Unit 1: Introduction

1.1 Network as an Infrastructure for 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. For data communications to occur, the communicating devices must be part of a communication system made up of a combination of hardware (physical equipment) and software (programs).

Data Communication Networking:

Data communications refers to the transmission of digital data between two or more computers and a computer network or data network is a telecommunications network that allows computers to exchange data. The physical connection between networked computing devices is established using either cable media or wireless media. The best-known computer network is the Internet.

The communication system that consists of the interconnection between two or more devices is referred to as a Network. 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.

A Basic Communication Model:

A data communications system has five components.

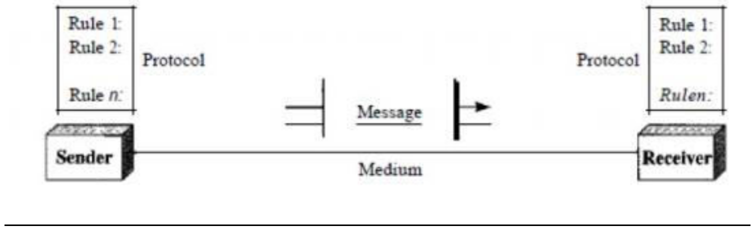


Fig: Data communication model

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.

Characteristics of Data Communication System:

The effectiveness of a data communications system depends on four fundamental characteristics: delivery, accuracy, timeliness, and jitter.

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

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

3. Timeliness. The system must deliver data in a timely manner. Data delivered late are useless. In the case of video and audio, timely delivery means delivering data as they are produced, in the same order that they are produced, and without significant delay. This kind of delivery is called real-time transmission.

4. Jitter. Jitter refers to the variation in the packet arrival time. It is the uneven delay in the delivery of audio or video packets.

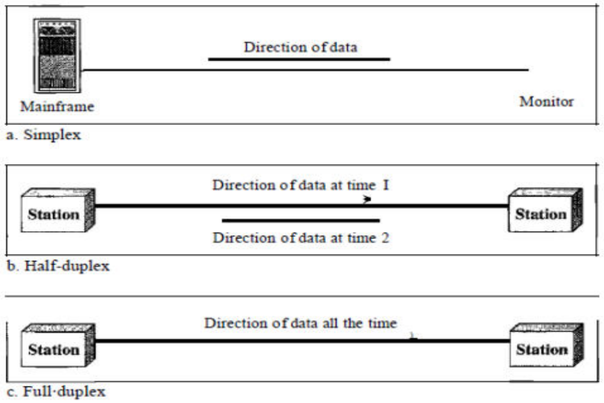
Data Transmission Modes

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

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. 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. 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 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. The full-duplex mode is like a two-way 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. 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.


Direction of data
Mainframe
Monitor
Direction of data at time 1
Station
Station
Direction of data at time 2
Direction of data all the time
Station
Station

a. Simplex

b. Half-duplex

c. Full-duplex

Network Criteria

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

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

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

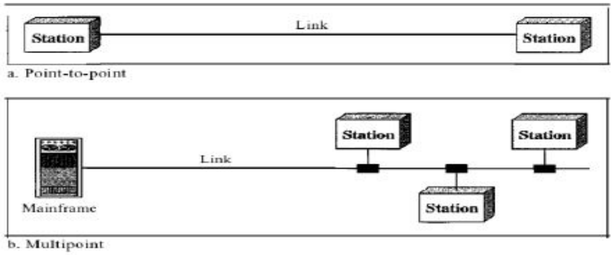
Security: Network security issues include protecting data from unauthorized access, protecting data from damage and development, and implementing policies and procedures for recovery from breaches and data losses.

