary data. When binary data represents digit 1, the amplitude is held; otherwise it is set to 0. Both frequency and phase remain same as in the original carrier signal. Below is figure for ASK.



- **Frequency Shift Keying**

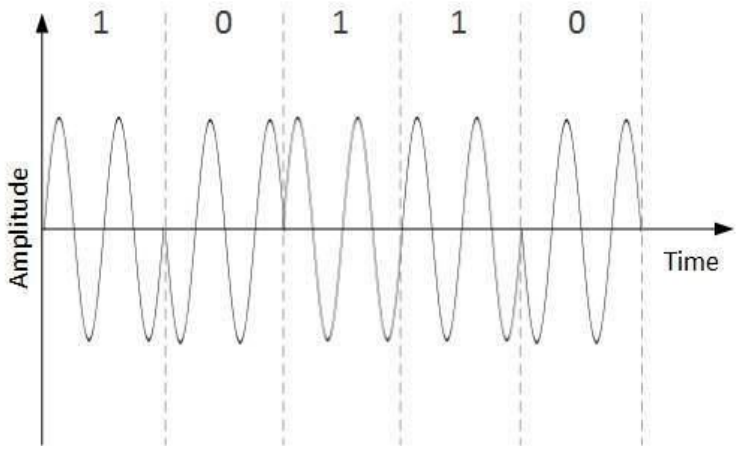
  In this conversion technique, the frequency of the analog carrier signal is modified to reflect binary data.



  This technique uses two frequencies, f1 and f2. One of them, for example f1, is chosen to represent binary digit 1 and the other one is used to represent binary digit 0. Both amplitude and phase of the carrier wave are kept intact.

- **Phase Shift Keying**

  In this conversion scheme, the phase of the original carrier signal is altered to reflect the binary data.



When a new binary symbol is encountered, the phase of the signal is altered. Amplitude and frequency of the original carrier signal is kept intact.

- **Quadrature Phase Shift Keying**

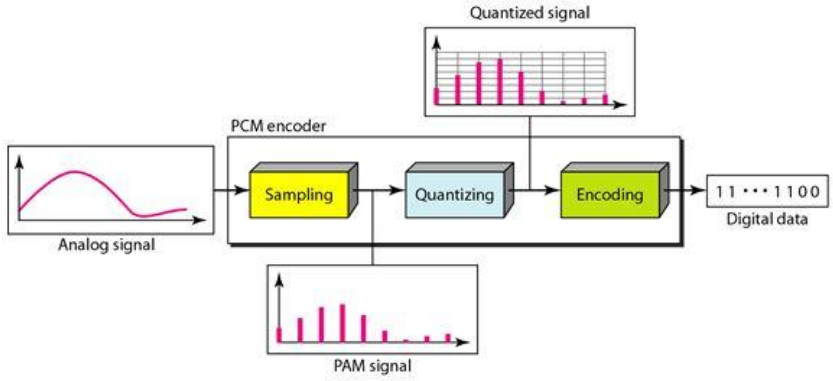
  QPSK alters the phase to reflect two binary digits at once. This is done in two different phases. The main stream of binary data is divided equally into two sub-streams. The serial data is converted in to parallel in both sub-streams and then each stream is converted to digital signal using NRZ technique. Later, both the digital signals are merged together.

3.4 Analog data Digital Signals

If we have an analog signal such as one created by a microphone or camera. To change an analog signal to digital data we use two techniques, pulse code modulation and delta modulation. After the digital data are created (digitization) then we convert the digital data to a digital signal.

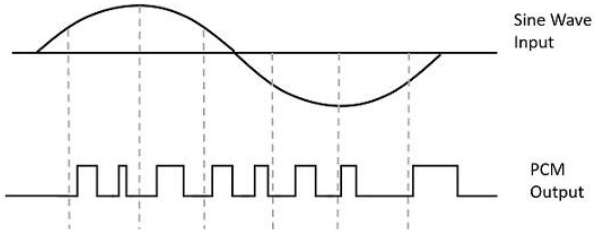
Pulse Code Modulation (PCM): Pulse Code Modulation (PCM) is the most common technique used to change an analog signal to digital data (digitization). A PCM encoder has three processes as shown in the following Figure.

The analog signal is sampled.
The sampled signal is quantized.
The quantized values are encoded as streams of bits.



Sampling: The first step in PCM is sampling. The analog signal is sampled every Ts s, where Ts is the sample interval or period. The inverse of the sampling interval is called the sampling rate or sampling frequency and denoted by ƒs, Where $\mathbf{f_s = 1/ T_s}$

A signal is pulse code modulated to convert its analog information into a binary sequence, i.e., **1s** and **0s**. The output of a PCM will resemble a binary sequence. The following figure shows an example of PCM output with respect to instantaneous values of a given sine wave.



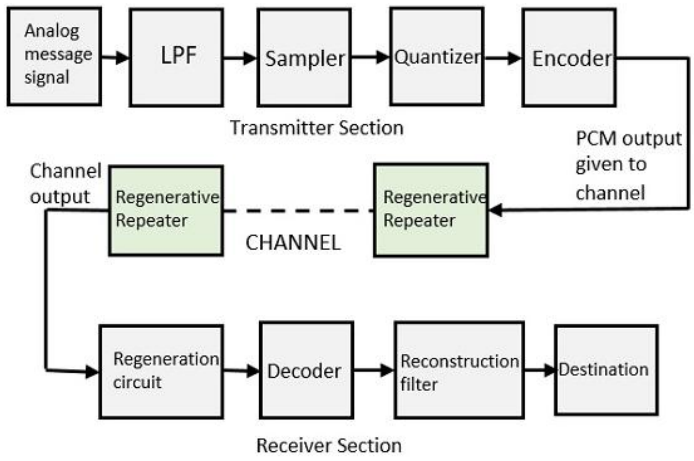
Instead of a pulse train, PCM produces a series of numbers or digits, and hence this process is called as **digital**. Each one of these digits, though in binary code, represent the approximate amplitude of the signal sample at that instant.

In Pulse Code Modulation, the message signal is represented by a sequence of coded pulses. This message signal is achieved by representing the signal in discrete form in both time and amplitude.

Basic Elements of PCM

The transmitter section of a Pulse Code Modulator circuit consists of **Sampling, Quantizing** and **Encoding**, which are performed in the analog-to-digital converter section. The low pass filter prior to sampling prevents aliasing of the message signal.

The basic operations in the receiver section are **regeneration of impaired signals, decoding,** and **reconstruction** of the quantized pulse train. Following is the block diagram of PCM which represents the basic elements of both the transmitter and the receiver sections.



Low Pass Filter: This filter eliminates the high frequency components present in the input analog signal which is greater than the highest frequency of the message signal, to avoid aliasing of the message signal.

Sampler: This is the technique which helps to collect the sample data at instantaneous values of message signal, so as to reconstruct the original signal. The sampling rate must be greater than twice the highest frequency component W of the message signal, in accordance with the sampling theorem.

Quantizer: Quantizing is a process of reducing the excessive bits and confining the data. The sampled output when given to Quantizer, reduces the redundant bits and compresses the value.

Encoder: The digitization of analog signal is done by the encoder. It designates each quantized level by a binary code. The sampling done here is the sample-and-hold process. These three sections LPF,Sampler,andQuantizerLPF,Sampler,andQuantizer will act as an analog to digital converter. Encoding minimizes the bandwidth used.

Regenerative Repeater : This section increases the signal strength. The output of the channel also has one regenerative repeater circuit, to compensate the signal loss and reconstruct the signal, and also to increase its strength.

Decoder: The decoder circuit decodes the pulse coded waveform to reproduce the original signal. This circuit acts as the demodulator.

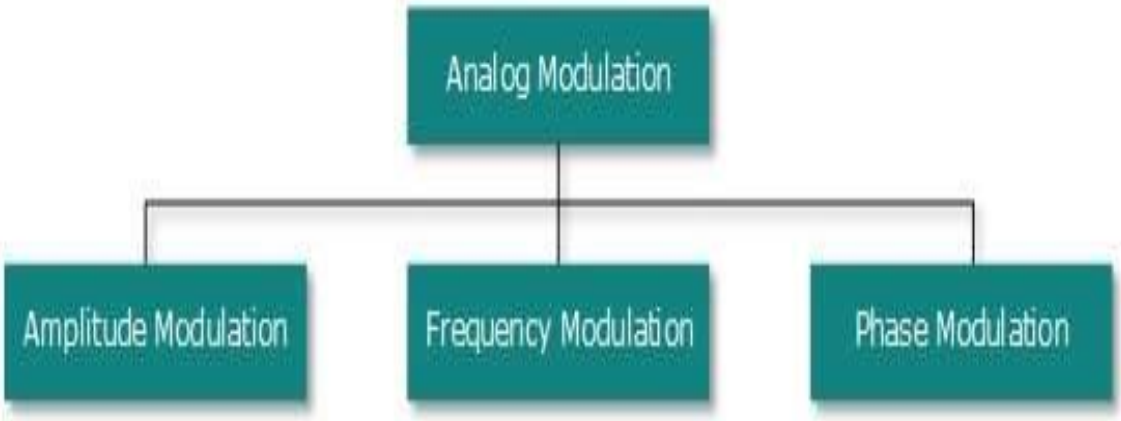
Reconstruction Filter: After the digital-to-analog conversion is done by the regenerative circuit and the decoder, a low-pass filter is employed, called as the reconstruction filter to get back the original signal.

Hence, the Pulse Code Modulator circuit digitizes the given analog signal, codes it and samples it, and then transmits it in an analog form. This whole process is repeated in a reverse pattern to obtain the original signal.

3.5 Analog data analog signals

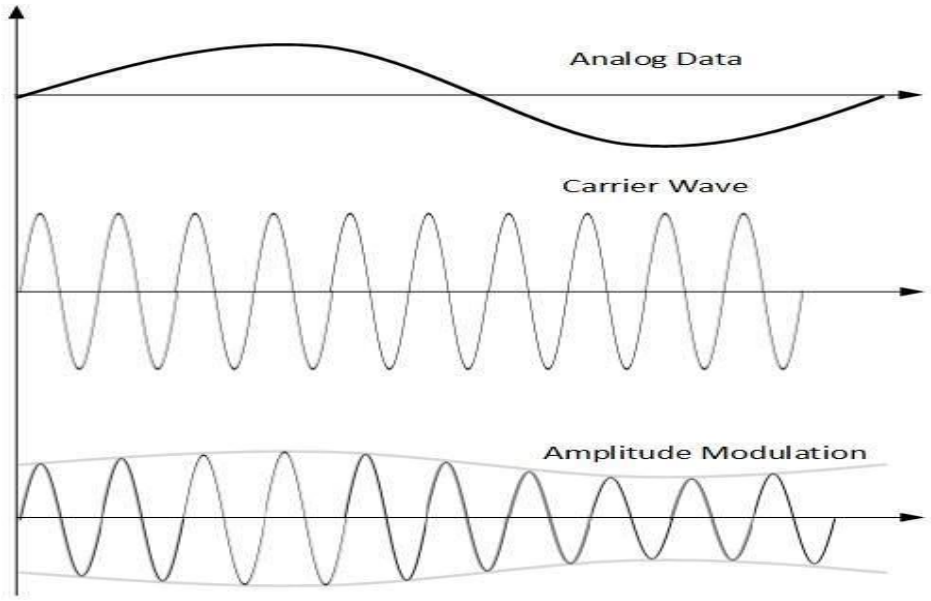
Analog-to-Analog Conversion

Analog signals are modified to represent analog data. This conversion is also known as Analog Modulation. Analog modulation is required when bandpass is used. Analog to analog conversion can be done in three ways:



- **Amplitude Modulation**

  In this modulation, the amplitude of the carrier signal is modified proportional to the amplitude of analog data.



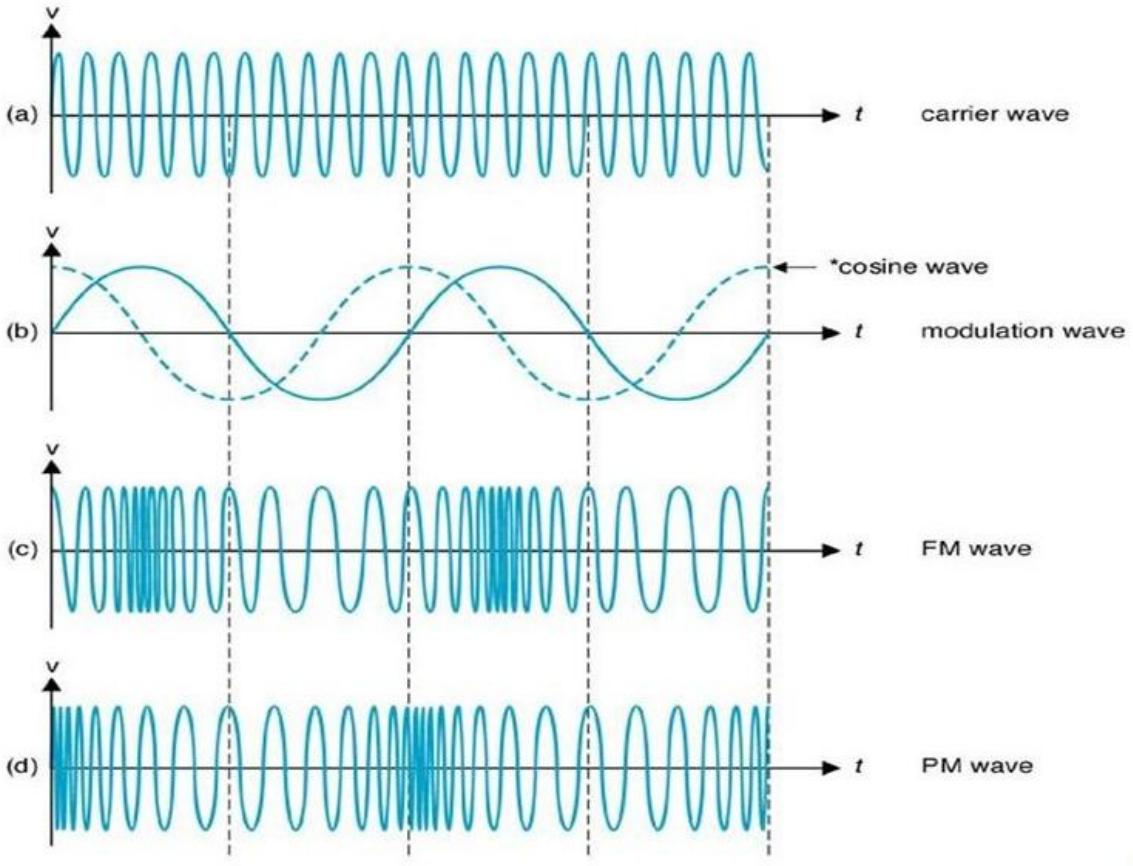
  Amplitude modulation is implemented by means of a multiplier. The amplitude of modulating signal (analog data) is multiplied by the amplitude of carrier frequency, which then reflects analog data.

  The frequency and phase of carrier signal remain unchanged.

- **Frequency Modulation**

  In frequency modulation (FM), the frequency of the carrier wave is varied in such a way that the change in frequency at any instant is proportional to another signal that varies with time. Its principal application is

also in radio, where it offers increased noise immunity and decreased distortion over the AM transmissions at the expense of greatly increased bandwidth.



- **Phase Modulation**

  In the modulation technique, the phase of carrier signal is modulated in order to reflect the change in voltage (amplitude) of analog data signal. Phase modulation is practically similar to Frequency Modulation, but in Phase modulation frequency of the carrier signal is not increased. Frequency of carrier is signal is changed (made dense and sparse) to reflect voltage change in the amplitude of modulating signal.

Unit-4. Data Communication & Data link control

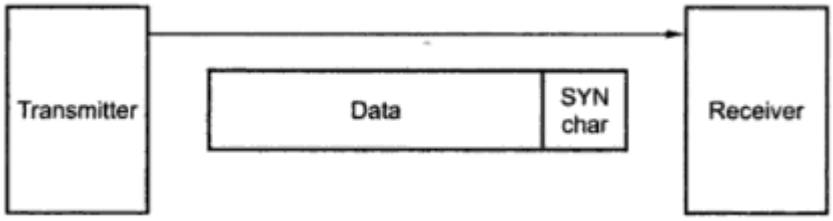
4.1 Asynchronous and Synchronous Transmission

Synchronous Transmission
In synchronous transmission, data moves in a completely paired approach, in the form of chunks or frames. Synchronisation between the source and target is required so that the source knows where the new byte begins, since there are no spaces included between the data.

Synchronous transmission is effective, dependable, and often utilised for transmitting a large amount of data. It offers real-time communication between linked devices.

An example of synchronous transmission would be the transfer of a large text file. Before the file is transmitted, it is first dissected into *blocks* of sentences. The blocks are then transferred over the communication link to the target location.

Because there are no beginning and end bits, the data transfer rate is quicker but there's an increased possibility of errors occurring. Over time, the clocks will get out of sync, and the target device would have the incorrect time, so some bytes could become damaged on account of lost bits. To resolve this issue, it's necessary to regularly re-synchronise the clocks, as well as to make use of check digits to ensure that the bytes are correctly received and translated.



Synchronous Data Format

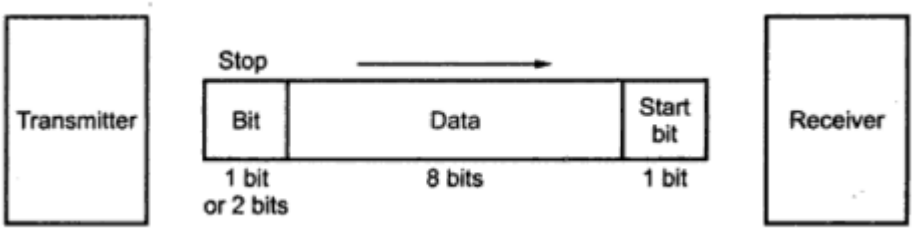
Characteristics of Synchronous Transmission

- There are no spaces in between characters being sent.
- Timing is provided by modems or other devices at the end of the transmission.
- Special 'syn' characters goes before the data being sent.
- The syn characters are included between chunks of data for timing functions.

Examples of Synchronous Transmission

- Chatrooms
- Video conferencing
- Telephonic conversations
- Face-to-face interactions

Asynchronous Transmission
In asynchronous transmission, data moves in a half-paired approach, 1 byte or 1 character at a time. It sends the data in a constant current of bytes. The size of a character transmitted is 8 bits, with a parity bit added both at the beginning and at the end, making it a total of 10 bits. It doesn't need a clock for integration—rather, it utilises the parity bits to tell the receiver how to translate the data.
It is straightforward, quick, cost-effective, and doesn't need 2-way communication to function.



Asynchronous Data Format

Characteristics of Asynchronous Transmission

- Each character is headed by a beginning bit and concluded with one or more end bits.
- There may be gaps or spaces in between characters.

Examples of Asynchronous Transmission

- Emails
- Forums
- Letters
- Radios
- Televisions

4.2 Error Detection

Error:

A condition when the receiver's information does not match with the sender's information. During transmission, digital signals suffer from noise that can introduce errors in the binary bits travelling from sender to receiver. That means a 0 bit may change to 1 or a 1 bit may change to 0.

Error Detecting Codes (Implemented either at Data link layer or Transport Layer of OSI Model)

Whenever a message is transmitted, it may get scrambled by noise or data may get corrupted. To avoid this, we use error-detecting codes which are additional data added to a given digital message to help us detect if any error has occurred during transmission of the message.

Basic approach used for error detection is the use of redundancy bits, where additional bits are added to facilitate detection of errors. Some popular techniques for error detection are :

1. Simple Parity check
2. Two-dimensional Parity check
3. Checksum
4. Cyclic redundancy check

1. Simple Parity check

Blocks of data from the source are subjected to a check bit or parity bit generator form, where a parity of :

- 1 is added to the block if it contains odd number of 1's, and
- 0 is added if it contains even number of 1's

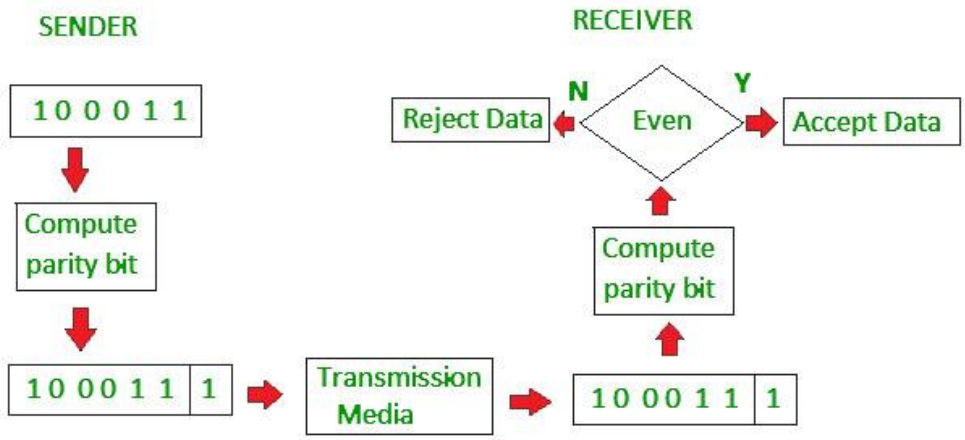
This scheme makes the total number of 1's even, that is why it is called even parity checking.