Physical Structures:

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

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

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

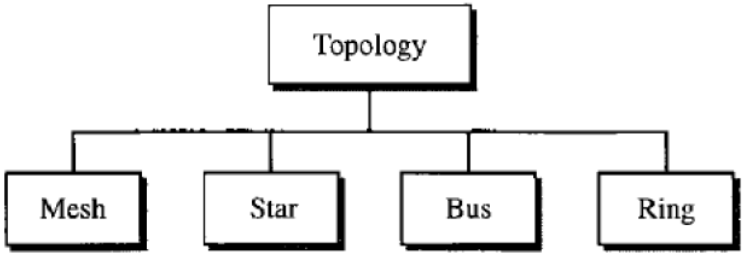


a. Point-to-point

b. Multipoint

Network Topology:

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



1. Mesh:

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

Advantages:

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

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

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

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

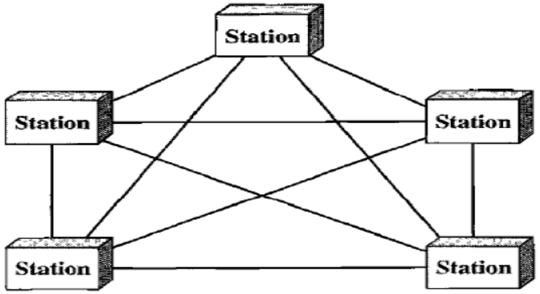
Disadvantages:

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

2. Installation and reconnection are difficult.

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

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



2. Star Topology:

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

Advantages:

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

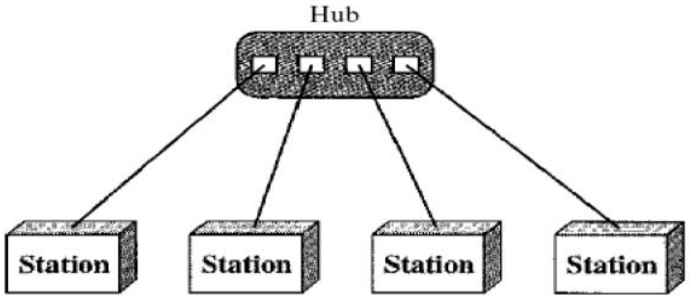
2. Easy to install and reconfigure.

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

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

Disadvantages:

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



3. BUS:

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

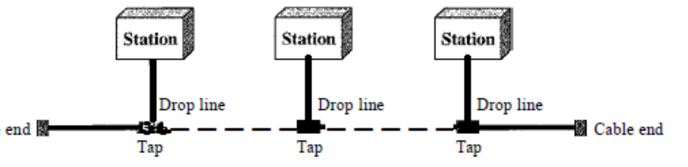
Advantages:

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

Disadvantages:

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

A bus topology connecting three stations



4. RING:

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

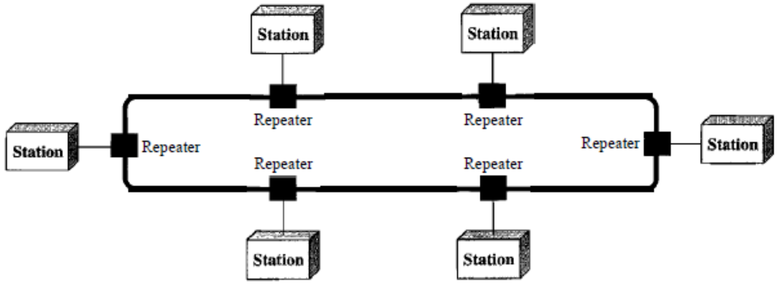
Advantages:

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

Disadvantages:

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

Figure 1.8 *A ring topology connecting six stations*

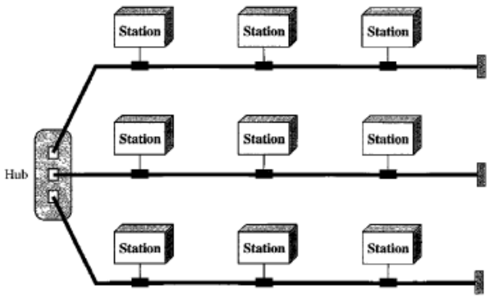


5. Hybrid Topology

A network can be hybrid. Hybrid topologies combine two or more different topology structures—the tree topology is a good example, integrating the bus and star layouts. Hybrid structures are most commonly found in larger companies where individual departments have personalized network topologies adapted to suit their needs and network usage.

For example, we can have a main star topology with each branch consisting several stations in a bus topology as shown below.

A hybrid topology: a star backbone with three bus networks



Advantages of Hybrid Topology

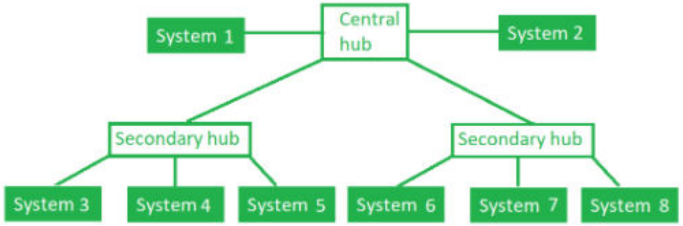
The main advantage of hybrid structures is the degree of flexibility they provide, as there are few limitations on the network structure itself that a hybrid setup can't accommodate.

Disadvantages of Hybrid Topology

However, each type of network topology comes with its own disadvantages, and as a network grows in complexity, so too does the experience and know-how required on the part of the admins to keep everything functioning optimally. There's also the monetary cost to consider when creating a hybrid network topology.

6. 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.



In this topology, the various secondary hubs are connected to the central hub which contains the repeater. The data flows from top to bottom i.e., from the central hub to secondary and then to the devices or from bottom to top i.e., devices to secondary hub and then to the central hub.

Advantages:

- It allows more devices to be attached to a single central hub thus it increases the distance that is travel by the signal to come to the devices.
- It allows the network to get isolate and also prioritize from different computers.

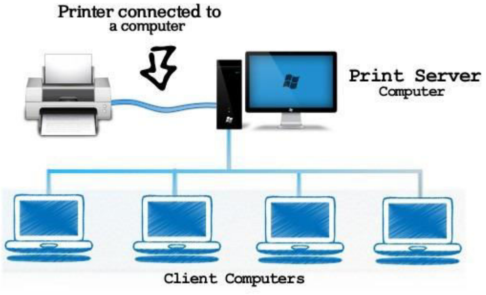
Disadvantages:

- If the central hub gets fails the entire system fails.
- The cost is high because of cabling

1.2 Applications of Computer Network

Networking of computers provides a communication link between the users, and provides access to information. Networking of computers has several applications, described as follows:

- Resource Sharing—In an organization, resources such as printers, fax machines and scanners are generally not required by each person at all times. Moreover, for small organizations it may not be feasible to provide such resources to each individual. Such resources can be made available to different users of the organization on the network. It results in availability of the resource to different users regardless of the physical location of the resource or the user, enhances optimal use of the resource, leads to easy maintenance, and saves cost too.



Access Shared Printer connected to a Computer/Server

- Sharing of Information—In addition to the sharing of physical resources, networking facilitates sharing of information. Information stored on networked computers located at same or different physical locations, becomes accessible to the computers connected to the network.
- As a Communication Medium—Networking helps in sending and receiving of electronic-mail (email) messages from anywhere in the world. Data in the form of text, audio, video and pictures can be sent via e-mail. This allows the users to communicate online in a faster and cost-effective manner. Video conferencing is another form of communication made possible via networking. People in distant locations can hold a meeting, and they can hear and see each other simultaneously.
- For Back-up and Support—Networked computers can be used to take back-up of critical data. In situations where there is a requirement of always-on computer, another computer on the network can take over in case of failure of one computer.
- Server-Client Communication- One can imagine a company's information system as consisting of one or more databases and some employees who need to access it remotely. In this model, the data is stored on powerful computers called Servers. Often these are centrally housed and maintained by a system administrator. In contrast, the employees have simple machines, called Clients, on their desks, using which they access remote data.
- eCommerce- A goal that is starting to become more important in businesses is doing business with consumers over the Internet. Airlines, bookstores and music vendors have discovered that many customers like the convenience of shopping from home.
- Interactive entertainment- Interactive entertainment includes:

   - Multi-person real-time simulation games.
   - Video on demand.
   - Participation in live TV programs likes quiz, contest, discussions etc.

  In short, the ability to merge information, communication and entertainment will surely give rise to a massive new industry based on computer networking.
- Retrieving Remote Information − Through computer networks, users can retrieve remote information on a variety of topics. The information is stored in remote databases to which the user gains access through information systems like the World Wide Web.
- VoIP – VoIP or Voice over Internet protocol has revolutionized telecommunication systems. Through this, telephone calls are made digitally using Internet Protocols instead of the regular analog phone lines.