1. Simple Parity check

Blocks of data from the source are subjected to a check bit or parity bit generator form, where a parity of :

- 1 is added to the block if it contains odd number of 1's, and
- 0 is added if it contains even number of 1's

This scheme makes the total number of 1's even, that is why it is called even parity checking.

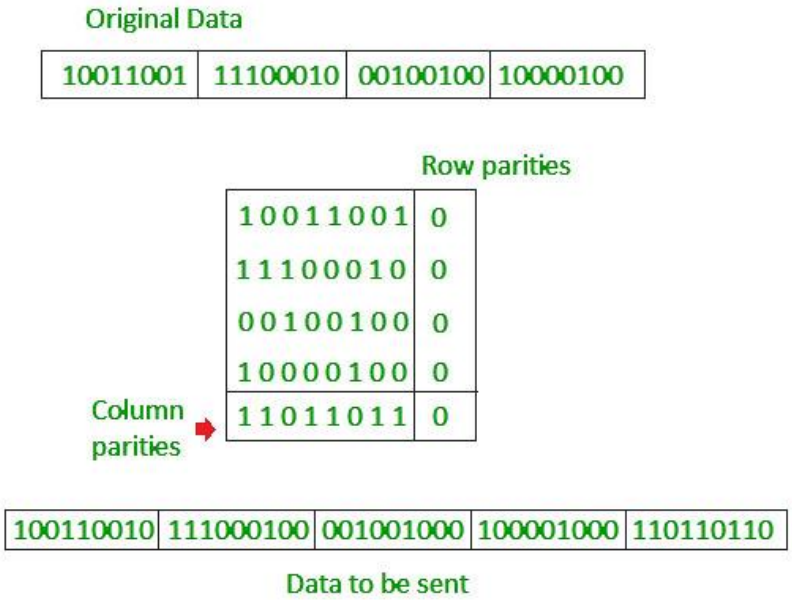


2. Two-dimensional Parity check

Parity check bits are calculated for each row, which is equivalent to a simple parity check bit. Parity check bits are also calculated for all columns, then both are sent along with the data. At the receiving end these are compared with the parity bits calculated on the received data.

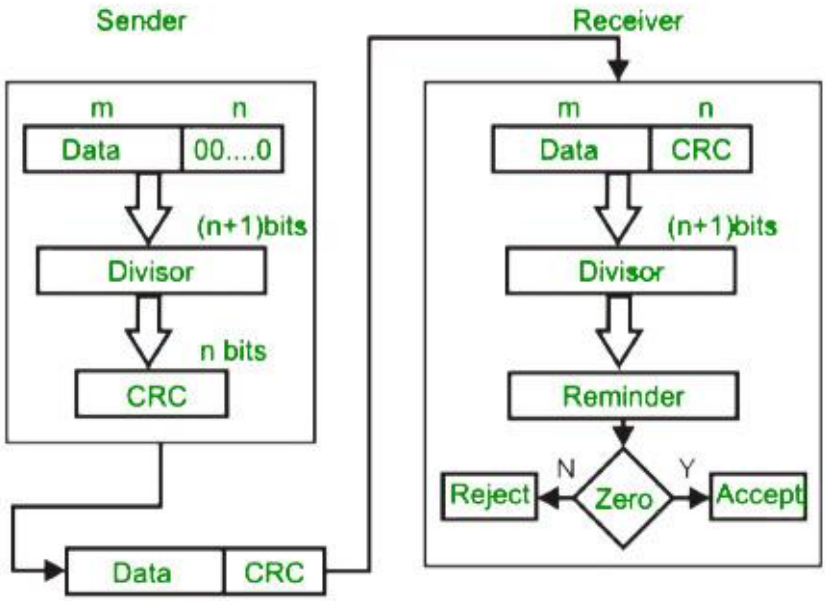
3. Checksum

- In checksum error detection scheme, the data is divided into k segments each of m bits.
- In the sender's end the segments are added using 1's complement arithmetic to get the sum. The sum is complemented to get the checksum.
- The checksum segment is sent along with the data segments.
- At the receiver's end, all received segments are added using 1's complement arithmetic to get the sum. The sum is complemented.
- If the result is zero, the received data is accepted; otherwise discarded.

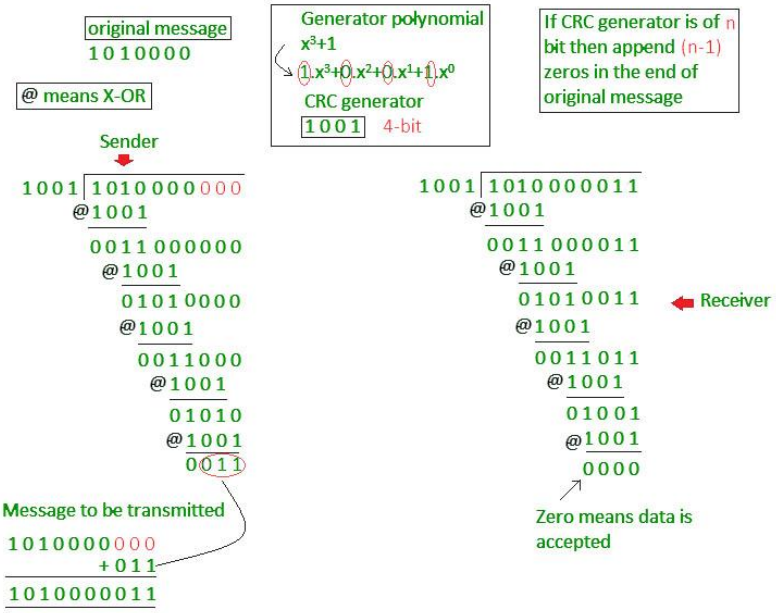


4. Cyclic redundancy check (CRC)

- Unlike checksum scheme, which is based on addition, CRC is based on binary division.
- In CRC, a sequence of redundant bits, called cyclic redundancy check bits, are appended to the end of data unit so that the resulting data unit becomes exactly divisible by a second, predetermined binary number.
- At the destination, the incoming data unit is divided by the same number. If at this step there is no remainder, the data unit is assumed to be correct and is therefore accepted.
- A remainder indicates that the data unit has been damaged in transit and therefore must be rejected.



Example :



4.3 Line Configuration

Line Configuration

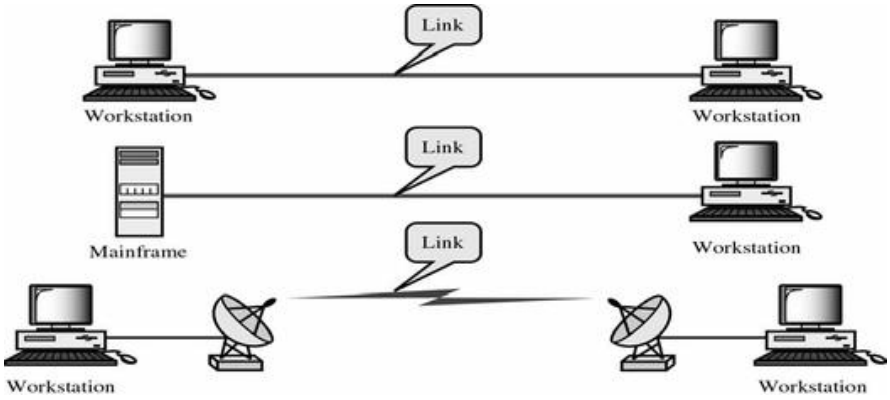
Line configuration refers to the way two or more communication devices attached to a link. Line configuration is also referred to as connection. A Link is the physical communication pathway that transfers data from one device to another. For communication to occur, two devices must be connected in same way to the same link at the same time. There are two possible line configurations.

1. Point-to-Point.
2. Multipoint.

Point-to-Point

A **Point to Point Line Configuration** Provide dedicated link between two devices use actual length of wire or cable to connect the two end including microwave & satellite link. Infrared remote control & tvs remote control.
The entire capacity of the channel is reserved for transmission between those two devices. Most point-to-point line configurations 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.

Point to point network topology is considered to be one of the easiest and most conventional network topologies. It is also the simplest to establish and understand. To visualize, one can consider point to point network topology as two phones connected end to end for a two way communication

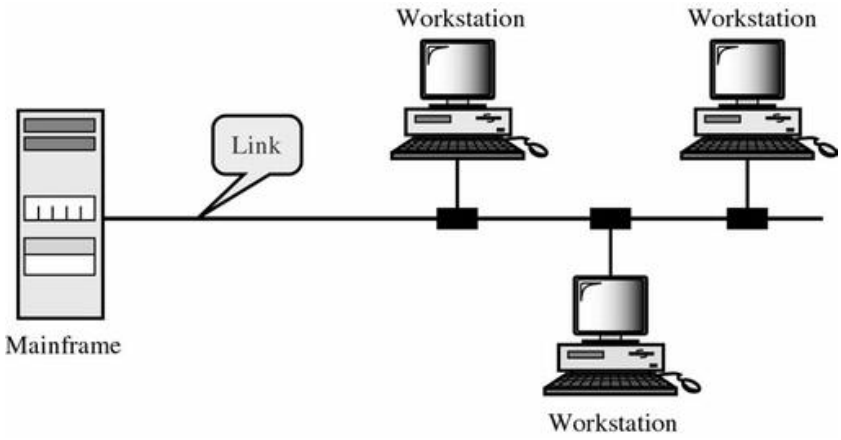


Multipoint Configuration

Multipoint Configuration also known as **Multidrop line configuration** one or more than two specific devices share a single link capacity of the channel is shared.

More than two devices share the Link that is the capacity of the channel is shared now. With shared capacity, there can be two possibilities in a Multipoint Line Config:

- **Spatial Sharing**: If several devices can share the link simultaneously, its called Spatially shared line configuration
- **Temporal (Time) Sharing**: If users must take turns using the link , then its called Temporally shared or Time Shared Line Configuration



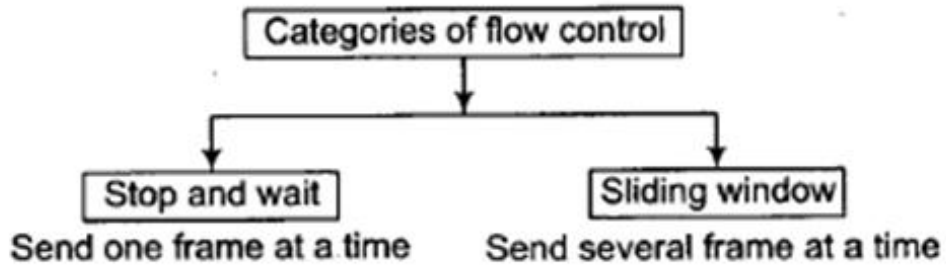
4.4 Flow Control

Flow Control: Flow control coordinates that amount of data that can be sent before receiving an acknowledgement.

- It is one of the most important duties of the data link layer.
- Flow control tells the sender how much data to send.
- It makes the sender wait for some sort of an acknowledgement (ACK) before continuing to send more data.
- Flow Control Techniques: **Stop-and-wait, and Sliding Window**

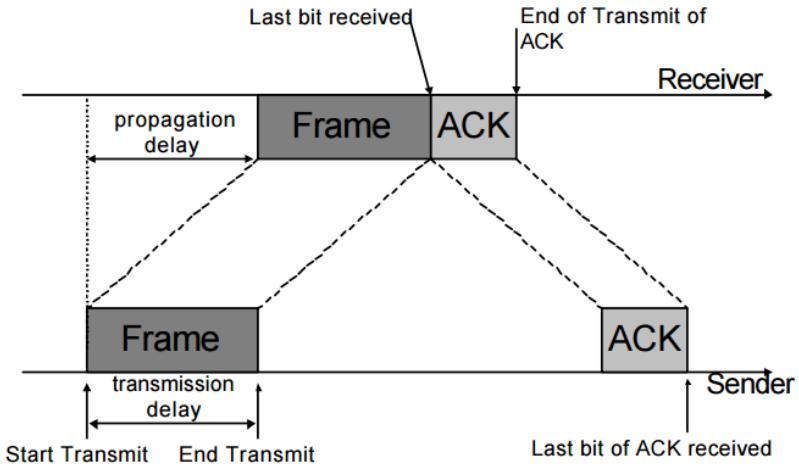
Flow Control Techniques:

- One important aspect of the data link layer is flow control.
- Flow control refers to a set of procedures used to restrict the amount of data the sender can send before waiting for acknowledgement.



Stop and Wait Flow control:

- The sender has to wait for an acknowledgment of every frame that it sends.
- Only when a acknowledgment has been received is the next frame sent. This process continues until the sender transmits an End of Transmission (EOT) frame.
- In Stop-and-Wait flow control, the receiver indicates its readiness to receive data for each frame.



- For every frame that is sent, there needs to be an acknowledgement, which takes a similar amount of propagation time to get back to the sender.
- Only one frame can be in transmission at a time. This leads to inefficiency if propagation delay is much longer than the transmission delay

Advantages: It's simple and each frame is checked and acknowledged well.

Disadvantage: Only one frame can be in transmission at a time. It is inefficient, if the distance between devices is long. Reason is propagation delay is much longer than the transmission delay.

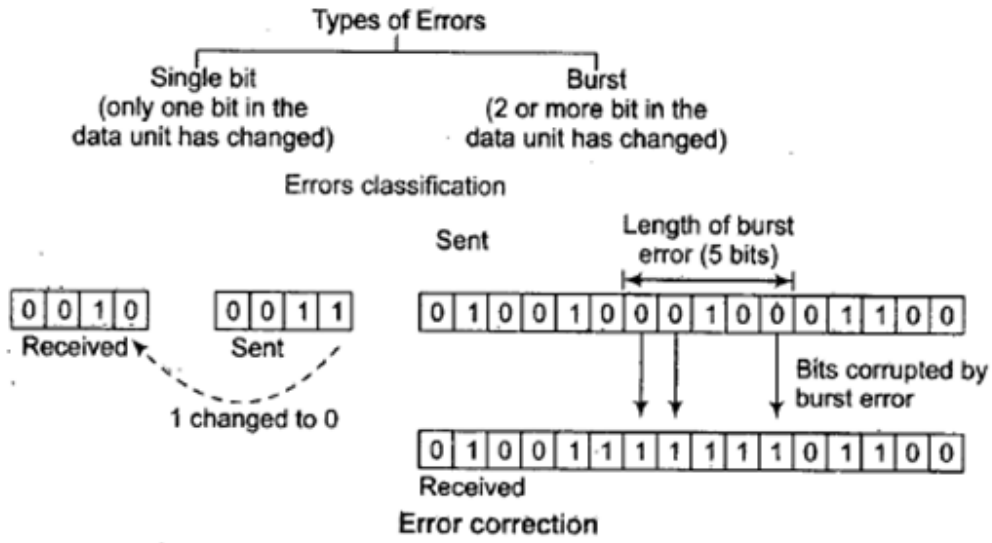
Sliding Window Flow Control:

- It works by having the sender and receiver have a "window" of frames.
- Each frame has to be numbered in relation to the sliding window. For a window of size n, frames get a number from 0 to n - 1. Subsequent frames get a number mod n.
- The sender can send as many frames as would fit into a window.
- The receiver, upon receiving enough frames, will respond with an acknowledgment of all frames up to a certain point in the window. It is called slide.
- This window can hold frames at either end and provides the upper limit on the number of frames that can be transmitted before requiring an acknowledgement. For example, if n = 8, the frames are numbered 0, 1, 2, 3, 4, 5, 6, 7, 0, 1, 2, 3, 4, 5, 6, 7, 0, 1...so on. The size of the window is (n -1) = 7.
- When the receiver sends an ACK, it includes the number of the next frame it expects to receive. When the receiver sends an ACK containing the number 5, it means all frames upto number 4 have been received.
- If the window size is sufficiently large the sender can continuously transmit packets:
    - If W >= (2a+1), sender can transmit continuously. **(Efficiency =1)**
    - If W < (2a+1), sender can transmit W frames every (2a+1) time units. **(Efficiency = W/(1+2a))**

4.5 Error Control

Error Control Techniques:

- Many factors including line noise can alter or wipe out one or more bits of a given data unit.



- Reliable systems must have mechanism for detecting and correcting such errors.
- Error detection and correction are implemented either at the data link layer or the transport layer of the OSI model.
- Error detection techniques have already been covered in 4.2 Section.

Error Correction:

- Error Correction in data link is implemented simply anytime.
- An error is detected in an exchange, a negative acknowledgement NAK is returned and the specified frames are retransmitted. This process is called Automatic Repeat Request(ARQ).
- Retransmission of data happens in three Cases: Damaged frame, Lost frame and Lost acknowledgement.

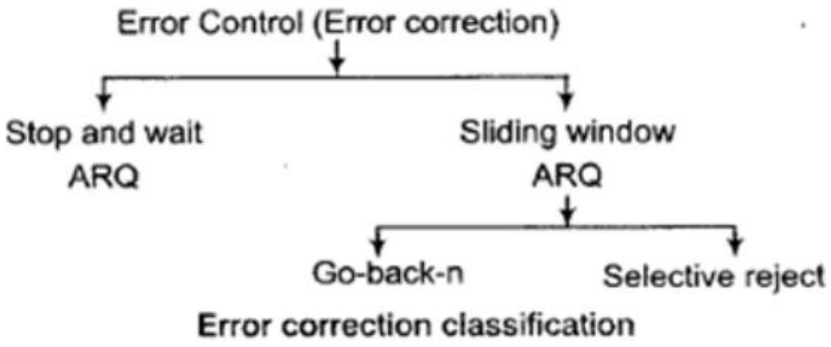


Fig.Different Error Control (or) Correction Techniques

Stop and Wait ARQ:

- Include re-transmission of data in case of lost or damaged framer.
- It is addition to the basic flow control mechanism with re-transmissions.
    - Sender sends an information frame to receiver.

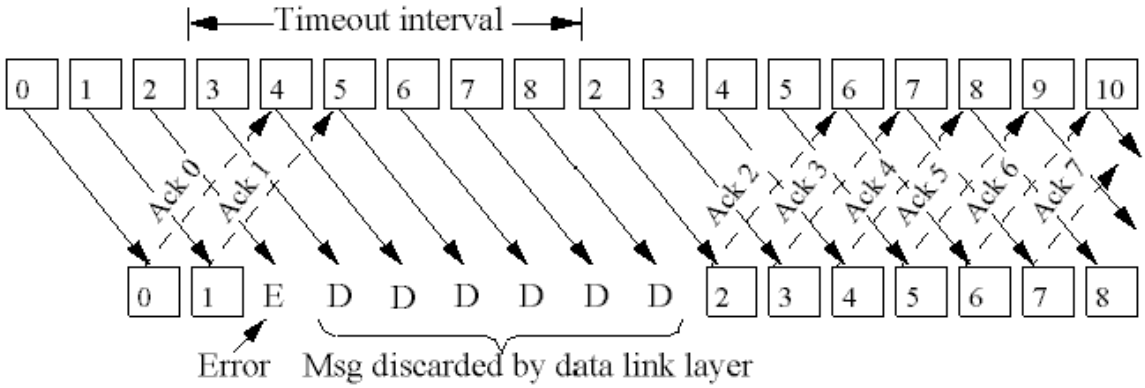
- Sender waits for an ACK before sending the next frame.
- Receiver sends an ACK if frame is correctly received.
- If no ACK arrives within time-out, sender will resend the frame.
    - **Time-out period > Rround trip time**
- If an error is discovered in a data frame, indicating that it has been corrupted in transit, a NAK frame is returned. NAK frames, which are numbered, tell the sender to retransmit the last frame sent.
- **Piggybacking:** In bidirectional communications, both parties send & acknowledge data, i.e. both parties implement flow control. Outstanding ACKs are placed in the header of information frames, piggybacking can save bandwidth since the overhead from a data frame and an ACK frame (addresses, CRC, etc) can be combined into just one frame.

Sliding Window ARQ:

- To cover retransmission of lost or damaged frames, some features are added to the basic flow control mechanism of sliding window.
- A Sender may send multiple frames as allowed by the window size.
- The sending device keeps copies of all transmitted frames, until they have been acknowledged. .
- In addition to ACK frames, the receiver has the option of returning a NAK frame, if the data have been received damaged. NAK frame tells the sender to retransmit a damaged frame.
- Here, both ACK and NAK frames must be numbered for identification.
- ACK frames carry the number of next frame expected.
- NAK frames on the other hand, carry the number of the damaged frame itself.
- If the last ACK was numbered 3, an ACK 6 acknowledges the receipt of frames 3, 4 and 5 as well.
- If data frames 4 and 5 are received damaged, both NAK 4 and NAK 5 must be returned.
- Like stop and wait ARQ, the sending device in sliding window ARQ is equipped with a timer to enable it to handle lost acknowledgements.
- Sliding window ARQ is two types: Go-back-n ARQ, and Selective Reject ARQ.
- There are two ACK processing methods in sliding windows:
    - Selective ACK: The **ACK N** message acknowledges **only** the frame with sequence number **N**
    - Cumulative ACK : The **ACK N** message acknowledges **all** frames with sequence number **<= N**

(i) Go-back-n ARQ:

- The sliding window method using **cumulative ACK** is known as the **Go-Back-N** ARQ protocol.
- Receiver window size is 1.
- In this method, if one frame is lost or damaged all frames sent, since the last frame acknowledged are retransmitted.
- For example, sender may send frames 1,2,3,4 and get an NAK with a value of 2. The NAK acknowledges everything that came before it, and asks for frame 2 (and subsequent frames) to be resent.
- NAK number refer to the next expected frame number.
- Example: In the following figure, frame 2 has an error, then all subsequent frames are discarded. After timeout sender sends all frames from frame 2.