1.3 Network Architecture

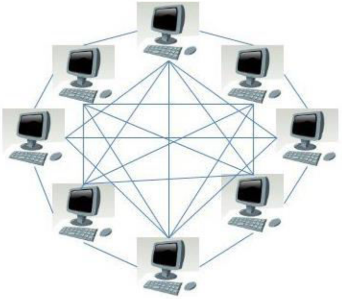
Network architecture is the design of a computer network. It is a framework for the specification of a network's physical components and their functional organization and configuration, its operational principles and procedures, as well as communication protocols used.

There are two major types of network architectures:

- Peer-To-Peer Architecture
- Client/Server Architecture

Peer-To-Peer Architecture

- In a peer-to-peer network, tasks are allocated to every device on the network.
- Furthermore, there is no real hierarchy in this network, all computers are considered equal and all have the same abilities to use the resources available on this network.
- Instead of having a central server which would act as the shared drive, each computer that’s connected to this network would act as the server for the files stored on it.



Peer-to-Peer Network Model

Advantages of Peer-To-Peer Network:

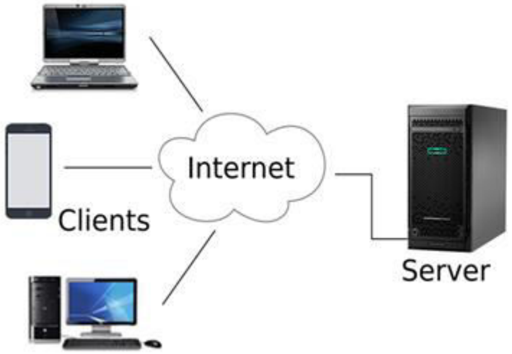
- It is less costly as it does not contain any dedicated server.
- If one computer stops working but, other computers will continue working.
- It is easy to set up and maintain as each computer manages itself.

Disadvantages of Peer-To-Peer Network:

- In the case of Peer-To-Peer network, it does not contain the centralized system. Therefore, it cannot back up the whole data as the data is different in different locations.
- Security and data backups are to be done to each individual computer.
- As the numbers of computers increases on a P2P network; performance, security, and access become a major headache.

Client/Server Architecture

- Client-server architecture, architecture of a computer network in which many clients (remote processors) request and receive service from a centralized server (host computer).
- In a client/server network, a centralized, really powerful computer(server) act as a hub in which other computers or workstations(clients) can connect to. This server is the heart of the system, which manages and provides resources to any client that requests them.
- A server performs all the major operations such as security and network management.
- A server is responsible for managing all the resources such as files, directories, printer, etc.
- All the clients communicate with each other through a server. For example, if client1 wants to send some data to client 2, then it first sends the request to the server for the permission. The server sends the response to the client 1 to initiate its communication with the client 2.



Advantages of a client/server network

- Resources and data security are controlled through the server.
- Not restricted to a small number of computers.
- Server can be accessed anywhere and across multiple platforms.

Disadvantages of a client/server network

- Can become very costly due to the need of a server as well as networking devices such as hubs, routers, and switches.
- If and when the server goes down, the entire network will be affected.
- Technical staff needed to maintain and ensure network functions efficiently.

Key Differences Between Client-Server and Peer-to-Peer Network

- The key difference between Client-Server and Peer-to-Peer network is that there is a dedicated server and specific clients in the client-server network model whereas, in peer-to-peer each node can act as both server and client.
- In the client-server model, the server provides services to the client. However, in peer-to-peer, each peer can provide services and can also request for the services.
- In the client-server model, sharing information is more important whereas, in peer-to-peer model connectivity between peers is more important.
- In the client-server model, data is stored on a centralized server whereas, in peer-to-peer each peer has its own data.
- In peer-to-peer model, the servers are distributed in a system, so there are fewer chances of server getting bottlenecked, but in the client-server model, there is a single server serving the clients, so there are more chances of server getting bottlenecked.
- The client-server model is more expensive to implement than peer-to-peer.
- The client-server model is more scalable and stable than peer-to-peer.

1.4 Types of Computer Networks

There are two primary types of computer networking: wired networking and wireless networking.