- Damaged/Error Frame :
  - In go-back-n ARQ, The receiver sends the NAK for this frame along with that frame number, that it expects to be retransmitted.
  - After sending NAK, the receiver discards all the frames that it receives, after a damaged frame.
  - The receiver does not send any ACK (acknowledgement) for the discarded frames.
    After the sender receives the NAK for the damaged frame, it retransmits all the frames onwards the frame number referred by NAK.
- Lost frame:
  - In go-back-n ARQ, Receiver easily detects the loss of a frame as the newly received frame is received out of sequence.
  - The receiver sends the NAK for the lost frame and then the receiver discards all the frames received after a lost frame.
  - The receiver does not send any ACK for that discarded frames.
  - After the sender receives the NAK for the lost frame, it retransmits the lost frame referred by NAK and also retransmits all the frames which it has sent after the lost frame.
- Lost Acknowledgement :
  - In go-back-n ARQ, If the sender does not receive any ACK or if the ACK is lost or damaged in between the transmission.
  - The sender waits for the time to run out and as the time run outs, the sender retransmits all the frames for which it has not received the ACK.
  - The sender identifies the loss of ACK with the help of a timer.
  - The ACK number, like NAK number, shows the number of the frame, that receiver expects to be the next in sequence.
  - The window size of the receiver is 1 as the data link layer only require the frame which it has to send next to the network layer.
  - The sender window size is equal to 'w'. If the error rate is high, a lot of bandwidth is lost wasted.

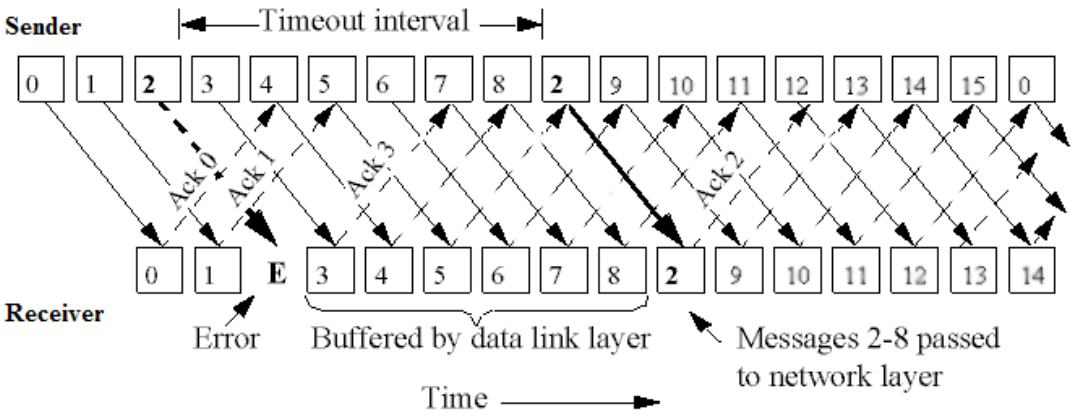
(ii) Selective Repeat ARQ:

- Selective Repeat ARQ overcomes the limitations of Go-Back-N by adding two new features:
  - Receiver window > 1 frame: Out-of-order but error-free frames can be accepted
  - Retransmission mechanism is modified: Only individual frames are retransmitted

- In this method, only specific damaged or lost frame is retransmitted
- Sender only retransmits frames for which a NAK is received.

- NAK number refer to the frame lost.
- If a frame is corrupted in transmit, a NAK is returned and the frame is resent out of sequence.
- The sender needs to maintain all data that hasn't been aknowledged yet.
- The receiving device must be able to sort the frames it has and insert the retransmitted frame into its proper place in the sequence.
- It has advantage that few re-transmissions than go-back-n. But complexity at sender and receiver is involved.

- Example: Frame 2 has an error, so receiver maintains buffer to store the next frames.



- Damaged frames :
  - In Selective reject, If a receiver receives a damaged frame, it sends the NAK for the frame in which error or damage is detected.
  - The NAK number, like in go-back-n also indicate the acknowledgement of the previously received frames and error in the current frame.
  - The receiver keeps receiving the new frames while waiting for the damaged frame to be replaced.
  - The frames that are received after the damaged frame are not be acknowledged until the damaged frame has been replaced.
- Lost Frame :
  - As in a selective repeat protocol, a frame can be received out of order and further they are sorted to maintain a proper sequence of the frames.
  - While sorting, if a frame number is skipped, the receiver recognise that a frame is lost and it sends NAK for that frame to the sender.
  - After receiving NAK for the lost frame the sender searches that frame in its window and retransmits that frame.
  - If the last transmitted frame is lost then receiver does not respond and this silence is a negative acknowledgement for the sender.
- Lost Acknowledgement :
  - In Selective reject, If the sender does not receive any ACK or the ACK is lost or damaged in between the transmission.
  - The sender waits for the time to run out and as the time run outs, the sender retransmit all the frames for which it has not received the ACK.
  - The sender identifies the loss of ACK with the help of a timer.

4.6 Multiplexing

Multiplexing:

Multiplexing is a technique used to combine and send the multiple data streams over a single medium. The process of combining the data streams is known as multiplexing and hardware used for multiplexing is known as a multiplexer.

Multiplexing is achieved by using a device called Multiplexer (**MUX**) that combines n input lines to generate a single output line. Multiplexing follows many-to-one, i.e., n input lines and one output line.

Demultiplexing is achieved by using a device called Demultiplexer (**DEMUX**) available at the receiving end. DEMUX separates a signal into its component signals (one input and n outputs). Therefore, we can say that demultiplexing follows the one-to-many approach.

Reason for using Multiplexing:

- The transmission medium is used to send the signal from sender to receiver. The medium can only have one signal at a time.
- If there are multiple signals to share one medium, then the medium must be divided in such a way that each signal is given some portion of the available bandwidth. For example: If there are 10 signals and bandwidth of medium is100 units, then the 10 unit is shared by each signal.
- When multiple signals share the common medium, there is a possibility of collision. Multiplexing concept is used to avoid such collision.
- Transmission services are very expensive.

History of Multiplexing:

- Multiplexing technique is widely used in telecommunications in which several telephone calls are carried through a single wire.
- Multiplexing originated in telegraphy in the early 1870s and is now widely used in communication.
- George Owen Squier developed the **telephone carrier multiplexing** in 1910.

Concept of Multiplexing:

- The 'n' input lines are transmitted through a multiplexer and multiplexer combines the signals to form a composite signal.
- The composite signal is passed through a Demultiplexer and demultiplexer separates a signal to component signals and transfers them to their respective destinations.

Advantages of Multiplexing:

- More than one signal can be sent over a single medium.
- The bandwidth of a medium can be utilized effectively.

4.7 FDM synchronous TDM

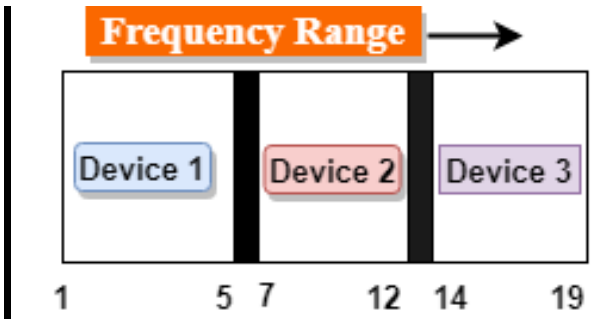
Multiplexing techniques can be classified as:

- Frequency Division Multiplexing(FDM)
- Time Division Multiplexing(TDM)
- Wavelength Division Multiplexing(WDM)

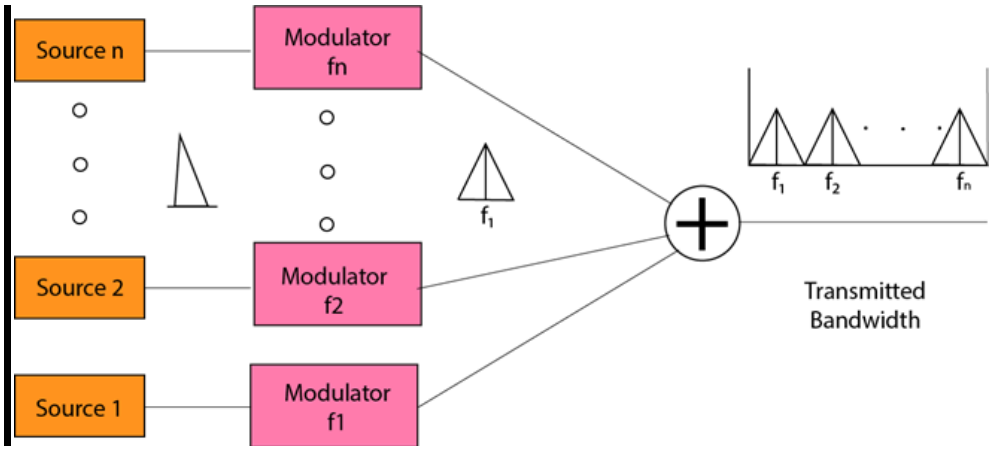
Here we will study about FDM and TDM in detail.

Frequency-division Multiplexing (FDM)

- It is an analog technique.
- **Frequency Division Multiplexing** is a technique in which the available bandwidth of a single transmission medium is subdivided into several channels.



- In the above diagram, a single transmission medium is subdivided into several frequency channels, and each frequency channel is given to different devices. Device 1 has a frequency channel of range from 1 to 5.
- The input signals are translated into frequency bands by using modulation techniques, and they are combined by a multiplexer to form a composite signal.
- The main aim of the FDM is to subdivide the available bandwidth into different frequency channels and allocate them to different devices.
- Using the modulation technique, the input signals are transmitted into frequency bands and then combined to form a composite signal.
- The carriers which are used for modulating the signals are known as **sub-carriers**. They are represented as f1,f2..fn.
- **FDM** is mainly used in radio broadcasts and TV networks.



Advantages Of FDM:

- FDM is used for analog signals.
- FDM process is very simple and easy modulation.
- A Large number of signals can be sent through an FDM simultaneously.
- It does not require any synchronization between sender and receiver.

Disadvantages Of FDM:

- FDM technique is used only when low-speed channels are required.
- It suffers the problem of crosstalk.
- A Large number of modulators are required.
- It requires a high bandwidth channel.

Applications Of FDM:

- FDM is commonly used in TV networks.
- It is used in FM and AM broadcasting. Each FM radio station has different frequencies, and they are multiplexed to form a composite signal. The multiplexed signal is transmitted in the **air.**

Time Division Multiplexing:

- It is a digital technique.
- In Frequency Division Multiplexing Technique, all signals operate at the same time with different frequency, but in case of Time Division Multiplexing technique, all signals operate at the same frequency with different time.
- In **Time Division Multiplexing technique**, the total time available in the channel is distributed among different users. Therefore, each user is allocated with different time interval known as a Time slot at which data is to be transmitted by the sender.
- A user takes control of the channel for a fixed amount of time.
- In Time Division Multiplexing technique, data is not transmitted simultaneously rather the data is transmitted one-by-one.
- In TDM, the signal is transmitted in the form of frames. Frames contain a cycle of time slots in which each frame contains one or more slots dedicated to each user.
- It can be used to multiplex both digital and analog signals but mainly used to multiplex digital signals.

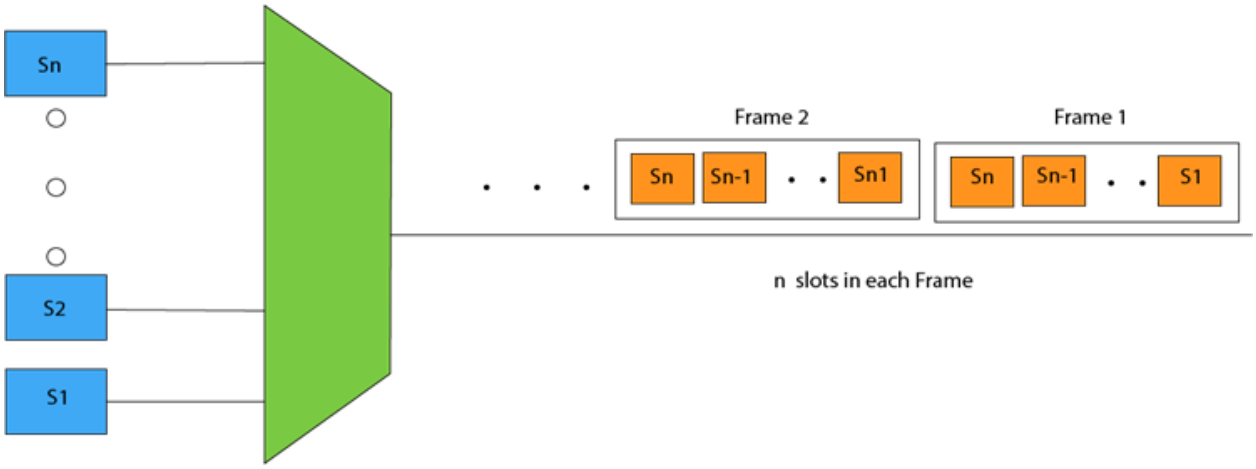
There are two types of TDM:

- Synchronous TDM
- Asynchronous TDM

Synchronous TDM

- A Synchronous TDM is a technique in which time slot is preassigned to every device.
- In Synchronous TDM, each device is given some time slot irrespective of the fact that the device contains the data or not.
- If the device does not have any data, then the slot will remain empty.
- In Synchronous TDM, signals are sent in the form of frames. Time slots are organized in the form of frames. If a device does not have data for a particular time slot, then the empty slot will be transmitted.
- The most popular Synchronous TDM are T-1 multiplexing, ISDN multiplexing, and SONET multiplexing.

- If there are n devices, then there are n slots.



Concept Of Synchronous TDM:

In the above figure, the Synchronous TDM technique is implemented. Each device is allocated with some time slot. The time slots are transmitted irrespective of whether the sender has data to send or not.

Disadvantages Of Synchronous TDM:

- The capacity of the channel is not fully utilized as the empty slots are also transmitted which is having no data. In the above figure, the first frame is completely filled, but in the last two frames, some slots are empty. Therefore, we can say that the capacity of the channel is not utilized efficiently.
- The speed of the transmission medium should be greater than the total speed of the input lines. An alternative approach to the Synchronous TDM is Asynchronous Time Division Multiplexing.

4.8 Statistical TDM

Statistical Time Division Multiplexing:

In case of TDM, time slots are allocated to channels, even if they have no information to transmit. This is just wastage of the bandwidth and to overcome this inefficiency of standard TDM, a technique known as STDM has been developed where time is allocated to lines only when it is required. This is achieved with the use of intelligent devices that are capable of identifying when a terminal is idle.

Because of the intelligent device statistically compensates for normal idle time, more lines can be connected to a transmission medium.

During the peak traffic period a buffer memory temporarily stores the data, so high-speed line time can be effectively utilized with active channels. It adopts a methodology where each transmission has an identification information (a channel identifier). This increases the overheads, which are handled by grouping a number of characters for each channel together for transmission. It is also referred as "Intelligent" TDM. In this case, data rate capacity is well below the sum of connected capacity of each channel because it utilizes the idle time very effectively. It is digital only and requires more complex framing of data.

It is widely used for remote communications with multiple terminals. The additional services such as data compression, line priority, mixed speed lines, host ports sharing; network port control, automatic speed detection etc are available with STDM techniques.

Unit 5. Switching & Routing

5.1 Circuit Switching networks

Routing:

- Routing deals with moving packets between different networks.
- Routers operate at layer 3 of the OSI model.
- A router can find where to send a packet using Network ID within the network layer header.
- It will use the routing table to determine the route to the destination host.

Switching:

- Switching deals with moving packets between devices on the same network.
- A switch operates at layer 2 of the OSI model.
- A switch is also referred to as Multi-Port Bridge, It allows us to find where the packet should be sent by examining the MAC Address within the data link header of the packet.
- A switch maintains a database of MAC Addresses and which port they are connected to.
- There are basically three types of switching methods are made available.
    1) Circuit Switching
    2) Packet Switching
    3) Message Switching

- Out of three methods, circuit switching and packet switching are commonly used but the message switching has been opposed out in the general communication procedure but is still used in the networking application. In this chapter we will deal with Circuit Switching and Packet Switching.

Circuit Switching:

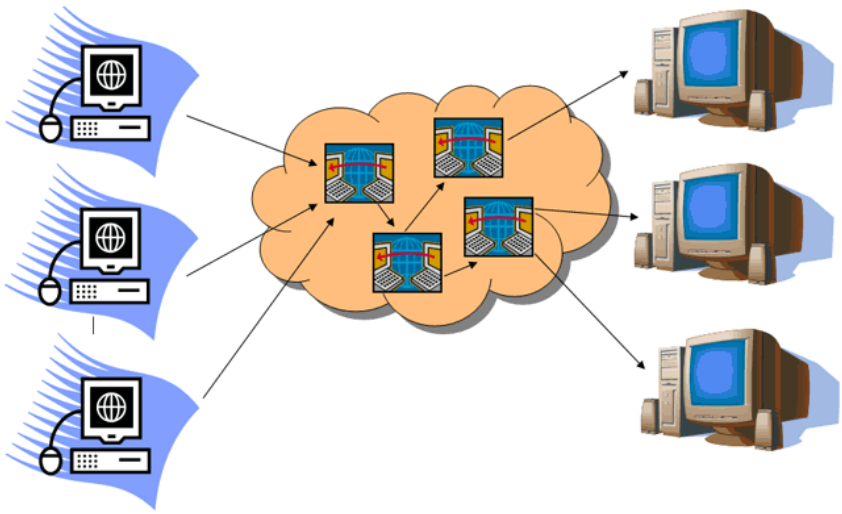
Circuit switching is a switching method where an end-to-end path is created between two stations within a network before starting the data transfer.

Circuit switching has **three** phases:

1. Circuit establishment,
2. Transferring the data
3. Circuit disconnect.

Circuit switching method has a fixed data rate and both the subscribers need to operate at this fixed rate. Circuit switching is the simplest method of data communication where dedicated physical connections are established between two individual senders and receiver. To create these dedicated connections, a set of switches are connected by physical links.

In the below image, three computers on the left side are connected with three desktops PCs on the right side with physical links, depending on the four circuit switchers. If the circuit switching is not used, they need to be connected with point-to-point connections, where many number of dedicated lines are required, which will not only increase the connection cost but also increase the complexity of the system.



The routing decision, in the case of circuit switching, is made when the routing path is being established in the network. After the dedicated routing path is established the data continuously submitted to the receiver destination. The connection is maintained until the end of the conversation.

Three Phases in Circuit switching Communication: The start to the end communication in Circuit Switching is done using this formation as below.

During the Setup phase, in the circuit switching network, a dedicated routing or connection path is established between the sender and the receiver. At this period End to End addressing, like source address, destination address is must create a connection between two physical devices. The circuit switching happens in the physical layers.

Data transfer only happens after the setup phase is completed and only when a physical, dedicated path is established. No addressing method is involved in this phase. The switches use time slot (TDM) or the occupied band (FDM) to route the data from the sender to the receiver. One thing needs to keep in mind that the data sending is continuous and there may be periods of silence in data transmitting. All internal connections are made in duplex form.

At the final Circuit disconnect phase, when any one of the subscriber in the network, sender or receiver needs to disconnect the path, a disconnect signal is sent to all involved switches to release the resource and break the connection. This phase also called as Teardown phase in circuit switching method.

A Circuit switch creates a temporary connection between an input link with an output link. There are various types of switches available with multiple inputs and output lines.

Generally, Circuit Switching is used in Telephone Lines.

Advantages of Circuit Switching

Circuit Switching Method provides large advantages in specific cases. The Advantages are as follows:

1. The data rate is fixed and dedicated because the connection is established using dedicated physical connection or circuits.
2. As there are dedicated transmission routing paths involved, it is a good choice for continuous transmission over a long duration.
3. The data transmission delay is negligible. No waiting time is involved in switches. So, the data gets transmitted without any prior delay in the transmission. This is definitely a positive advantage of Circuit Switching method.

Disadvantages of Circuit Switching

Other than the Advantages, Circuit switching also have some disadvantages.

1. Whether the communication channel is free or busy, the dedicated channel could not be used for other data transmission.
2. It requires more bandwidth, and continuous transmission offers wastage of bandwidth when there is a silence period.
3. It is highly inefficient when utilizing the system resource. We cannot use the resource for other connection as it is allocated for the entire conversation.
4. It takes huge time during the establishment of physical links between senders and receivers.

5.2 Packet Switching Principles

Packet switching is a method of data transfer where the data is broken into small pieces of variable lengths and then transmitted to the network line. Broken pieces of data are called as **packets**. After receiving those broken data or packets, all are reassembled at the destination and thus making a complete file. Due to this method, the data gets transferred fast and in an efficient manner. In this method, no pre-setup or resource reservation is required like circuit switching method.

This method use Store and Forward techniques. So each hop will store the packet first and then forward the packets to the next host destination. Each packet contains control information, source address and destination address. Due to this packets can use any route or paths in an existing network.

Advantages of Packet Switching:

Packet switching offers advantages over the circuit switching. Packet switching network is designed to overcome the drawbacks of Circuit Switching method.

1. Efficient in terms of Bandwidth.
2. Transmission delay is minimum
3. Missing packets can be detected by the destination.
4. Cost-effective implementation.
5. Reliable when busy path or links breakdown is detected in the network. Packets can be transmitted by other links or can use a different path.

Disadvantages of Packet Switching:

Packet switching also encounters few drawbacks.

1. Packet switching does not follow any particular order to transmit the packet one by one.
2. Packet missing occurs in large data transmission.
3. Each packet needs to be encoded with sequence numbers, Receiver and Senders address, and other information.
4. Routing is complex in the nodes as packets can follow multiple paths.
5. When rerouting occurs for some reason, the delay in receiving the packets is increased.

Differences between Circuit Switching and Packet Switching:

We already got an idea about what are the differences between circuit switching and packet switching. Let's see the differences in a table format for better understanding-

Differences	**Circuit switching**	**Packet Switching**
Steps Involvement	In circuit switching, 3 phase setup is required for total conversation. Connection Establishment, Data Transfer, Connection Teardown	In the case of Packet Switching, we can make data transfer directly.
Destination Address	Entire Path Address is provided by the source.	Each Data packets knows only the final destination address, the routing path depends on routers decision.
Data Processing	Data Processing takes place at the Source system.	Data Processing takes place at nodes and source systems.
Uniform Delay	Uniform Delay happens.	The delay between data units is not uniform.

Between data Units		
Reliability	Circuit Switching is more reliable compared with Packet Switching	Packet Switching is less reliable compared with Circuit Switching.
Resource Wasteage	Resource Wastage is high in Circuit Switching.	Resource wastage is Less in Packet Switching.
Store and Forward Technique	It does not use store and forward technique	It uses store and forward technique
Congestion	Congestion occurs at only connection Establishment time.	Contestation can occur on data transfer phase.
Transmission Data	The source makes the transmission of the data.	Transmission of data is done by the source, routers.

5.3 X.25

X.25

X.25 is a protocol suite defined by ITU-T for packet switched communications over WAN (Wide Area Network). It was originally designed for use in the 1970s and became very popular in 1980s. Presently, it is used for networks for ATMs and credit card verification. It allows multiple logical channels to use the same physical line. It also permits data exchange between terminals with different communication speeds.

X.25 has three protocol layers

- **Physical Layer:** It lays out the physical, electrical and functional characteristics that interface between the computer terminal and the link to the packet switched node. X.21 physical implementer is commonly used for the linking.
- **Data Link Layer:** It comprises the link access procedures for exchanging data over the link. Here, control information for transmission over the link is attached to the packets from the packet layer to form the LAPB frame (Link Access Procedure Balanced). This service ensures a bit-oriented, error-free, and ordered delivery of frames.
- **Packet Layer:** This layer defines the format of data packets and the procedures for control and transmission of the data packets. It provides external virtual circuit service. Virtual circuits may be of two types: virtual call and permanent virtual circuit. The virtual call is established dynamically when needed through call set up procedure, and the circuit is relinquished through call clearing procedure. Permanent virtual circuit, on the other hand, is fixed and network assigned.

Equipment used

- **X.21** implementer
- **DTE** : Data Terminal Equipment

- **DCTE** : Data Circuit Terminating Equipment

Characteristics of X.25:

In addition to the characteristics of the packet switched network, X.25 has the following characteristics:

1. Multiple logical channels can be set on a single physical line
2. Terminals of different communication speeds can communicate
3. The procedure for transmission controls can be changed.

X.25 Performance:

During 1976, when the X.25standard was released, it was capable of supporting a transmission speed of 64 Kbps. Unfortunately most of this bandwidth was used only for error checking. During 1992, ITU revised and issued a new X.25version, which was able to support a speed of 2.048Mbps. The France Telecom has been offering 2.08Mbps X.25 for many years. The packet based nature of the X.25 affect the performance very badly. During times when the traffic is extremely heavy, packet delivery delay is inevitable. Even though the routers are directing the packets around the congested areas, users still experience a poor performance. On the other hand packet-switching can accommodate burst of traffic over and above the maximum bandwidth however, circuit switching can accommodate inflexible and finite amount of traffic only.

X.25 protocol is highly reliable than any other protocol of this kind. Every router in the network, on receiving the packet, performs a complete check-up for the presence of errors before sending it to the next router. As a result, each node maintains a table, containing management, flow control, and error checking information against which, each packet is checked. In addition to this, the destination stations take responsibility to detect the lost or damaged packet and requests for a retransmission.

For every transmission of a packet to the next node, an acknowledgment is received from the next node in the data link layer. At the network layer, acknowledgment is sent back by the destination station to the source station as soon as the packet is received.

5.4 Routing in Packet switching

Centralized vs Distributed Routing

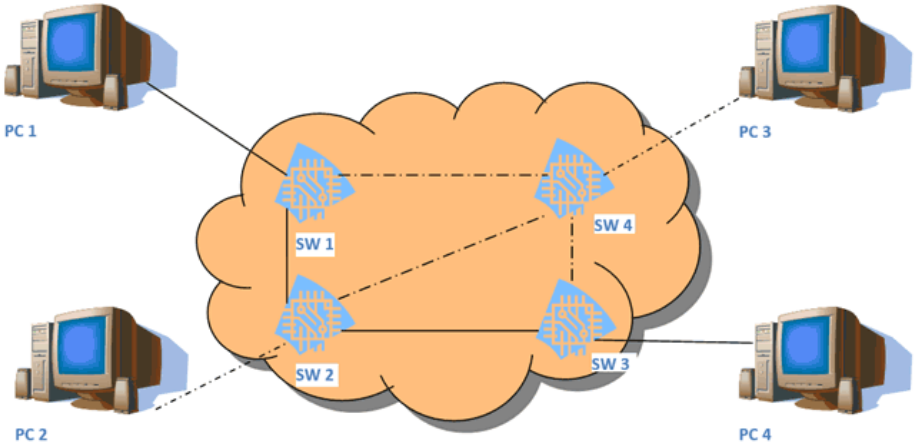
- **Centralized Routing :**
  - All routes determined by a central node.
  - All state information sent to central node.
  - Problems adapting to frequent topology changes
  - Does not scale
- **Distributed Routing:**
  - Routes determined by routers using distributed algorithm
  - State information exchanged by routers
  - Adapts to topology and other changes
  - Better scalability

Static vs Dynamic Routing

- **Static Routing:**
  - Set up manually, do not change; requires administration
  - Works when traffic predictable & network is simple
  - Used to override some routes set by dynamic algorithm
  - Used to provide default router
- **Dynamic Routing:**
  - Adapt to changes in network conditions
  - Automated
  - Calculates routes based on received updated network state information

Virtual Circuit Based Packet Switching:

VC based package switching is a mode of packet switching where a logical path or virtual circuit connection is done between sender and receiver. **VC stands for Virtual Circuit**. In this mode of packet switching operation, a predefined route is created and all packets will follow the predefined paths. All routers or switches which are involved in the logical connection are provided a unique Virtual Circuit ID to uniquely identify the virtual connections. It also **has the same three-phase protocol used in circuit switching, Setup Phase, Data Transfer Phase and Tear down Phase**.

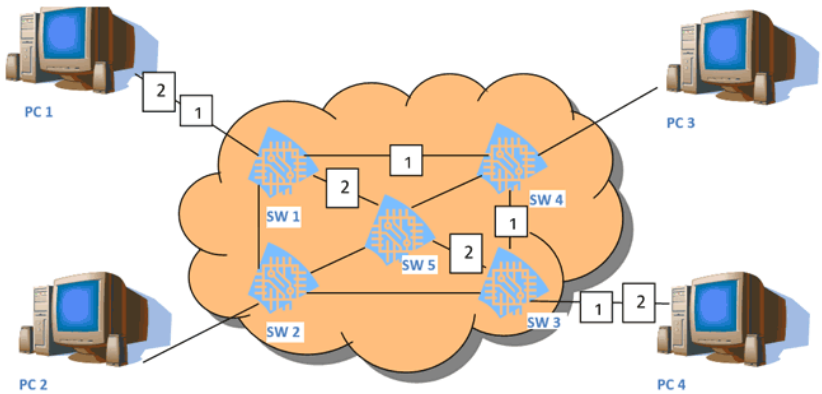


In the above image, 4 PCs are connected with a 4 switch network and the **data flow will be packet switching in Virtual circuit mode**. As we can see switches are connected with each other and share the communications path with each other. Now in the virtual circuit, a predefined route needs to be established. If we want to transfer data from PC1 to the PC 4 the path will be directed from the SW1 to SW2 to SW3 and then finally at PC4. This route is predefined and All SW1, SW2, SW3 are provided with a unique ID to identify the data paths, so the data is bound by the paths and could not choose another route.

Datagram Based Packet Switching:

Datagram switching is completely different from VC based packet switching technology. **In Datagram switching, the path is dependent on the data**. The Packets have all the necessary information like Source address, destination address and Port identity etc. So in connectionless datagram-based packet switching mode, each packet is treated independently. They can choose different routes and the routing decisions are made dynamically when data are transmitting inside the network. So, at the destination, the packets can be received out of order or in any sequence, there is no predefined route and the guaranteed packet delivery is not possible. To secure guaranteed packet receiving, additional end system protocols need to be configured.

In this mode of packet switching, there is no setup, transmit and teardown phase is involved.

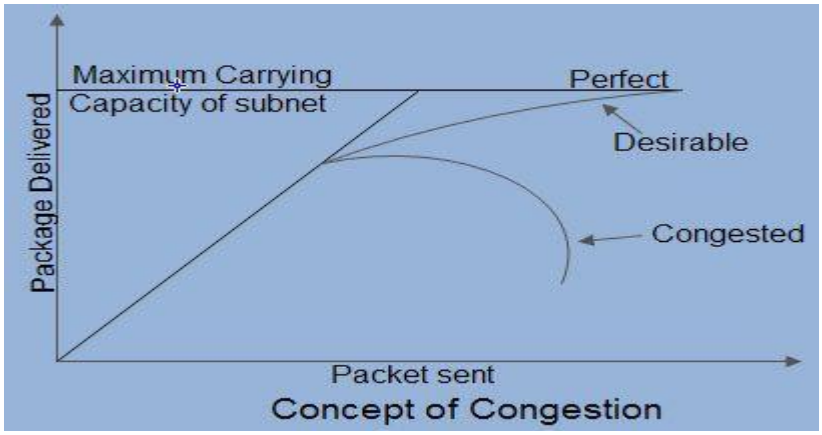


Again in the above image, 4 computers are connected and we transferring data from PC1 to PC4. The data contains two packets labeled as 1 and 2. As we can see, in Datagram mode, packet 1 chose to follow the SW1- SW4-SW3 path whereas Packet 2 chose the route path of SW1- SW5- SW3 and finally reached PC4. The packets can choose different path depending on the delay time and congestion on other paths in Datagram packet switching network.

5.5 Congestion & 5.6 Effects of congestion, congestion control

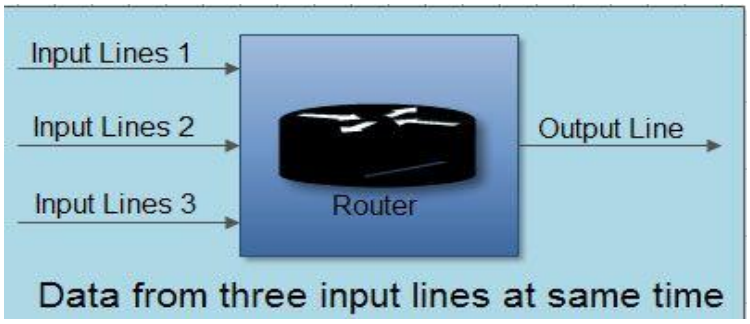
Congestion is an important issue that can arise in packet switched network. Congestion is a situation in Communication Networks in which too many packets are present in a part of the subnet, performance degrades. Congestion in a network may occur when the load on the network (i.e. the number of packets sent to the network) is greater than the capacity of the network (i.e. the number of packets a network can handle.). Network congestion occurs in case of traffic overloading.

In other words when too much traffic is offered, congestion sets in and performance degrades sharply



Causing of Congestion: The various causes of congestion in a subnet are:

• The input traffic rate exceeds the capacity of the output lines. If suddenly, a stream of packet start arriving on three or four input lines and all need the same output line. In this case, a queue will be built up. If there is insufficient memory to hold all the packets, the packet will be lost. Increasing the memory to unlimited size does not solve the problem. This is because, by the time packets reach front of the queue, they have already timed out (as they waited the queue). When timer goes off source transmits duplicate packet that are also added to the queue. Thus same packets are added again and again, increasing the load all the way to the destination.

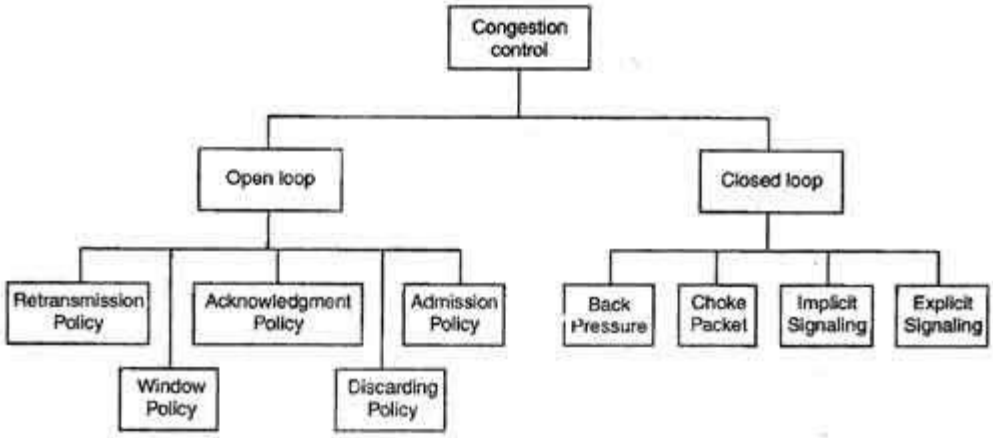


• The routers are too slow to perform bookkeeping tasks (queuing buffers, updating tables, etc.).
• The routers' buffer is too limited.
• Congestion in a subnet can occur if the processors are slow. Slow speed CPU at routers will perform the routine tasks such as queuing buffers, updating table etc slowly. As a result of this, queues are built up even though there is excess line capacity.
• Congestion is also caused by slow links. This problem will be solved when high speed links are used. But it is not always the case. Sometimes increase in link bandwidth can further deteriorate the congestion problem as higher speed links may make the network more unbalanced. Congestion can make itself worse. If a route!" does not have free buffers, it start ignoring/discarding the newly arriving packets. When these packets are discarded, the sender may

retransmit them after the timer goes off. Such packets are transmitted by the sender again and again until the source gets the acknowledgement of these packets. Therefore multiple transmissions of packets will force the congestion to take place at the sending end.

How to correct the Congestion Problem:

Congestion Control refers to techniques and mechanisms that can either prevent congestion, before it happens, or remove congestion, after it has happened. Congestion control mechanisms are divided into two categories, one category prevents the congestion from happening and the other category removes congestion after it has taken place.



Types of Congestion Control Methods

These two categories are:

1. Open loop

2. Closed loop

Open Loop Congestion Control

- In this method, policies are used to prevent the congestion before it happens.
- Congestion control is handled either by the source or by the destination.
- The various methods used for open loop congestion control are:

Retransmission Policy

- The sender retransmits a packet, if it feels that the packet it has sent is lost or corrupted.
- However retransmission in general may increase the congestion in the network. But we need to implement good retransmission policy to prevent congestion.
- The retransmission policy and the retransmission timers need to be designed to optimize efficiency and at the same time prevent the congestion.

Window Policy

- To implement window policy, selective reject window method is used for congestion control.
- Selective Reject method is preferred over Go-back-n window as in Go-back-n method, when timer for a packet times out, several packets are resent, although some may have arrived safely at the receiver. Thus, this duplication may make congestion worse.
- Selective reject method sends only the specific lost or damaged packets.

Acknowledgement Policy

- The acknowledgement policy imposed by the receiver may also affect congestion.
- If the receiver does not acknowledge every packet it receives it may slow down the sender and help prevent congestion.
- Acknowledgments also add to the traffic load on the network. Thus, by sending fewer acknowledgements we can reduce load on the network.
- To implement it, several approaches can be used:
    1. A receiver may send an acknowledgement only if it has a packet to be sent.

2. A receiver may send an acknowledgement when a timer expires.
3. A receiver may also decide to acknowledge only N packets at a time.

Discarding Policy

- A router may discard less sensitive packets when congestion is likely to happen.
- Such a discarding policy may prevent congestion and at the same time may not harm the integrity of the transmission.

Admission Policy

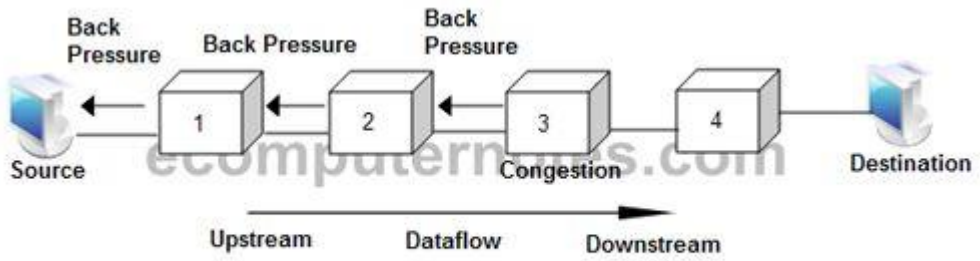
- An admission policy, which is a quality-of-service mechanism, can also prevent congestion in virtual circuit networks.
- Switches in a flow first check the resource requirement of a flow before admitting it to the network.

Closed Loop Congestion Control

- Closed loop congestion control mechanisms try to remove the congestion after it happens.
- The various methods used for closed loop congestion control are:

Backpressure

- Backpressure is a node-to-node congestion control that starts with a node and propagates, in the opposite direction of data flow.

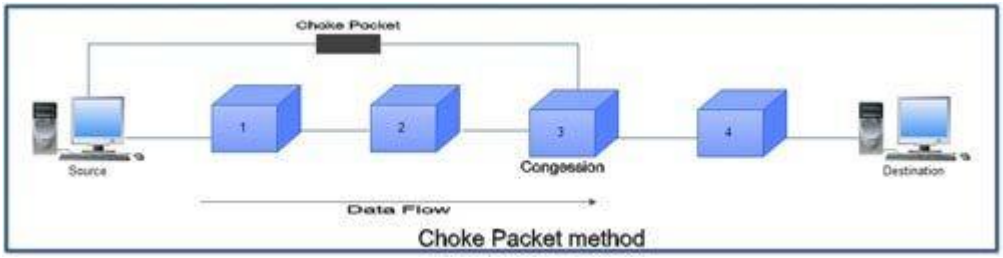


Backpressure Method

- The backpressure technique can be applied only to virtual circuit networks. In such virtual circuit each node knows the upstream node from which a data flow is coming.
- In this method of congestion control, the congested node stops receiving data from the immediate upstream node or nodes.
- This may cause the upstream node on nodes to become congested, and they, in turn, reject data from their upstream node or nodes.
- As shown in fig node 3 is congested and it stops receiving packets and informs its upstream node 2 to slow down. Node 2 in turns may be congested and informs node 1 to slow down. Now node 1 may create congestion and informs the source node to slow down. In this way the congestion is alleviated. Thus, the pressure on node 3 is moved backward to the source to remove the congestion.

Choke Packet

- In this method of congestion control, congested router or node sends a special type of packet called choke packet to the source to inform it about the congestion.
- Here, congested node does not inform its upstream node about the congestion as in backpressure method.
- In choke packet method, congested node sends a warning directly to the source station i.e. the intermediate nodes through which the packet has traveled are not warned.



Choke Packet method

Implicit Signaling

- In implicit signaling, there is no communication between the congested node or nodes and the source.
- The source guesses that there is congestion somewhere in the network when it does not receive any acknowledgment. Therefore the delay in receiving an acknowledgment is interpreted as congestion in the network.
- On sensing this congestion, the source slows down.
- This type of congestion control policy is used by TCP.

Explicit Signaling

- In this method, the congested nodes explicitly send a signal to the source or destination to inform about the congestion.
- Explicit signaling is different from the choke packet method. In choke packed method, a separate packet is used for this purpose whereas in explicit signaling method, the signal is included in the packets that carry data .
- Explicit signaling can occur in either the forward direction or the backward direction .
- In backward signaling, a bit is set in a packet moving in the direction opposite to the congestion. This bit warns the source about the congestion and informs the source to slow down.
- In forward signaling, a bit is set in a packet moving in the direction of congestion. This bit warns the destination about the congestion. The receiver in this case uses policies such as slowing down the acknowledgements to remove the congestion.