1. Wired networking requires the use of a physical medium for transport between nodes. Copper-based Ethernet cabling, popular due to its low cost and durability, is commonly used for digital communications in businesses and homes. Alternatively, optical fiber is used to transport data over greater distances and at faster speeds, but it has several tradeoffs, including higher costs and more fragile components.
2. Wireless networking uses radio waves to transport data over the air, enabling devices to be connected to a network without any cabling. Wireless LANs are the most well-known and widely deployed form of wireless networking. Alternatives include microwave, satellite, cellular and Bluetooth, among others.

As a general rule, wired networking offers greater speed, reliability and security compared to wireless networks; wireless networking tends to provide more flexibility, mobility and scalability.

There are several different types of computer networks which can be characterized by their size as well as their purpose. The size of a network can be expressed by the geographic area they occupy and the number of computers that are part of the network. Networks can cover anything from a handful of devices within a single room to millions of devices spread across the entire globe.

Some of the different networks based on size are:

- Personal area network, or PAN
- Local area network, or LAN
- Metropolitan area network, or MAN
- Wide area network, or WAN

Personal Area Network (PAN):

The smallest and most basic type of network, a PAN is made up of a wireless modem, a computer or two, phones, printers, tablets, etc., and revolves around one person in one building. These types of networks are typically found in small offices or residences, and are managed by one person or organization from a single device.

Local Area Networks (LAN):

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 organizations 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

(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. 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.

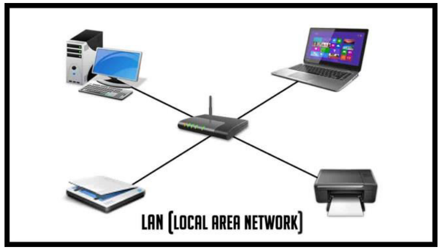


Fig: Local Area Network

Characteristics of LAN:

- LANs are private networks, not subject to external control
- Simple and better performance
- Work in a restricted geographical area

Advantages:

- Resource sharing
- Software applications sharing
- Easy and Cheap communication
- Data Security
- Internet sharing

Disadvantages

- Restricted to local area

Metropolitan Area Network (MAN):

A metropolitan area network, or MAN, covers a city. A MAN is a computer network that interconnects users with computer resources in a geographical area or region larger than that covered by a LAN. It can be an interconnection between several LANs by bridging them with backbone lines.

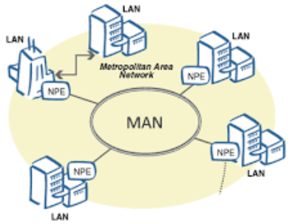


Fig: Metropolitan Area Network

Characteristics:

- Generally, covers towns and cities (up to 50km)
- Transmission medium used for MAN is optical fiber, coaxial cable etc.
- Data rates adequate for distributed computing applications

Advantages

- Extremely efficient and provide fast communication via high-speed carriers, such as fiber optic cables
- Good backbone for larger networks and provides greater access to WAN

Disadvantages

- Complex, more cabling required and expensive

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. 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.

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. WANs are typically used to connect two or more LANs or MANs which are located relatively very far from each other.

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. Not all WANs are packet switched. A second possibility for a WAN is a satellite system. Each router has an antenna through which it can send and receive. All routers can hear the output from the satellite, and in some cases, they can also hear the upward transmissions of their fellow routers to the satellite as well. Sometimes the routers are connected to a substantial point-to-point subnet, with only some of them having a satellite antenna. Satellite networks are inherently broadcast and are most useful when the broadcast property is important.

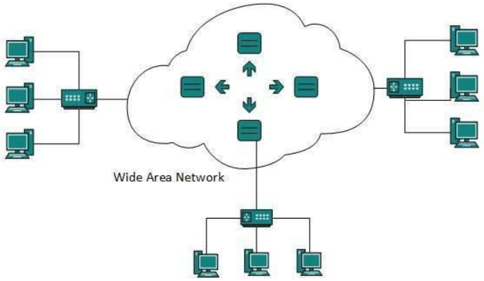


Fig: Wide Area Network

Characteristics

- Covers large distances (states, countries, continents)
- Communication medium used are satellite, public telephone networks which are connected by routers

Advantages

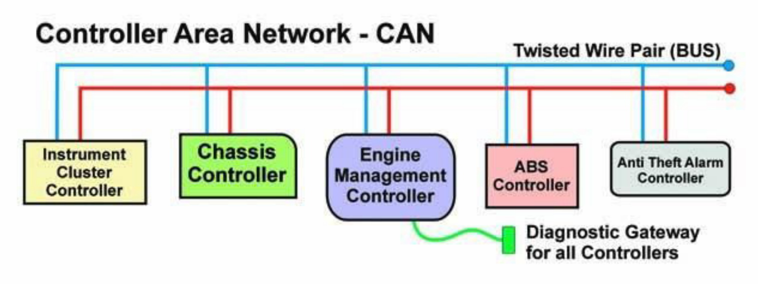
- Covers large geographical area
- Shared software and resources with connecting workstations
- Information can be exchanged to anyone else worldwide in the network

Disadvantages

- Data security
- Network is very complex and management is difficult
- As size increases, the networks become more expensive

Controller Area Network (CAN):

A Controller Area Network (CAN bus) is a robust vehicle bus standard designed to allow microcontrollers and devices to communicate with each other's applications without a host computer. It is a message-based protocol, designed originally for multiplex electrical wiring within automobiles to save on copper, but can also be used in many other contexts. For each device the data in a frame is transmitted sequentially but in such a way that if more than one device transmits at the same time the highest priority device is able to continue while the others back off. Frames are received by all devices, including by the transmitting device.


Controller Area Network - CAN
Twisted Wire Pair (BUS)
Instrument Cluster Controller
Chassis Controller
Engine Management Controller
ABS Controller
Anti Theft Alarm Controller
Diagnostic Gateway for all Controllers

Storage-Area Network (SAN)

As a dedicated high-speed network that connects shared pools of storage devices to several servers, these types of networks don't rely on a LAN or WAN. Instead, they move storage resources away from the network and place them into their own high-performance network. SANs can be accessed in the same fashion as a drive attached to a server. Types of storage-area networks include converged, virtual and unified SANs.

Enterprise Private Network (EPN)

These types of networks are built and owned by businesses that want to securely connect its various locations to share computer resources.

Virtual Private Network (VPN)

By extending a private network across the Internet, a VPN lets its users send and receive data as if their devices were connected to the private network – even if they're not. Through a virtual point-to-point connection, users can access a private network remotely.

1.5 Protocols and Standards

Protocol is the set of rule and standard is agreed upon rules. These are the two widely used terms in networking.

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.

1.6 The OSI Reference Model

An Open Systems Interconnection (OSI) Model is a set of protocols that allows any two different systems to communicate regardless of their underlying architecture. The purpose of the OSI Model is to show how to facilitate communication between different systems without requiring changes to the logic of the underlying hardware and software. The OSI Model is not a protocol; it is a model for understanding and designing a network architecture that is flexible, robust, and interoperable. The OSI model is a layered framework for the design of network systems that allows communication between all types of computer systems. It consists of seven separate but related layers, each of which defines a part of the process of moving information across a network.

Seven layers of the OSI model

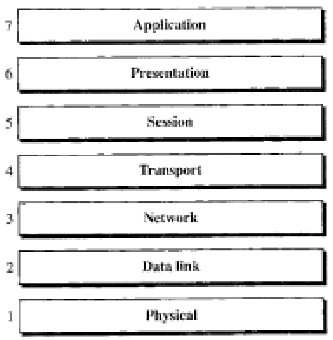
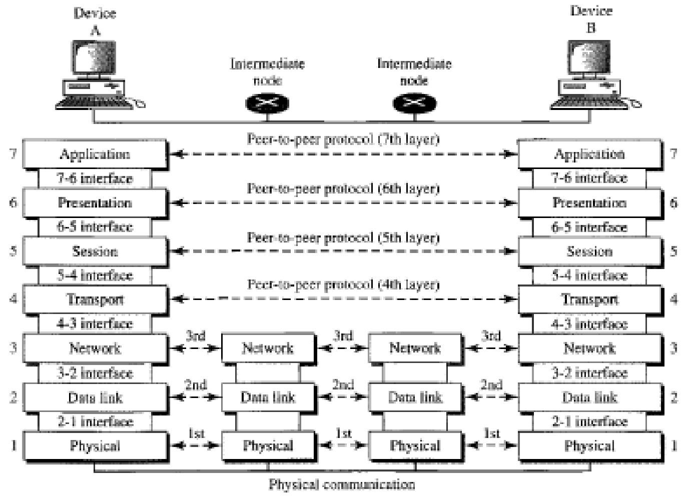


Figure 2.3 *The interaction between layers in the OSI model*



1. **Physical Layer:**
   The physical layer coordinates the functions required to carry a bit stream over a physical medium. It deals with the mechanical and electrical specifications of the interface and

transmission medium. It also defines the procedures and functions that physical devices and interfaces have to perform for transmission to occur.