5.7 Traffic Management

Network traffic management deals with the process of monitoring and controlling the activities of network besides transforming the network into a managed resource by improving performance, efficiency, and security. It also helps to operate, administer, and maintain the network systems.

TYPES OF NETWORK TRAFFIC

Networks accommodate an increasingly complex set of data traffic. Identifying the type of traffic will help network administrators to facilitate the optimization of the network. Various types of traffic are:

Traffic Type	Example	Problem	Solution
Bursty Traffic	Downloads of FTP, graphic, video content	Consumes high bandwidth and Starves applications	Set constraint to limit access to bandwidth
Interactive Traffic	SSL transactions, IM, Telnet sessions	Susceptible to competition for bandwidth and results in poor response time	Prioritize over less essential traffic
Latency Sensitive Traffic	Streaming applications, Voice over IP, video conferencing	Susceptible to competition for bandwidth and results in poor response time	Set minimum and maximum bandwidth range based on priority
Non-Real Time Traffic	Email, batch processing applications	Consumes bandwidth during business hours	Schedule bandwidth during non-business hours

BANDWIDTH MANAGEMENT

The process of measuring and controlling the traffic on a network link so that overfilling the link can be avoided is called bandwidth management. It helps to keep the internet connection working fast and smooth. To manage bandwidth, traffic is measured, analyzed and the cause for heavy traffic is identified. Then network traffic control tool is used to avoid unwanted traffic and schedule bandwidth usage.
Due to various aspects, the communication links will not reach maximum capacity. The link utilization should be below the maximum theoretical capacity of the link so that fast responsiveness can be ensured by eliminating bottleneck queues at the link endpoints. Issues that limit the performance of a link are:

- The capacity of a connection is determined by TCP through flooding until packets are dropped.
- Higher latency is created due to queueing in routers.
- With the network reaching its capacity, TCP global synchronization results in waste of bandwidth
- Bursty traffic requires spare bandwidth
- No proper support for explicit congestion notification
- Queue management is controlled by Internet Service Providers

Traffic Measuring – Packet sniffers look at individual packets and help to track tricky problems. But they are voluminous and require the knowledge of network protocols. So traffic measuring tools are used to get broader view of the amount and type of traffic on a particular network. Some of the tools include:

- Caligare Flow for NetFlow monitoring and detecting network anomalies
- Exbander Precision for monitoring and analyzing
- FireBeast for bandwidth management and traffic shaping.
- PRTG for monitoring bandwidth usages.
- Sandvine Intelligent Network Solutions for measuring and managing network traffic

Traffic Shaping – An action on a set of packets to impose additional delay on those packets so that traffic on the network can be controlled for optimized and guaranteed performance is referred to as traffic shaping. It helps to control the volume of traffic sent in a specific period. Normally, it is applied at the edges of the network to control the entry of traffic, but sometimes, it is applied at the source or by an element in the network.
Rate Limiting – A method to control the rate of traffic sent or received on a network interface is referred to as rate limiting. When a traffic is less than or equal to the specified rate, it is sent and when it exceeds the specified rate, it is dropped or delayed. It is performed in the following ways.

- Policing – It discards excess packets and can be applied to any network protocol including IPv6.
- Queuing – It delays packets in transit and can be applied to any network protocol including IPv6.
- Congestion control – It manipulates the protocol's congestion mechanism and can be applied to TCP.

IPV6 TRAFFIC MANAGEMENT TECHNIQUES

- XRMON – XRMON captures network traffic levels and users across the network with a very little overhead to the infrastructure.
- sflow – sflow is a standard for network traffic monitoring and accounting as per the specification in RFC 3176. It gives complete visibility into the use of network and helps in optimizing the network resource.
- NetFlow – Netflow is an open network protocol developed by Cisco Systems to collect IP traffic information.
- IPFIX – Internet Protocol Flow Information eXport (IPFIX) is an IETF standard based on Netflow version 9. It exports
- Internet Protocol flow information from routers, probes, and other devices.

OTHER NETWORK TRAFFIC MANAGEMENT TOOLS

- **Traffic Prioritization Tool** – It improves the quality of service on the network besides blocking the competing services. It minimizes latency and allocates bandwidth on data networks.
- **Network Resource Management Tool** – It tracks, records, and monitors the current capacity of the system besides controlling the resource allocations.
- **Network Inventory Tool** – It helps in the audit of software and hardware elements of the network computers.
- **System Management Tool** – It is used for steady monitoring of server performance and availability.

5.8 Congestion Control in Packet Switching Network

Congestion has been considered as one of the basic important issue in packet switched network. Congestion control refers to the mechanisms and techniques to control the congestion and keep the load below the capacity. This paper provides an overview of category provided by congestion control .It also includes how TCP uses congestion control to avoid congestion or alleviate congestion in network. Computer networks have experienced an explosive growth over the past few years and with that growth have come severe congestion problems. This paper also concentrates on avoidance of congestion. This scheme allows a network to operate in the region of low delay and high throughput. In this paper, a survey on various mechanisms of congestion control and avoidance has been done.

Congestion in a network may occur if the load on the network-the number of packets sent to the network is greater than the capacity of the network-the number of packets a network can handle..Network congestion occurs when a link or node is carrying so much data that its quality of service deteriorates. Typical effects include queuing, packet loss or the blocking of new connections. Congestion control is a method used for monitoring the process of regulating the total amount of data entering the network .so as to keep traffic levels at an acceptable value. This is done in order to avoid the telecommunication network reaching what is termed congestive collapse. Modern networks use congestion control and network congestion avoidance techniques to try to avoid congestion collapse. These include: exponential back off in protocols such as 802.11's CSMA/CA and the original Ethernet, window reduction in TCP, and fair queuing in devices such as routers Congestion avoidance techniques monitor network traffic loads in an effort to anticipate and avoid congestion a common network bottleneck.

Congestion affects two vital parameters of the network performance, namely throughput and delay. The throughput can be defined as the percentage utilization of the network capacity. Throughput is affected as offered load increases. Initially throughput increases linearly with offered load, because utilization of the network increases. However, as the offered load increases beyond certain limit, say 60% of the capacity of the network, the throughput drops. If the offered load increases further, a point is reached when not a single packet is delivered to any destination, which is commonly known as deadlock situation. The ideal one corresponds to the situation when all the packet introduced are delivered to their destination up to the maximum capacity of the network. The second one corresponds to the situation when there is no congestion control. The third one is the case when some congestion control technique is used. This prevents the throughput collapse, but provides lesser throughput than the ideal condition due to overhead of the congestion control technique.

CONGESTION CONTROL MECHANISMS:

Congestion control mostly applies to packet-switching network. A wide variety of approaches have been proposed, however the "objective is to maintain the number of packets within the network below the level at which performance falls off dramatically.
Basic TCP congestion control schemes:

- Slow start
- Fast retransmission and Fast Recovery(Reno)

Slow Start: Slow start is part of the congestion control strategy used by TCP in conjunction with other algorithms to avoid sending more data than the network is capable of forwarding, that is, to avoid causing network congestion. The algorithm is specified by RFC 5681. Although the strategy is referred to as slow start, its congestion window growth is quite aggressive, more aggressive than the congestion avoidance phase.[1] Before slow start was introduced in TCP, the initial pre-congestion avoidance phase was even faster.

Fast retransmission and Fast Recovery(Reno): Fast retransmit is an enhancement to TCP that reduces the time a sender waits before retransmitting a lost segment. A TCP sender normally uses a simple timer to recognize lost segments. If an acknowledgement is not received for a particular segment within a specified time (a function of the estimated round-trip delay time), the sender will assume the segment was lost in the network, and will retransmit the segment.

Unit-6. LAN Technology

6.1. Topology and Transmission Media

Network 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. It is the schematic description of a network arrangement, connecting various nodes(sender and receiver) through lines of connection.

Types of Network Topology:

1. BUS Topology:
Bus topology is a network type in which every computer and network device is connected to single cable. When it has exactly two endpoints, then it is called **Linear Bus topology**.

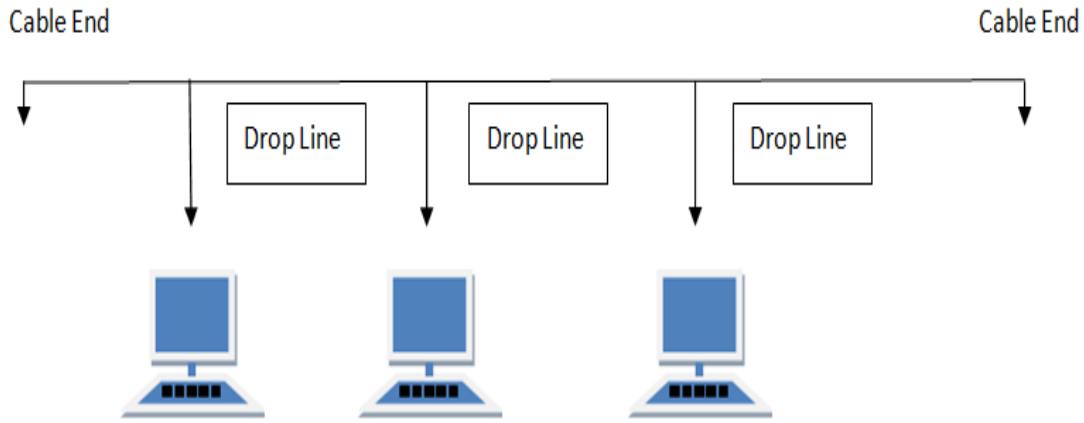


Fig.6.1(a) Bus Topology

Features of Bus Topology:

1. It transmits data only in one direction.
2. Every device is connected to a single cable

Advantages of Bus Topology

1. It is cost effective.
2. Cable required is least compared to other network topology.
3. Used in small networks.
4. It is easy to understand.
5. Easy to expand joining two cables together.

Disadvantages of Bus Topology

1. Cables fails then whole network fails.
2. If network traffic is heavy or nodes are more the performance of the network decreases.
3. Cable has a limited length.
4. It is slower than the ring topology.

2. RING Topology:

It is called ring topology because it forms a ring as each computer is connected to another computer, with the last one connected to the first. Exactly two neighbors for each device.

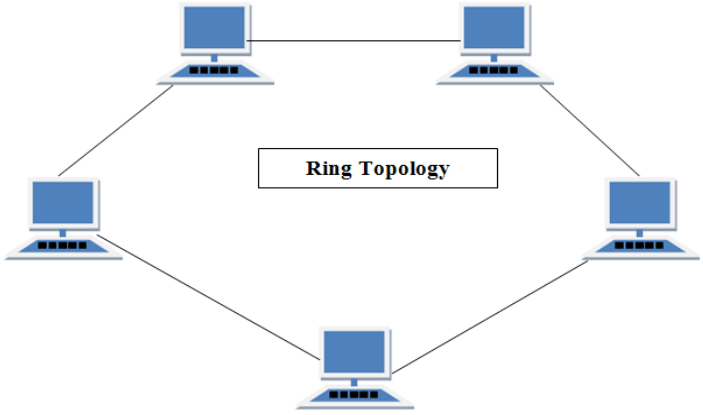


Fig 6.1(b) Ring Topology

Features of Ring Topology

1. A number of repeaters are used for Ring topology with large number of nodes, because if someone wants to send some data to the last node in the ring topology with 100 nodes, then the data will have to pass through 99 nodes to reach the 100th node. Hence to prevent data loss repeaters are used in the network.
2. The transmission is unidirectional, but it can be made bidirectional by having 2 connections between each Network Node, it is called **Dual Ring Topology**.
3. In Dual Ring Topology, two ring networks are formed, and data flow is in opposite direction in them. Also, if one ring fails, the second ring can act as a backup, to keep the network up.
4. Data is transferred in a sequential manner that is bit by bit. Data transmitted, has to pass through each node of the network, till the destination node.

Advantages of Ring Topology

1. Transmitting network is not affected by high traffic or by adding more nodes, as only the nodes having tokens can transmit data.
2. Cheap to install and expand

Disadvantages of Ring Topology

1. Troubleshooting is difficult in ring topology.
2. Adding or deleting the computers disturbs the network activity.
3. Failure of one computer disturbs the whole network.

3. STAR Topology:

In this type of topology all the computers are connected to a single hub through a cable. This hub is the central node and all others nodes are connected to the central node.

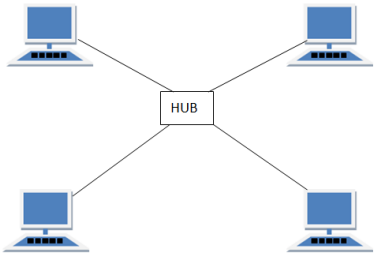


Fig 6.1(c) Star Topology

Features of Star Topology

1. Every node has its own dedicated connection to the hub.
2. Hub acts as a repeater for data flow.
3. Can be used with twisted pair, Optical Fiber or coaxial cable.

Advantages of Star Topology

1. Fast performance with few nodes and low network traffic.
2. Hub can be upgraded easily.
3. Easy to troubleshoot.
4. Easy to setup and modify.
5. Only that node is affected which has failed, rest of the nodes can work smoothly.

Disadvantages of Star Topology

1. Cost of installation is high.
2. Expensive to use.
3. If the hub fails then the whole network is stopped because all the nodes depend on the hub.
4. Performance is based on the hub that is it depends on its capacity

4. MESH Topology:

It is a point-to-point connection to other nodes or devices. All the network nodes are connected to each other. Mesh **has n(n-1)/2** physical channels to link n devices.

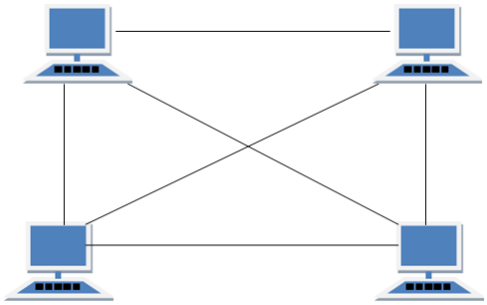


Fig 6.1(d) Mesh Topology

Types of Mesh Topology

1. Partial Mesh Topology : In this topology some of the systems are connected in the same fashion as mesh topology but some devices are only connected to two or three devices.
2. Full Mesh Topology : Each and every nodes or devices are connected to each other.

Features of Mesh Topology

1. Fully connected.
2. Robust.
3. Not flexible.

Advantages of Mesh Topology

1. Each connection can carry its own data load.
2. It is robust.
3. Fault is diagnosed easily.
4. Provides security and privacy.

Disadvantages of Mesh Topology

1. Installation and configuration is difficult.
2. Cabling cost is more.
3. Bulk wiring is required.

5. TREE Topology:

It has a root node and all other nodes are connected to it forming a hierarchy. It is also called hierarchical topology. It should at least have three levels to the hierarchy.

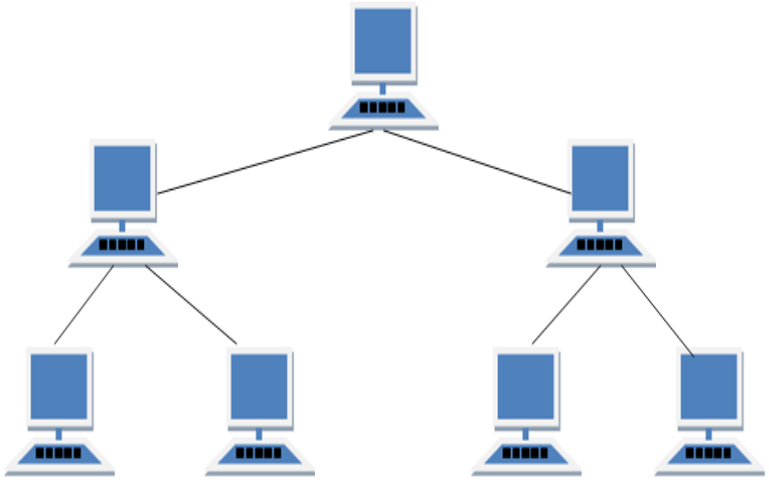


Fig 6.1(e) Tree Topology

Features of Tree Topology

1. Ideal if workstations are located in groups.
2. Used in Wide Area Network.

Advantages of Tree Topology

1. Extension of bus and star topologies.
2. Expansion of nodes is possible and easy.
3. Easily managed and maintained.
4. Error detection is easily done.

Disadvantages of Tree Topology

1. Heavily cabled.
2. Costly.
3. If more nodes are added maintenance is difficult.
4. Central hub fails, network fails.

6. HYBRID Topology:

It is two different types of topologies which is a mixture of two or more topologies. For example if in an office in one department ring topology is used and in another star topology is used, connecting these topologies will result in Hybrid Topology (ring topology and star topology).

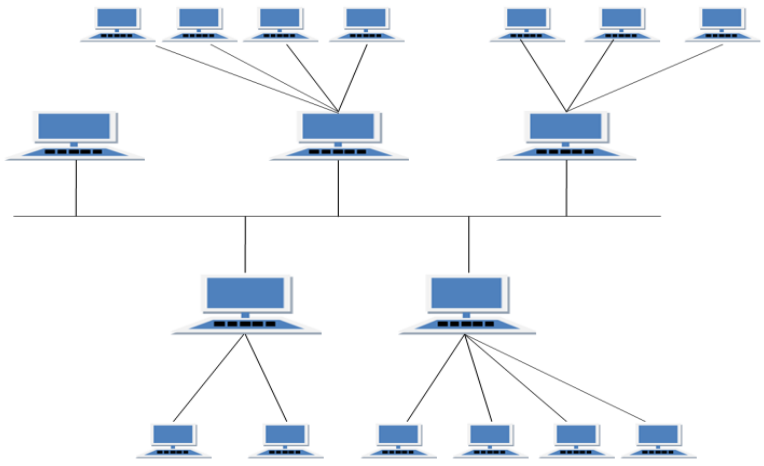


Fig 6.1(f) Hybrid Topology

Features of Hybrid Topology

1. It is a combination of two or topologies
2. Inherits the advantages and disadvantages of the topologies included

Advantages of Hybrid Topology

1. Reliable as Error detecting and trouble shooting is easy.
2. Effective.
3. Scalable as size can be increased easily.
4. Flexible.

Disadvantages of Hybrid Topology

1. Complex in design.
2. Costly.

6.2 LAN protocol architecture

A Local Area Network (LAN) is a private network that connects computers and devices within a limited area like a residence, an office, a building or a campus. On a small scale, LANs are used to connect personal computers to printers. However, LANs can also extend to a few kilometers when used by companies, where a large number of computers share a variety of resources like hardware (e.g. printers, scanners, audiovisual devices etc), software (e.g. application programs) and data.

The distinguishing features of LAN are

- Network size is limited to a small geographical area, presently to a few kilometers.
- Data transfer rate is generally high. They range from 100 Mbps to 1000 Mbps.
- In general, a LAN uses only one type of transmission medium, commonly category 5 coaxial cables.
- A LAN is distinguished from other networks by their topologies. The common topologies are bus, ring, mesh, and star.
- The number of computers connected to a LAN is usually restricted. In other words, LANs are limitedly scalable.
- IEEE 802 LAN/MAN standards are specific to the type of network (Ethernet, WLAN, WPAN, etc).
- IEEE 802.3 or Ethernet is the most common LAN. They use a wired medium in conjuncture with a switch or a hub. Originally, coaxial cables were used for communications. But now twisted pair cables and fiber optic cables are also used. Ethernet's speed has increased from 2.9 Mbps to 400 Gbps.

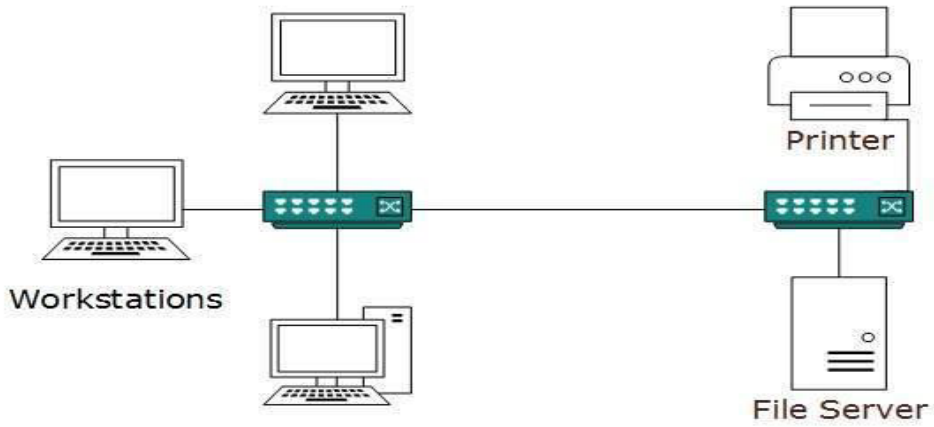


Fig 6.2(a) Sample LAN Network

Ethernet: Ethernet is most widely used LAN Technology, which is defined under IEEE standards 802.3. The reason behind its wide usability is Ethernet is easy to understand, implement, maintain and allows low-cost network implementation. Also, Ethernet offers flexibility in terms of topologies which are allowed.Ethernet generally uses Bus Topology. Ethernet operates in two layers of the OSI model, Physical Layer, and Data Link Layer. For Ethernet, the protocol data unit is Frame since we mainly deal with DLL. In order to handle collision, the Access control mechanism used in Ethernet is CSMA/CD. Manchester Encoding Technique is used in Ethernet.

A sample encoded data using Manchester Encoding Scheme is shown in below fig.

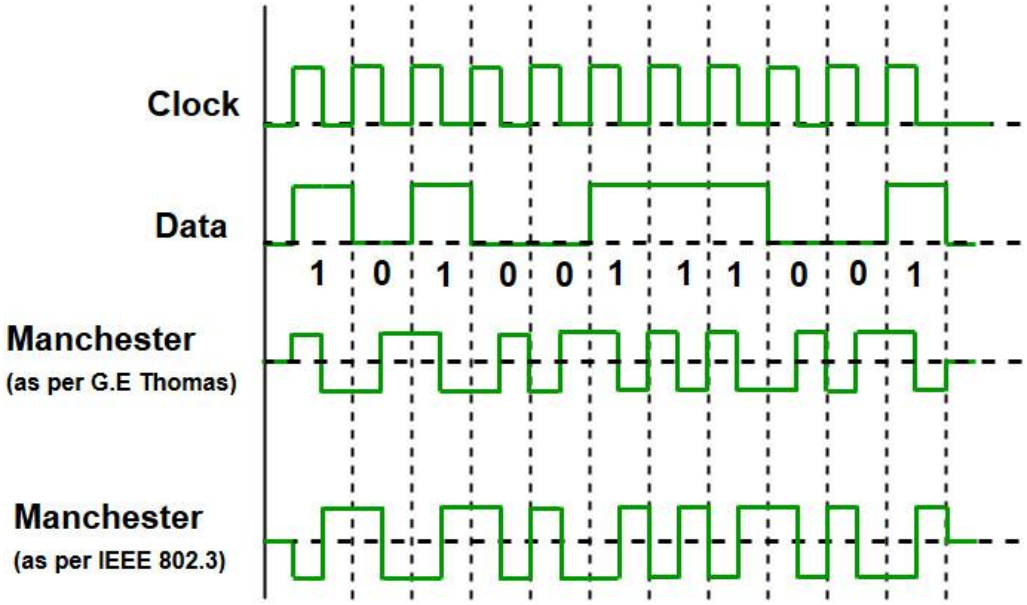


Fig 6.2(b) Manchester Coded Data Representation

Since we are talking about IEEE 802.3 standard Ethernet therefore, 0 is expressed by a high-to-low transition, a 1 by the low-to-high transition. In both Manchester Encoding and Differential Manchester, Encoding Baud rate is double of bit rate.

Baud rate = 2* Bit rate

Ethernet LANs consist of network nodes and interconnecting media or link. The network nodes can be of two types:
Data Terminal Equipment (DTE): Generally, DTEs are the end devices that convert the user information into signals or reconvert the received signals. DTEs devices are: personal computers, workstations, file servers or print servers also referred to as end stations. These devices are either the source or the destination of data frames. The data terminal equipment may be a single piece of equipment or multiple pieces of equipment that are interconnected and perform all the required functions to allow the user to communicate. A user can interact to DTE or DTE may be a user.

Data Communication Equipment (DCE):- DCEs are the intermediate network devices that receive and forward frames across the network. They may be either standalone devices such as repeaters, network switches, routers or maybe communications interface units such as interface cards and modems. The DCE performs functions such as signal conversion, coding and may be a part of the DTE or intermediate equipment.

Currently, these data rates are defined for operation over optical fibers and twisted-pair cables:

1. **Fast Ethernet:** Fast Ethernet refers to an Ethernet network that can transfer data at a rate of 100 Mbit/s.
2. **Gigabit Ethernet:** Gigabit Ethernet delivers a data rate of 1,000 Mbit/s (1 Gbit/s).

3. **10 Gigabit Ethernet:** 10 Gigabit Ethernet is the recent generation and delivers a data rate of 10 Gbit/s (10,000 Mbit/s). It is generally used for backbones in high-end applications requiring high data rates.

The IEEE has subdivided the data link layer into two sublayers:

- Logical Link Control (LLC) -(IEEE 802.2)It Places information in the frame to identify which network layer protocol is used for the frame.
- Media Access Control (MAC)-(IEEE 802.3, 802.11, or 802.15)

IEEE has also created several physical layer standards for different LAN protocols.

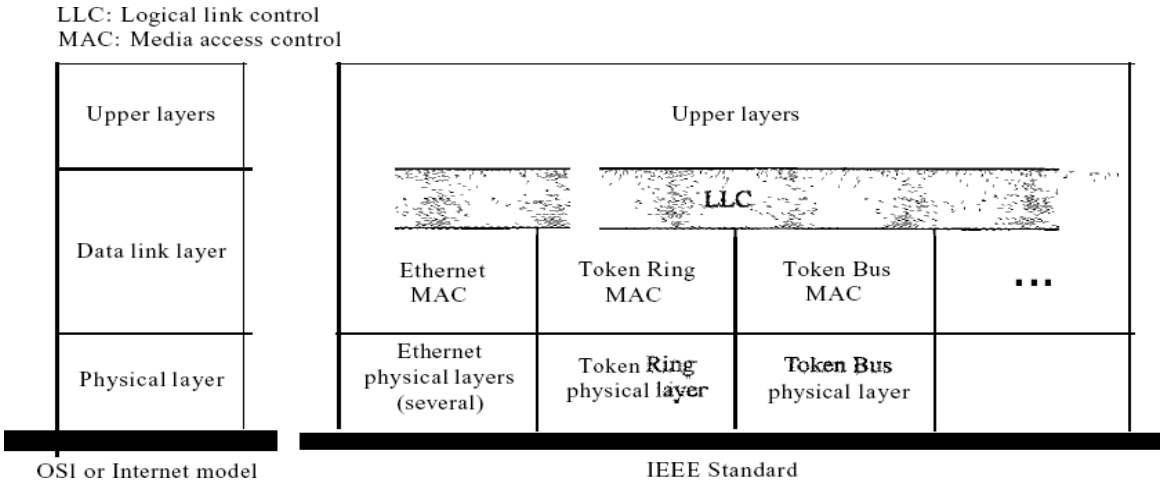


Fig 6.2(c) IEEE standardfor LANs

Data Link Layer

As we mentioned before, the data link layer in the IEEE standard is divided into two sublayers: LLC and MAC.

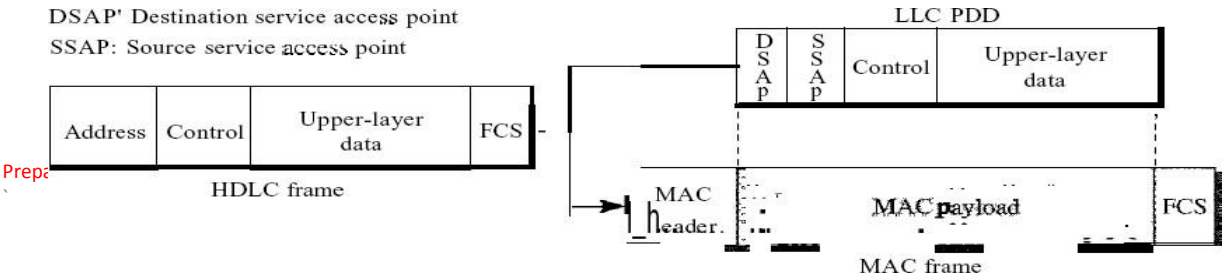
Logical Link Control (LLC)

We said that data link control handles framing, flow control, and error control. In IEEE Project 802, **flow control, error control,and part of the framing duties are collected into one sublayer called the logical link control**. Framing is handled in both the LLC sublayer and the MAC sublayer.

The LLC provides one single data link control protocol for all IEEE LANs. In this way, the LLC is different from the media access control sublayer, which provides different protocols for different LANs. A single LLC protocol can provide interconnectivity between different LANs because it makes the MAC sublayer transparent. Fig 6.2(c) shows one single LLC protocol serving several MAC protocols. Framing LLC defines a protocol data unit (PDU) that is somewhat similar to that of HDLC. The header contains a control field like the one in HDLC; this field is used for flow and error control. The two other header fields define the upper-layer protocol at the source and destination that uses LLC. These fields are called the destination service access point (DSAP) and the source service access point (SSAP). The other fields defined in a typical data link control protocol such as HDLC are moved to the MAC sublayer. In other words, a frame defined in HDLC is divided into a PDU at the LLC sublayer and a frame at the MAC sublayer, as shown in Fig 6.2(d).

Need for LLC The purpose of the LLC is to provide flow and error control for the upper-layer protocols that actually demand these services. For example, if a LAN or several LANs are used in an isolated system, LLC may be needed to provide flow and error control for the application layer protocols. However, most upper-layer protocols such as IP, do not use the services of LLC.

Fig 6.2(d)



6.3. Medium Access control

The medium access control (MAC) is a sublayer of the data link layer of the open system interconnections (OSI) reference model for data transmission. It is responsible for flow control and multiplexing for transmission medium. It controls the transmission of data packets via remotely shared channels. It sends data over the network interface card.

MAC Layer in the OSI Model

The Open System Interconnections (OSI) model is a layered networking framework that conceptualizes how communications should be done between heterogeneous systems. The data link layer is the second lowest layer. It is divided into two sublayers:

- The logical link control (LLC) sublayer
- The medium access control (MAC) sublayer

Functions of MAC Layer

It provides an abstraction of the physical layer to the LLC and upper layers of the OSI network.

- It is responsible for encapsulating frames so that they are suitable for transmission via the physical medium.
- It resolves the addressing of source station as well as the destination station, or groups of destination stations.
- It performs multiple access resolutions when more than one data frame is to be transmitted. It determines the channel access methods for transmission.
- It also performs collision resolution and initiating retransmission in case of collisions.
- It generates the frame check sequences and thus contributes to protection against transmission errors.

Channel Allocation Problem:

In broadcast networks, single channel is shared by several stations. This channel can be allocated to only one transmitting user at a time. There are **two** different methods of channel allocations:

- Static Channel Allocation
- Dynamic Channel Allocation

Static Channel Allocations: In this method, a single channel is divided among various users either on the basis of frequency or on the basis of time. It either uses FDM (Frequency Division Multiplexing) or TDM (Time Division Multiplexing). In FDM, fixed frequency is assigned to each user, whereas, in TDM, fixed time slot is assigned to each user.

Dynamic Channel Allocation: In this method, no user is assigned fixed frequency or fixed time slot. All users are dynamically assigned frequency or time slot, depending upon the requirements of the user.

Multiple Access Protocols:

Many protocols have been defined to handle the access to shared link. These protocols are organized in **three** different groups.:

1. Random Access Protocols
2. Controlled Access Protocols
3. Channelization Protocols

Refer below figure for further classification.

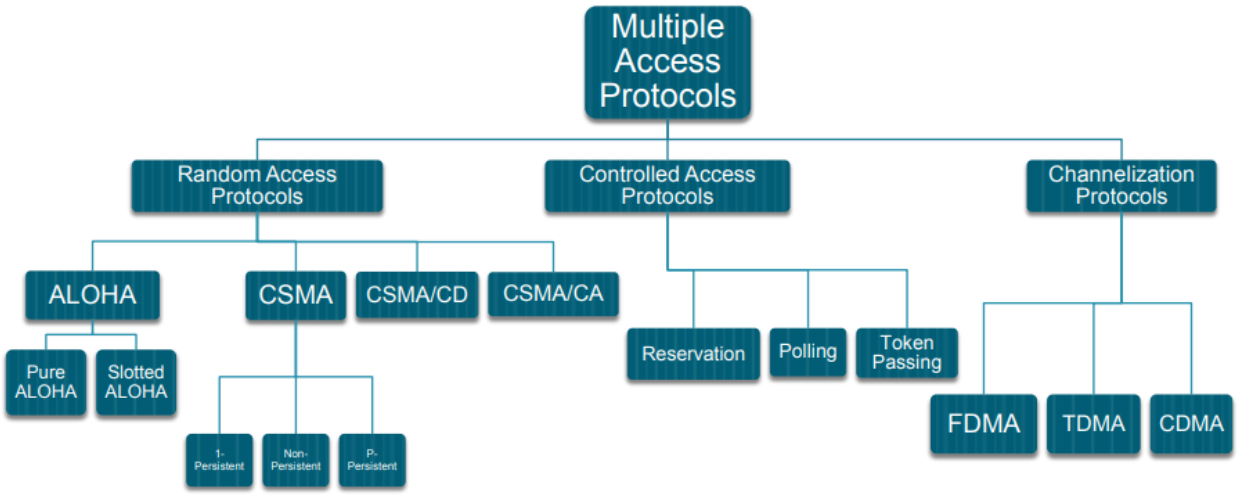


Fig 6.3(a) Classification of Multiple Access Protocols

Random Access Protocols: It is also called Contention Method. In this method, there is no control station. Any station can send the data. The station can make a decision on whether or not to send data. This decision depends on the state of the channel, i.e. channel is busy or idle. There is no scheduled time for a stations to transmit. They can transmit in random order. There is no rule that decides which station should send next. If two stations transmit at the same time, there is collision and the frames are lost. The various random access methods are:

- ALOHA
- CSMA (Carrier Sense Multiple Access)
- CSMA/CD (Carrier Sense Multiple Access with Collision Detection)
- CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance)

ALOHA: ALOHA was developed at University of Hawaii in early 1970s by Norman Abramson. It was used for ground based radio broadcasting. In this method, stations share a common channel. When two stations transmit simultaneously, collision occurs and frames are lost. There are two different versions of ALOHA:

- Pure ALOHA
- Slotted ALOHA

Pure ALOHA: In pure ALOHA, stations transmit frames whenever they have data to send. When two stations transmit simultaneously, there is collision and frames are lost. In pure ALOHA, whenever any station transmits a frame, it expects an acknowledgement from the receiver. If acknowledgement is not received within specified time, the station assumes that the frame has been lost. If the frame is lost, station waits for a random amount of time and sends it again. This waiting time must be random, otherwise, same frames will collide again and again. Whenever two frames try to occupy the channel at the same time, there will be collision and both the frames will be lost. If first bit of a new frame overlaps with the last bit of a frame almost finished, both frames will be lost and both will have to be retransmitted.

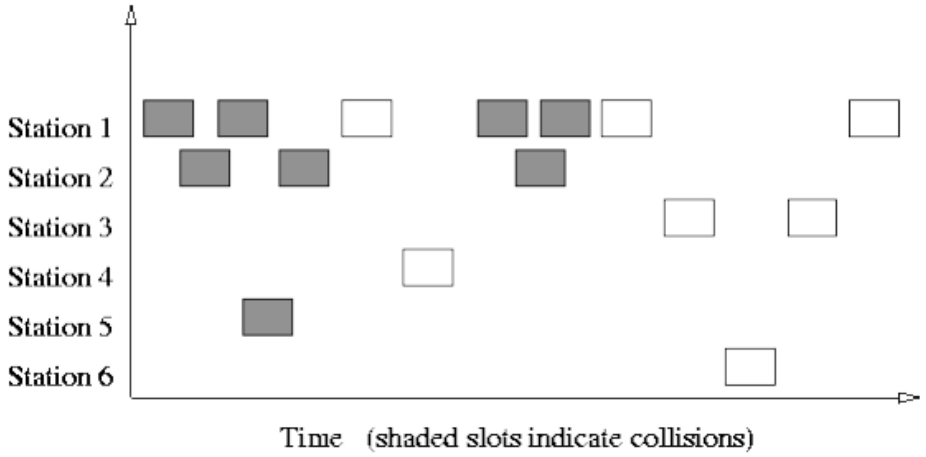


Fig 6.3(b)Pure Aloha based transmission

Slotted ALOHA: Slotted ALOHA was invented to improve the efficiency of pure ALOHA. In slotted ALOHA, time of the channel is divided into intervals called slots. The station can send a frame only at the beginning of the slot and only one frame is sent in each slot. If any station is not able to place the frame onto the channel at the beginning of the slot, it has to wait until the next time slot. There is still a possibility of collision if two stations try to send at the beginning of the same time slot.

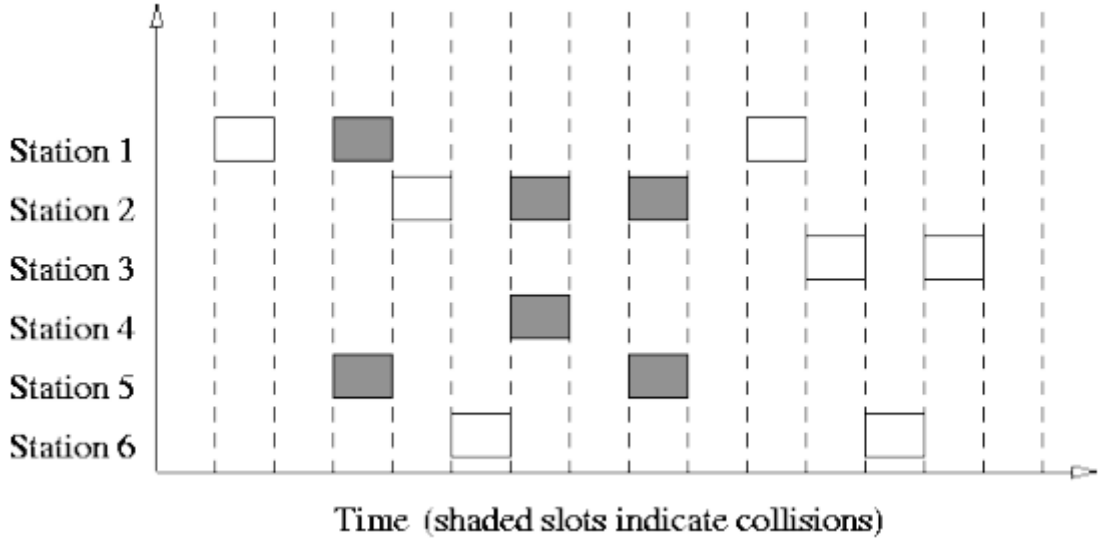
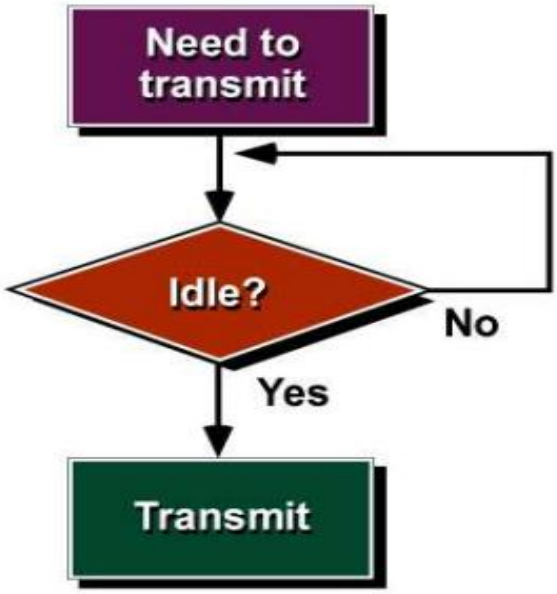


Fig 6.3(c)Slotted Aloha based transmission

Carrier Sense Multiple Access (CSMA): CSMA was developed to overcome the problems of ALOHA i.e. to minimize the chances of collision. CSMA is based on the principle of "carrier sense". The station sense the carrier or channel before transmitting a frame. It means the station checks whether the channel is idle or busy. The chances of collision reduces to a great extent if a station checks the channel before trying to use it.



The chances of collision still exist because of propagation delay. The frame transmitted by one station takes some time to reach the other station. In the meantime, other station may sense the channel to be idle and transmit its frames. This results in the collision.

Media Access Control (MAC) Address :

MAC Addresses are unique **48-bits** hardware number of a computer, which is embedded into network card (known as **Network Interface Card**) during the time of manufacturing. MAC Address is also known as **Physical Address** of a network device. MAC address is used by Media Access Control (MAC) sublayer of Data-Link Layer. MAC Address is word wide unique, since millions of network devices exists and we need to uniquely identify each. MAC Address is a 12-digit hexadecimal number (6-Byte binary number), which is mostly represented by Colon-Hexadecimal notation. First 6-digits (say 00:40:96) of MAC Address identifies the manufacturer, called as OUI (**Organizational Unique Identifier**). IEEE Registration Authority Committee assign these MAC prefixes to its registered vendors.

Eg:00-0a-83-b1-c0-8e 9(or) 00:0a:83:bq:c0:8e

6.4 Bridges, Hub, Switch

The operation of devices is related to different layers as illustrated below:

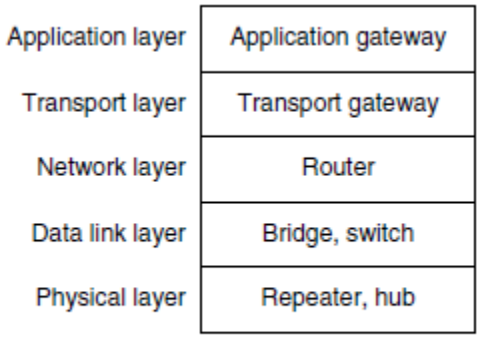


Fig: Devices in different layers

Repeaters:

- The repeater operates in the physical layer.
- These are analog devices that work with signals on the cables to which they are connected.
- A signal appearing on one cable is regenerated and put out on another cable. Hence it extends the physical length of LAN.
- Repeaters do not understand frames, packets or headers. They understand the symbols that encode bit as volts. Classic Ethernet, for example, was designed to allow four repeaters that would boost the signal to extend the maximum cable length from 500 meters to 2500 meters.

Hub:

- A hub has a number of input lines that it joins electrically. Active hub and passive hub are two types of hubs.
- Frames arriving on any of the lines are sent out on all the others. It is broadcast device. If two frames arrive at the same time, they will collide, just as on a coaxial cable.
- All the lines coming into a hub must operate at the same speed. Hubs differ from repeaters in that they do not boost the incoming signals and are designed for multiple input lines, but the differences are slight.
- Like repeaters, hubs are physical layer devices that do not examine the link layer addresses or use them in any way. It is not an intelligent device.

Bridge:

- A bridge connects two or more LANs. It operates **at data link layer**.
- Like a hub, a modern bridge has multiple ports, usually enough for 4 to 48 input lines of a certain type. Unlike in a hub, each port is isolated to be its own collision domain.
- When a frame arrives, the bridge extracts the destination address (for Ethernet, it is 48 bit) from the frame header and looks it up in a table to see where to send the frame.
- The bridge only outputs the frame on the port where it is needed and can forward multiple frames at the same time. Filtering, forwarding and blocking of frames are functions of bridges.
- Bridges offer much better performance than hubs and the isolation between bridge ports also means that the input lines may run at different speeds, possibly even with different network types. A common example is a bridge with ports that connect to 10-, 100-, and 1000-Mbps Ethernet.
- Buffering within the bridge is needed to accept a frame on one port and transmit the frame out on a different port.
- Bridges were originally intended to be able to join different kinds of LANs, for example, an Ethernet and a Token Ring LAN. However, this never worked well because of differences between the LANs such as frame formats, maximum frame lengths, security and Quality of service.

Switch:

- Switches are modern bridges by another name. It acts as multiport bridge to connect devices or segments in a LAN. It operates at **data link layer**.
- It is point to point device.
- It is an intelligent device. It uses switching table to find the correct destination.
- Switches are of two types:
    i. Store-and-forward switch: It stores the frame in the input buffer until the whole packet has arrived.
    ii. Cut-through switch: It forwards the packet to the output buffer as soon as the destination address is received.
- Also there are layer 2 (bridge) and layer 3 switches (kind of router). It is sophisticated and expensive device.

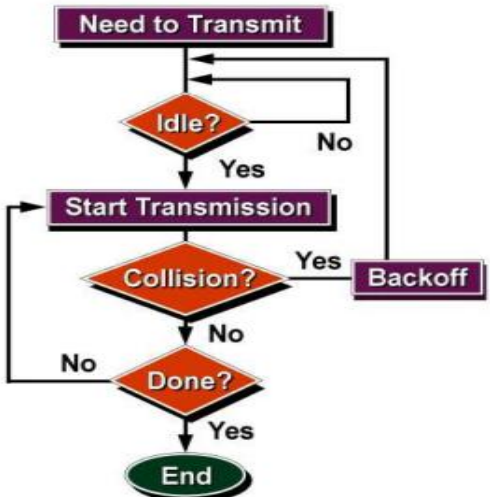
Router:

- Routers are devices that connect two or more networks. It operates at **network layer.**
- They consist of a combination of hardware and software.
- The hardware can be a network server, a separate computer or a special device. The hardware includes the physical interfaces to the various networks in the internetwork.
- These interfaces can be Token Ring, Ethernet, T1, Frame Relay, ATM or any other technology.
- The software in a router are the operating system and the routing protocol. Management software can also be used.
- Routers use logical and physical addressing to connect two or more logically separate networks.
- The network address allows routers to calculate the optimal path to a workstation or computer.
- The two methods of route discovery are Distance vector routing and Link state routing.

6.5 Ethernet (CSMA/CD), Fiber Channel

Ethernet(CSMA/CD):

It's CSMA with Collision Detection. In this protocol, the station senses the channel before transmitting the frame. If the channel is busy, the station waits. Additional feature in CSMA/CD is that the stations can detect collisions. The stations abort their transmission as soon as they detect collision. This feature is not present in CSMA. The stations continue to transmit even though they find that collision has occurred.



In CSMA/CD, the station that sends its data on the channel continues to sense the channel even after data transmission. If collision is detected, the station aborts its transmission and waits for a random amount of time & sends its data again. As soon as a collision is detected, the transmitting station release a jam signal. Jam signal alerts other stations. Stations are not supposed to transmit immediately after the collision has occurred.

Fiber Channel:

Fiber Channel is the general name of an integrated set of standards being developed by the American National Standards Institute (ANSI). It is designed to significantly improve the speed at which data is transmitted between workstations, mainframes, supercomputers, storage devices and displays, while providing one standard for networking, storage and data transfer. In general Fiber Channel provides speed, reliability, distance, and connectivity; all at a low cost. It guarantees bandwidth and delivery while being able to handle any protocol or topology.

Fiber Channel is a high-speed networking technology primarily used for transmitting data among data centers, computer servers, switches and storage at data rates of up to 128 Gbps. It was developed to overcome the shortcomings of the Small Computer System Interface (SCSI) and High-Performance Parallel Interface (HIPPI) by filling the need for a reliable and scalable high-throughput and low-latency protocol and interface. Fiber Channel is especially suited for connecting servers to shared storage devices and interconnecting storage controllers and drives. The Fiber Channel interface was created for storage area networks (SANs).

Fiber Channel devices can be as far as 10 kilometers (approximately six miles) apart if multimodal optical fiber is used as the physical medium. Optical fiber is not required for shorter distances. Fiber Channel also works using coaxial cable and ordinary telephone twisted pair. When using copper cabling, however, it is recommended distances do not exceed 100 feet.

Fiber Channel offers point-to-point, switched and loop interfaces to deliver lossless, in-order, raw block data. Because Fiber Channel today is many times faster than SCSI, it has replaced that technology as the transmission interface between servers and clustered storage devices. Fiber Channel networks can transport SCSI commands and information units using the Fiber Channel Protocol (FCP), however. It is designed to not just interoperate with SCSI but with the Internet Protocol (IP) and other protocols.

6.6 Wireless LAN Technology

Wireless LAN stands for **Wireless Local Area Network**. It is also called LAWN (**Local Area Wireless Network**). WLAN is one in which a mobile user can connect to a Local Area Network (LAN) through a wireless connection.

The IEEE 802.11 group of standards defines the technologies for wireless LANs. For path sharing, 802.11 standard uses the Ethernet protocol and CSMA/CA (carrier sense multiple access with collision avoidance). It also uses an encryption method i.e. wired equivalent privacy algorithm.

Wireless LANs provide high speed data communication in small areas such as building or an office. WLANs allow users to move around in a confined area while they are still connected to the network.

In some instance wireless LAN technology is used to save costs and avoid laying cable, while in other cases, it is the only option for providing high-speed internet access to the public. Whatever the reason, wireless solutions are popping up everywhere.

Examples of WLANs that are available today are NCR's waveLAN and Motorola's ALTAIR.

Advantages of WLANs

- **Flexibility:** Within radio coverage, nodes can communicate without further restriction. Radio waves can penetrate walls, senders and receivers can be placed anywhere (also non-visible, e.g., within devices, in walls etc.).
- **Planning:** Only wireless ad-hoc networks allow for communication without previous planning, any wired network needs wiring plans.
- **Design:** Wireless networks allow for the design of independent, small devices which can for example be put into a pocket. Cables not only restrict users but also designers of small notepads, PDAs, etc.
- **Robustness:** Wireless networks can handle disasters, e.g., earthquakes, flood etc. whereas, networks requiring a wired infrastructure will usually break down completely in disasters.
- **Cost:** The cost of installing and maintaining a wireless LAN is on average lower than the cost of installing and maintaining a traditional wired LAN, for two reasons. First, after providing wireless access to the wireless network via an access point for the first user, adding additional users to a network will not increase the cost. And second, wireless LAN eliminates the direct costs of cabling and the labor associated with installing and repairing it.
- **Ease of Use:** Wireless LAN is easy to use and the users need very little new information to take advantage of WLANs.

Disadvantages of WLANs

- **Quality of Services:** Quality of wireless LAN is typically lower than wired networks. The main reason for this is the lower bandwidth due to limitations is radio transmission, higher error rates due to interference and higher delay/delay variation due to extensive error correction and detection mechanisms.
- **Proprietary Solutions:** Due to slow standardization procedures, many companies have come up with proprietary solutions offering standardization functionality plus many enhanced features. Most components today adhere to the basic standards IEEE 802.11a or 802.11b.
- **Restrictions:** Several govt. and non-govt. institutions world-wide regulate the operation and restrict frequencies to minimize interference.
- **Global operation:** Wireless LAN products are sold in all countries so, national and international frequency regulations have to be considered.

- **Low Power:** Devices communicating via a wireless LAN are typically power consuming, also wireless devices running on battery power. Whereas the LAN design should take this into account and implement special power saving modes and power management functions.
- **License free operation:** LAN operators don't want to apply for a special license to be able to use the product. The equipment must operate in a license free band, such as the 2.4 GHz ISM band.
- **Robust transmission technology:** If wireless LAN uses radio transmission, many other electrical devices can interfere with them (such as vacuum cleaner, train engines, hair dryers, etc.).Wireless LAN transceivers cannot be adjusted for perfect transmission is a standard office or production environment.

UNIT-7. TCP/IP

7.1 TCP/IP Protocol Suite

The TCP/IP Reference Model:

The TCP/IP reference model was developed prior to OSI model. The major design goals of this model were,

- To connect multiple networks together so that they appear as a single network.
- To survive after partial subnet hardware failures.
- To provide a flexible architecture.

Unlike OSI reference model, TCP/IP reference model has only 4 layers. They are,

- Network Access Layer
- Internet Layer
- Transport Layer
- Application Layer

Network Access Layer:

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 to host and network to network.

Internet Layer:

This layer, called the 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 they 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 present in the Internet.

The internet layer defines an official packet format and protocol called IP (Internet Protocol). 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. Fig. shows this correspondence.

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 peer entities on the source and destination hosts to carry on a conversation, just as in the OSI transport layer. Two end-to-end transport protocols have been defined here. The first one, 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 stream into discrete messages and passes each one on to the internet layer. At the destination, the receiving TCP process reassembles the received messages into the output stream. TCP also handles flow control to make sure a fast sender cannot swamp a slow receiver with more messages than it can handle.

The second protocol in this layer, 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 is also widely used for one-shot, client-server-type request-reply queries and applications in which prompt delivery is more important than accurate delivery, such as transmitting speech or video. The relation of IP, TCP, and UDP is shown in Fig.2. Since the model was developed, IP has been implemented on many other networks.

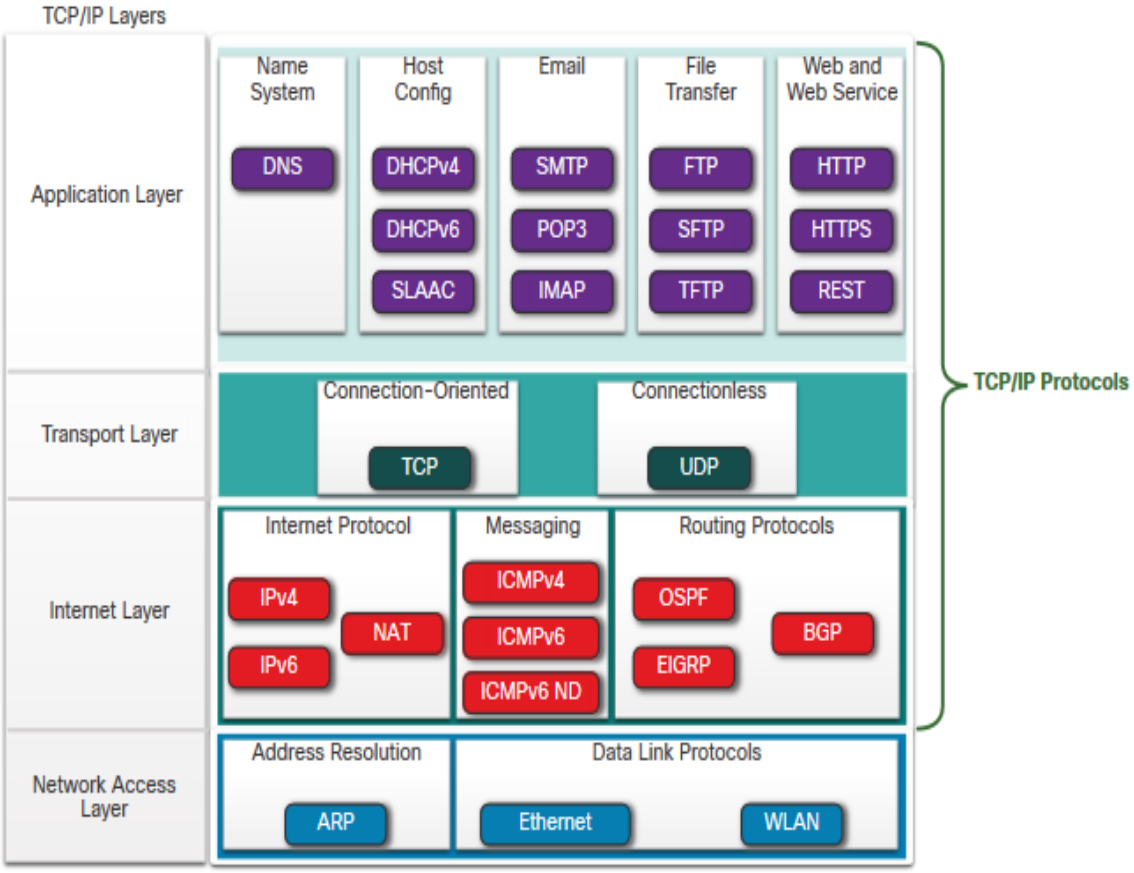


Fig. 7.1(a) The TCP/IP Protocol Suite

The Application Layer:

The TCP/IP model does not have session or presentation layers. On top of the transport layer is the application layer. It contains all the higher-level protocols. The early ones included virtual terminal (TELNET), file transfer (FTP), and electronic mail (SMTP), as shown in Fig.6.2. 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 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.

7.2 Basic Protocol Functions

1. Network Access Layer :
This layer corresponds to the combination of Data Link Layer and Physical Layer of the OSI model. It looks out for hardware addressing and the protocols present in this layer allows for the physical transmission of data.
We just talked about ARP being a protocol of Internet layer, but there is a conflict about declaring it as a protocol of Internet Layer or Network access layer. It is described as residing in layer 3, being encapsulated by layer 2 protocols.

2. Internet Layer :
This layer parallels the functions of OSI's Network layer. It defines the protocols which are responsible for logical transmission of data over the entire network. The main protocols residing at this layer are :

1. **IP –** stands for Internet Protocol and it is responsible for delivering packets from the source host to the destination host by looking at the IP addresses in the packet headers. IP has 2 versions:
   IPv4 and IPv6. IPv4 is the one that most of the websites are using currently. But IPv6 is growing as the number of IPv4 addresses are limited in number when compared to the number of users.
2. **ICMP –** stands for Internet Control Message Protocol. It is encapsulated within IP datagrams and is responsible for providing hosts with information about network problems.
3. **ARP –** stands for Address Resolution Protocol. Its job is to find the hardware address of a host from a known IP address. ARP has several types: Reverse ARP, Proxy ARP, Gratuitous ARP and Inverse ARP.

3. Host-to-Host Layer(Transport Layer):
This layer is analogous to the transport layer of the OSI model. It is responsible for end-to-end communication and error-free delivery of data. It shields the upper-layer applications from the complexities of data. The two main protocols present in this layer are :

1. **Transmission Control Protocol (TCP) –** It's a connection oriented protocol hence it is known to provide reliable and error-free communication between end systems. It performs sequencing and segmentation of data. It also has acknowledgment feature and controls the flow of the data through flow control mechanism. It is a very effective protocol but has a lot of overhead due to such features. Increased overhead leads to increased cost.

2. **User Datagram Protocol (UDP) –** On the other hand does not provide any such features as it is a connectionless protocol. It is the go-to protocol if your application does not require reliable transport as it is very cost-effective. It's faster than TCP.

4. Application Layer
This layer performs the functions of top three layers of the OSI model: Application, Presentation and Session Layer. It is responsible for node-to-node communication and controls user-interface specifications. Some of the protocols present in this layer are: HTTP, HTTPS, FTP, TFTP, Telnet, SSH, SMTP, SNMP, NTP, DNS, DHCP, NFS, X Window, LPD. Have a look at Protocols in Application Layer for some information about these protocols. Protocols other than those present in the linked article are :

1. **HTTP and HTTPS :** HTTP stands for Hypertext transfer protocol. It is used by the World Wide Web to manage communications between web browsers and servers. HTTPS stands for HTTP-Secure. It is a combination of HTTP with SSL(Secure Socket Layer). It is efficient in cases where the browser need to fill out forms, sign in, authenticate and carry out bank transactions.

2. **DNS:** DNS stands for Domain Name Server protocol. It is a host name to IP address translation service. DNS is a distributed database implemented in a hierarchy of name servers. It is an application layer protocol for message exchange between clients and servers.

3. **FTP:** FTP stands for file transfer protocol. It is the protocol that actually lets us transfer files.It can facilitate this between any two machines using it. But FTP is not just a protocol but it is also a program.FTP promotes sharing of files via remote computers with reliable and efficient data transfer. Port number for FTP is 20 for data and 21 for control.

4. **TFTP:** The Trivial File Transfer Protocol (TFTP) is the stripped-down, stock version of FTP, but it's the protocol of choice if you know exactly what you want and where to find it. It's a technology for transferring files between network devices and is a simplified version of FTP.

5. **SMTP:** It stands for Simple Mail Transfer Protocol. It is a part of the TCP/IP protocol. Using a process called "store and forward," SMTP moves your email on and across networks. It works closely with something called the Mail Transfer Agent (MTA) to send your communication to the right computer and email inbox. Port number for SMTP is 25.

6. **TELNET:** Telnet stands for the **TEL**ecomunications **NET**work. It helps in terminal emulation. It allows Telnet client to access the resources of the Telnet server. It is used for managing the files on the internet. It is used for initial set up of devices like switches. The telnet command is a command that uses the Telnet protocol to communicate with a remote device or system. Port number of telnet is 23.
7. **SSH –** SSH stands for Secure Shell. It is a terminal emulations software similar to Telnet. The reason SSH is more preferred is because of its ability to maintain the encrypted connection. It sets up a secure session over a TCP/IP connection.

8. **NTP –** NTP stands for Network Time Protocol. It is used to synchronize the clocks on our computer to one standard time source. It is very useful in situations like bank transactions. Assume the following situation without the presence of NTP. Suppose you carry out a transaction, where your computer reads the time at 2:30 PM while the server records it at 2:28 PM. The server can crash very badly if it's out of sync.

7.3 Principles of Internetworking

Internetworking is combined of 2 words, inter and networking which implies an association between totally different nodes or segments. This connection area unit is established through intercessor devices akin to routers or gateway. The first term for associate degree internetwork was catenet. This interconnection is often among or between public, private, commercial, industrial, or governmental networks. Thus, associate degree internetwork could be an assortment of individual networks, connected by intermediate networking devices, that functions as one giant network. Internetworking refers to the trade, products, and procedures that meet the challenge of making and administering internetworks.

To enable communication, every individual network node or phase is designed with similar protocol or communication logic, that is Transfer Control Protocol (TCP) or Internet Protocol (IP). Once a network communicates with another network having constant communication procedures, it's called Internetworking. Internetworking was designed to resolve the matter of delivering a packet of information through many links.

There a minute difference between extending the network and Internetworking. Merely exploitation of either a switch or a hub to attach 2 local area networks is an extension of LAN whereas connecting them via the router is associate degree example of Internetworking. Internetworking is enforced in Layer three (Network Layer) of OSI-ISO model. The foremost notable example of internetworking is that the Internet.There are chiefly 3 unit of Internetworking:

1. Extranet
2. Intranet
3. Internet

Intranets and extranets might or might not have connections to the net. If there is a connection to the net, the computer network or extranet area unit is usually shielded from being accessed from the net if it is not authorized. The net isn't thought-about to be a section of the computer network or extranet, though it should function a portal for access to parts of associate degree extranet.

1. **Extranet:** It's a network of the internetwork that's restricted in scope to one organization or entity however that additionally has restricted connections to the networks of one or a lot of different sometimes, however not essential. It's very lowest level of Internetworking, usually enforced in an exceedingly personal area. Associate degree extranet may additionally be classified as a Man, WAN, or different form of network however it cannot encompass one local area network i.e. it should have a minimum of one reference to associate degree external network.
2. **Intranet:** This associate degree computer network could be a set of interconnected networks, which exploits the Internet Protocol and uses IP-based tools akin to web browsers and FTP tools, that's underneath the management of one body entity. That body entity closes the computer network to the remainder of the planet and permits solely specific users. Most typically, this network is the internal network of a corporation or different enterprise. An outsized computer network can usually have its own internet server to supply users with browseable data.
3. **Internet:** A selected Internetworking, consisting of a worldwide interconnection of governmental, academic, public, and personal networks based mostly upon the Advanced analysis comes Agency Network (ARPANET) developed by ARPA of the U.S. Department of Defense additionally home to the World Wide Web (WWW) and cited as the 'Internet' to differentiate from all different generic Internetworks. Participants within the web, or their service suppliers, use IP Addresses obtained from address registries that management assignments.

Internetworking has evolved as an answer to a few key problems: isolated LANs, duplication of resources, and an absence of network management. Isolated LANs created transmission problem between totally different offices or departments. Duplication of resources meant that constant hardware and code had to be provided to every workplace or department, as did a separate support employee. This lack of network management meant that no centralized methodology of managing and troubleshooting networks existed.

One more form of interconnection of networks usually happens among enterprises at the Link Layer of the networking model, i.e. at the hardware-centric layer below the amount of the TCP/IP logical interfaces. Such interconnection is accomplished through network bridges and network switches. This can be typically incorrectly termed internetworking, however, the ensuing system is just a bigger, single subnetwork, and no internetworking protocol, akin to web Protocol, is needed to traverse these devices.

However, one electronic network is also reborn into associate degree internetwork by dividing the network into phases and logically dividing the segment traffic with routers. The Internet Protocol is meant to supply an associate degree unreliable packet service across the network. The design avoids intermediate network components maintaining any state of the network. Instead, this task is allotted to the endpoints of every communication session. To transfer information correctly, applications should utilize associate degree applicable Transport Layer protocol, akin to Transmission management Protocol (TCP), that provides a reliable stream. Some applications use a less complicated, connection-less transport protocol, User Datagram Protocol (UDP), for tasks that don't need reliable delivery of information or that need period of time service, akin to video streaming or voice chat.

Internetwork Addressing :

Internetwork addresses establish devices severally or as members of a bunch. Addressing schemes differ based on the protocol family and therefore the OSI layer. Three kinds of internetwork addresses area unit ordinarily used: data-link layer addresses, Media Access control (MAC) addresses, and network-layer addresses.

1. **Data Link Layer addresses:** A data-link layer address unambiguously identifies every physical network association of a network device. Data-link addresses typically area unit cited as physical or hardware addresses. Data-link addresses sometimes exist among a flat address area and have a pre-established and usually fastened relationship to a selected device. End systems usually have just one physical network association, and therefore have just one data-link address. Routers and different internetworking devices usually have multiple physical network connections and so eventually have multiple data-link addresses.
2. **MAC Addresses:** Media Access management (MAC) addresses encompass a set of data-link layer addresses. MAC addresses establish network entities in LANs that implement the IEEE MAC addresses of the data-link layer. MAC addresses different area unit distinctively for every local area network interface. MAC addresses are forty-eight bits long and are expressed in form of twelve hexadecimal digits. The primary half dozen hexadecimal digits, that are usually administered by the IEEE, establish the manufacturer or merchant and therefore comprise the Organizational Unique Identifier (OUI). The last half dozen positional notation digits comprise the interface serial variety or another price administered by the particular merchant. MAC addresses typically area unit referred to as burned-in addresses (BIAs) as a result of burned into read-only memory(ROM) and are traced into random-access memory (RAM) once the interface card initializes.
3. **Network-Layer Addresses:** Network addresses sometimes exist among a gradable address area and typically area unit referred to as virtual or logical addresses. the connection between a network address and a tool is logical and unfixed, it usually relies either on physical network characteristics or on groupings that don't have any physical basis. finish systems need one network-layer address for every network-layer protocol they support. Routers and different Internetworking devices need one network-layer address per physical network association for every network-layer protocol supported.

Challenges to Internetworking :

Implementing a useful internetwork isn't at any certainty. There are several challenging fields, particularly in the areas of dependableness, connectivity, network management, and adaptability and each and every space is essential in establishing associate degree economical and effective internetwork. Few of them are:-

- The initial challenge lies when we are trying to connect numerous systems to support communication between disparate technologies. For example, Totally different sites might use different kinds of media, or they could operate at variable speeds.
- Another essential thought is reliable service that should be maintained in an internetwork. Individual users and whole organizations depend upon consistent, reliable access to network resources.
- Network management should give centralized support associate degreed troubleshooting capabilities in an internetwork. Configuration, security, performance, and different problems should be adequately addressed for the internetwork to perform swimmingly.
- Flexibility, the ultimate concern, is important for network enlargement and new applications and services, among different factors.

7.4 Internet Protocol Operations

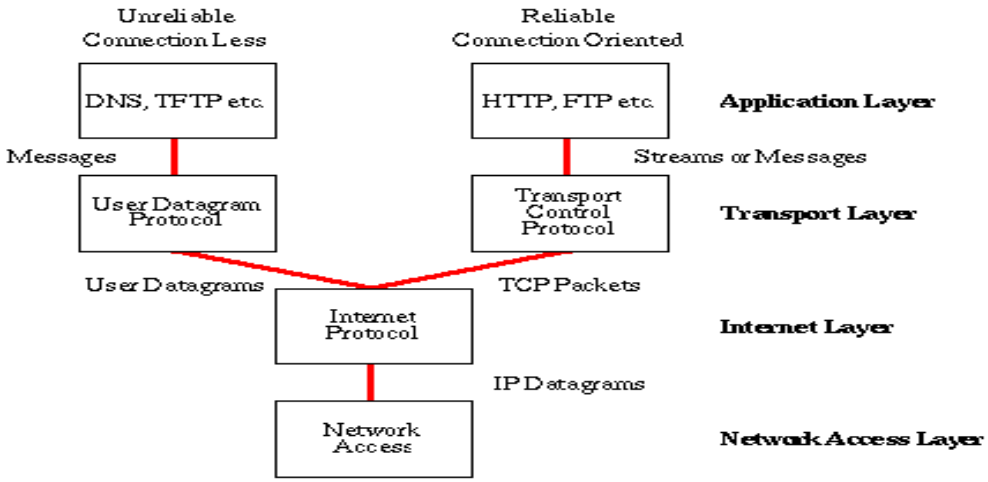
An **Internet Protocol address** (**IP address**) is a unique numerical label assigned to each device connected to a computer network that uses the Internet Protocol for communication. An IP address serves two main functions: host or network interface identification and location addressing.

Internet Protocol is **connectionless** and **unreliable** protocol. It ensures no guarantee of successfully transmission of data. In order to make it reliable, it must be paired with reliable protocol such as TCP at the transport layer.

The IP protocol and its associated routing protocols are possibly the most significant of the entire TCP/IP suite. IP is responsible for:

- **IP addressing** - The IP addressing conventions are part of the IP protocol.
- **Host-to-host communications** - IP determines the path a packet must take, based on the receiving host's IP address.
- **Packet formatting** - IP assembles packets into units known as **IP datagrams**.
- **Fragmentation** - If a packet is too large for transmission over the network media, IP on the sending host breaks the packet into smaller fragments. IP on the receiving host then reconstructs the fragments into the original packet.

Internet is an abstraction from the underlying network technologies and physical address resolution. This section introduces the basic components of the Internet protocol stack and relates the stack to the OSI reference protocol stack model. The model of the Internet protocol stack is illustrated in the figure below.



As seen in the figure above, the Internet protocol stack provides a connection oriented reliable branch (TCP) and an connectionless unreliable branch (UDP) both build on top of the Internet Protocol.

The Internet Protocol layer in the TCP/IP protocol stack is the first layer that introduces the virtual network abstraction that is the basic principle of the Internet model. All physical implementation details (ideally even though this is not quite true) are hidden below the IP layer. The IP layer provides an unreliable, connectionless delivery system. The reason why it is *unreliable* stem from the fact the protocol does not provide any functionality for error recovering for datagrams that are either duplicated, lost or arrive to the remote host in another order than they are send. If no such errors occur in the physical layer, the IP protocol guarantees that the transmission is terminated successfully.

The basic unit of data exchange in the IP layer is the Internet Datagram. The format of an IP datagram and a short description of the most important fields are included below:

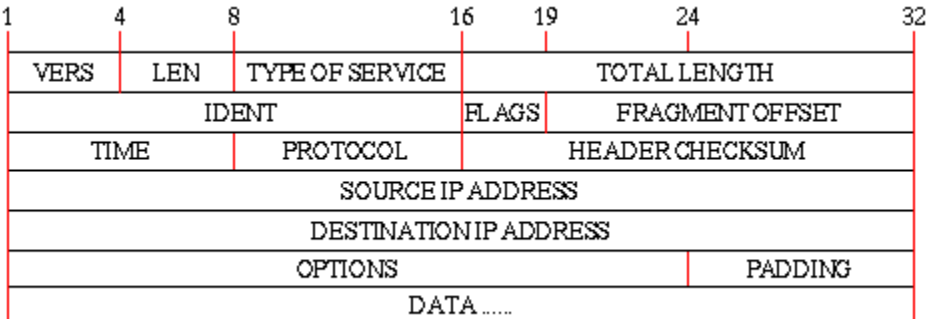


Fig. IP Header

IPV4 Header:

LEN: The number of 32 bit-segments in the IP header. Without any OPTIONS, this value is 5

TYPE OF SERVICE: Each IP datagram can be given a precedence value ranging from 0-7 showing the importance of the datagram. This is to allow *out-of-band* data to be routed faster than normal data. This is very important as Internet Control Message Protocol (ICMP) messages travels as the data part of an IP datagram. Even though an ICMP message is encapsulated in a IP datagram, the ICMP protocol is normally thought of as a integral part of the IP layer and not the UDP or TCP layer. Furthermore, the TYPE OF SERVICE field allows a classification of the datagram in order to specify is the service desired requires short delay time, high reliability or high throughput. However, in order for this to have any effect, the gateways must know more than one route to the remote host and as described in the Introduction, this is not the case.

IDENT, FLAGS, and FRAGMENT OFFSET: These fields are used to describe fragmentation of a datagram. The actual length of an IP datagram is in principle independent of the length of the physical frames being transferred on the network, referred to as the network's *Maximum Transfer Unit (MTU)*. If a datagram is longer than the MTU then it is divided in to a set of fragments having almost the same header as the original datagram but only the amount of data that fits into a physical frame. The IDENT flag is used to identify segments belonging to the same datagram, and the FRAGMENT OFFSET is the relative position of the fragment within the original datagram. Once a datagram is fragmented it stays like that until it receives the final destination. If one or more segments are lost or erroneous the whole datagram is discarded.

However, the underlying network technology is not completely hidden below the IP layer in spite of the fragmentation functionality. The reason is that the MTU can vary from 128 or less to several thousands of bytes dependent of the physical network (Ethernet has a MTU of 1500 bytes). It is hence question of efficiency when choosing the right datagram size so that fragmentation is minimized. It is recommended that gateways are capable of handling datagrams of at least 576 bytes without having to use fragmentation.

TIME: This is the remaining **Time To Live (TTL)** for a datagram when it travels on the Internet. The Routing Information Protocol (RIP) specifies that at most 15 hops are allowed.

SOURCE IP-ADDRESS and DESTINATION IP-ADDRESS: Both the source and destination address is indicated in the datagram header so that the recipient can send an answer back to the transmitting host. However, note that only the host address is specified - not the port number. This is because the IP protocol is an IMP-to-IMP protocol - it is *not* an end-to-end protocol. A layer more is needed to actually specify which two processes on the transmitting host and the final destination that should receive the datagrams.

Note that the IP-datagram only leaves space for the original source IP-address and the original destination IP-address. The *next hop* address is specified by encapsulation. The **Internet Layer** passes the IP-address of the *next hop* address to the **Network Layer**. This IP-address is bound to a physical address and a new frame is formed with this address. The rest of the original frame is then encapsulated in the new frame before it is send over the communication channel.

7.5 Internet Protocol

An IP address *(*internet protocol address*)* is a numerical representation that uniquely identifies a specific interface on the network. Addresses in IPv4 are 32-bits long. This allows for a maximum of 4,294,967,296 (2^{32}) unique addresses. Addresses in IPv6 are 128-bits, which allows for 3.4 x 10^{38} (2^{128}) unique addresses. The total usable address pool of both versions is reduced by various reserved addresses and other considerations. IP addresses are binary numbers but are typically expressed in decimal form (IPv4) or hexadecimal form (IPv6) to make reading and using them easier for humans.

IP versions: There are two versions of IP in use today, IPv4 and IPv6. The original IPv4 protocol is still used today on both the internet, and many corporate networks. However, the IPv4 protocol only allowed for 2^{32} addresses. This, coupled with how addresses were allocated, led to a situation where there would not be enough unique addresses for all devices connected to the internet.

IPv6 was developed by the Internet Engineering Task Force (IETF), and was formalized in 1998. This upgrade substantially increased the available address space and allowed for 2^{128} addresses. In addition, there were changes to improve the efficiency of IP packet headers, as well as improvements to routing and security.
IPv4 addresses: IPv4 addresses are actually 32-bit binary numbers, consisting of the two subaddresses (identifiers) mentioned above which, respectively, identify the network and the host to the network, with an imaginary boundary separating the two. An IP address is, as such, generally shown as 4 octets of numbers from 0-255 represented in decimal form instead of binary form.

For example, the address 168.212.226.204 represents the 32-bit binary number 10101000.11010100.11100010.11001100.

The binary number is important because that will determine which class of network the IP address belongs to. An IPv4 address is typically expressed in dotted-decimal notation, with every eight bits (octet) represented by a number from one to 255, each separated by a dot. An example IPv4 address would look like this: **192.168.17.43**

IPv4 addresses are composed of two parts. The first numbers in the address specify the network, while the latter numbers specify the specific host. A subnet mask specifies which part of an address is the network part, and which part addresses the specific host.

A packet with a destination address that is not on the same network as the source address will be forwarded, or routed, to the appropriate network. Once on the correct network, the host part of the address determines which interface the packet gets delivered to.
Subnet masks: A single IP address identifies both a network, and a unique interface on that network. A subnet mask can also be written in dotted decimal notation and determines where the network part of an IP address ends, and the host portion of the address begins.

When expressed in binary, any bit set to one means the corresponding bit in the IP address is part of the network address. All the bits set to zero mark the corresponding bits in the IP address as part of the host address.

The bits marking the subnet mask must be consecutive ones. Most subnet masks start with 255. and continue on until the network mask ends. A Class C subnet mask would be 255.255.255.0.

IP address classes:

Class	Leading bits	Size of *network number* bit field	Size of *rest* bit field	Number of networks	Addresses per network	Total addresses in class	Start address	End address
Class A	0	8	24	128 (2^7)	16,777,216 (2^{24})	2,147,483,648 (2^{31})	0.0.0.0	127.255.255.255
Class B	10	16	16	16,384 (2^{14})	65,536 (2^{16})	1,073,741,824 (2^{30})	128.0.0.0	191.255.255.255
Class C	110	24	8	2,097,152 (2^{21})	256 (2^8)	536,870,912 (2^{29})	192.0.0.0	223.255.255.255
Class D (multicast)	1110	not defined	not defined	not defined	not defined	268,435,456 (2^{28})	224.0.0.0	239.255.255.255
Class E (reserved)	1111	not defined	not defined	not defined	not defined	268,435,456 (2^{28})	240.0.0.0	255.255.255.255

Before variable length subnet masks allowed networks of any size to be configured, the IPv4 address space was broken into five classes.

Class A: In a Class A network, the first eight bits, or the first dotted decimal, is the network part of the address, with the remaining part of the address being the host part of the address. There are 128 possible Class A networks.
0.0.0.0 to 127.0.0.0
However, any address that begins with 127. is considered a loopback address.
Example for a Class A IP address:
2.134.213.2

Class B: In a Class B network, the first 16 bits are the network part of the address. All Class B networks have their first bit set to 1 and the second bit set to 0. In dotted decimal notation, that makes 128.0.0.0 to 191.255.0.0 as Class B networks. There are 16,384 possible Class B networks.
Example for a Class B IP address:
135.58.24.17

Class C: In a Class C network, the first two bits are set to 1, and the third bit is set to 0. That makes the first 24 bits of the address the network address and the remainder as the host address. Class C network addresses range from 192.0.0.0 to 223.255.255.0. There are over 2 million possible Class C networks.
Example for a Class C IP address:
192.168.178.1

Class D: Class D addresses are used for multicasting applications. Unlike the previous classes, the Class D is not used for "normal" networking operations. Class D addresses have their first three bits set to "1" and their fourth bit set to "0". Class D addresses are 32-bit network addresses, meaning that all the values within the range of 224.0.0.0 – 239.255.255.255 are used to uniquely identify multicast groups. There are no host addresses within the Class D address space, since all the hosts within a group share the group's IP address for receiver purposes.
Example for a Class D IP address:
227.21.6.173

Class E: Class E networks are defined by having the first four network address bits as 1. That encompasses addresses from 240.0.0.0 to 255.255.255.255. While this class is reserved, its usage was never defined. As a result, most network implementations discard these addresses as illegal or undefined. The exception is 255.255.255.255, which is used as a broadcast address.
Example for a Class D IP address:

```
243.164.89.28
```

Overview: IP address classes and bit-wise representations

```
Class A
  0.  0.  0.  0 = 00000000.00000000.00000000.00000000
127.255.255.255 = 01111111.11111111.11111111.11111111
                  0nnnnnnn.HHHHHHHH.HHHHHHHH.HHHHHHHH

Class B
128.  0.  0.  0 = 10000000.00000000.00000000.00000000
191.255.255.255 = 10111111.11111111.11111111.11111111
                  10nnnnnn.nnnnnnnn.HHHHHHHH.HHHHHHHH

Class C
192.  0.  0.  0 = 11000000.00000000.00000000.00000000
223.255.255.255 = 11011111.11111111.11111111.11111111
                  110nnnnn.nnnnnnnn.nnnnnnnn.HHHHHHHH

Class D
224.  0.  0.  0 = 11100000.00000000.00000000.00000000
239.255.255.255 = 11101111.11111111.11111111.11111111
                  1110XXXX.XXXXXXXX.XXXXXXXX.XXXXXXXX

Class E
240.  0.  0.  0 = 11110000.00000000.00000000.00000000
255.255.255.255 = 11111111.11111111.11111111.11111111
                  1111XXXX.XXXXXXXX.XXXXXXXX.XXXXXXXX
```

Private addresses: Within the address space, certain networks are reserved for private networks. Packets from these networks are not routed across the public internet. This provides a way for private networks to use internal IP addresses without interfering with other networks. The private networks are

```
10.0.0.1 - 10.255.255.255

172.16.0.0 - 172.32.255.255

192.168.0.0 - 192.168.255.255
```

Special addresses: Certain IPv4 addresses are set aside for specific uses:

127.0.0.0	Loopback address (the host's own interface)
224.0.0.0	IP Multicast
255.255.255.255	Broadcast (sent to all interfaces on network)

IPv4 address exhaustion: The original IPv4 specification was designed for the DARPA network that would eventually become the internet. Originally a test network, no one contemplated how many addresses might be needed in the future. At the time, the 2^{32} addresses (4.3 billion) were certainly considered sufficient. However, over time, it became apparent that as currently implemented, the IPv4 address space would not be big enough for a worldwide internet with numerous connected devices per person. The last top-level address blocks were allocated in 2011.

IPv6 addresses: To avoid the seemingly reoccurring issue in technology, where a specification's limitation seems more than sufficient at the time, but inevitably becomes too small, the designers of IPv6 created an enormous address space for IPv6. The address size was increased from 32 bits in IPv4 to 128 bits in IPv6.

The IPv6 has a theoretical limit of 3.4×10^{38} addresses. That's over 340 undecillion addresses, which is reportedly enough addresses to assign one to every single atom on the surface of the earth.

IPv6 addresses are represented by eight sets of four hexadecimal digits, and each set of numbers is separated by a colon. An example IPv6 address would look like this:

2DAB:FFFF:0000:3EAE:01AA:00FF:DD72:2C4A

IPv6 address abbreviation: With IPv6 addresses being so long, there are conventions to allow for their abbreviation. First, leading zeros from any one group of numbers may be eliminated. For example, :0033: can be written as :33:

Second, any consecutive sections of zeros can be represented by a double colon. This may be done only once in any address. The number of sections removed using this abbreviation can be determined as the number required to bring the address back up to eight sections. For example, 2DAB::DD72:2C4A would need to have five sections of zeroes added back in place of the double colon.

(2DAB:0000:0000:0000:0000:0000:DD72:2C4A)

The loopback address

0000:0000:0000:0000:0000:0000:0000:0001

may be abbreviated as ::1.

IPv6 private addresses: Like in IPv4 certain address blocks are reserved for private networks. These addresses are not routed over the public internet. In IPv6, private addresses are called Unique Local Addresses (ULA). Addresses from the FC00:: /7 block are ignored and not routed by default.

Name resolution: In both IPv4 and IPv6, remembering the IP address of every device is not possible, except on the smallest of networks. Name resolution provides a way to lookup an IP address from an easier to use name.

On the internet, name resolution is handled by the Domain Name System (DNS). With DNS, a name in the format *host.domain* can be used in place of the destination's IP address. When the connection is initiated, the source host will request the IP address of the destination host from a DNS server. The DNS server will reply with the destination's IP address. This IP address will then be used for all communications sent to that name.

Author:
Deepika Panda