The physical layer is also concerned with the following:

- Physical characteristics of interfaces and medium. The physical layer defines the characteristics of the interface between the devices and the transmission medium. It also defines the type of transmission medium.
- Representation of bits. The physical layer data consists of a stream of bits (sequence of 0s or 1s) with no interpretation. To be transmitted, bits must be encoded into signals--electrical or optical. The physical layer defines the type of encoding.
- Data rate. The transmission rate-the number of bits sent each second-is also defined by the physical layer. In other words, the physical layer defines the duration of a bit, which is how long it lasts.
- Synchronization of bits. The sender and receiver not only must use the same bit rate but also must be synchronized at the bit level. In other words, the sender and the receiver clocks must be synchronized.
- Line configuration. The physical layer is concerned with the connection of devices to the media. In a point-to-point configuration, two devices are connected through a dedicated link. In a multipoint configuration, a link is shared among several devices.
- Physical topology. The physical topology defines how devices are connected to make a network. Devices can be connected by using a mesh topology (every device is connected to every other device), a star topology (devices are connected through a central device), a ring topology (each device is connected to the next, forming a ring), a bus topology (every device is on a common link), or a hybrid topology (this is a combination of two or more topologies).
- Transmission mode. The physical layer also defines the direction of transmission between two devices: simplex, half-duplex, or full-duplex. In simplex mode, only one device can send; the other can only receive. The simplex mode is a one-way communication. In the half-duplex mode, two devices can send and receive, but not at the same time. In a full-duplex (or simply duplex) mode, two devices can send and receive at the same time.

2. Data Link Layer

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

Other responsibilities of the data link layer include the following:

- Framing. The data link layer divides the stream of bits received from the network layer into manageable data units called frames.
- Physical addressing. If frames are to be distributed to different systems on the network, the data link layer adds a header to the frame to define the sender and/or receiver of the frame. If the frame is intended for a system outside the sender's network, the receiver address is the address of the device that connects the network to the next one.
- Flow control. If the rate at which the data are absorbed by the receiver is less than the rate at which data are produced in the sender, the data link layer imposes a flow control mechanism to avoid overwhelming the receiver.
- Error control. The data link layer adds reliability to the physical layer by adding mechanisms to detect and retransmit damaged or lost frames. It also uses a mechanism to recognize duplicate frames. Error control is normally achieved through a trailer added to the end of the frame.
- Access control. When two or more devices are connected to the same link, data link layer protocols are necessary to determine which device has control over the link at any given time.

3. Network Layer

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

Other responsibilities of the network layer include the following:

- Logical addressing. The physical addressing implemented by the data link layer handles the addressing problem locally. If a packet passes the network boundary, we need another addressing system to help distinguish the source and destination systems. The network layer adds a header to the packet coming from the upper layer that, among other things, includes the logical addresses of the sender and receiver.
- Routing. When independent networks or links are connected to create internetworks (network of networks) or a large network, the connecting devices (called routers or switches) route or switch the packets to their final destination. One of the functions of the network layer is to provide this mechanism.

4. Transport Layer

The transport layer is responsible for process-to-process delivery of the entire message. A process is an application program running on a host. Whereas the network layer oversees source-to-destination delivery of individual packets, it does not recognize any relationship between those packets. It treats each one independently, as though each piece belonged to a separate message, whether or not it does. The transport layer, on the other hand, ensures that the whole message arrives intact and in order, overseeing both error control and flow control at the source-to-destination level.

Other responsibilities of the transport layer include the following:

- Service-point addressing. Computers often run several programs at the same time. For this reason, source-to-destination delivery means delivery not only from one computer to the next but also from a specific process (running program) on one computer to a specific process (running program) on the other. The transport layer header must therefore include a type of address called a service-point address (or port address). The network layer gets each packet to the correct computer; the transport layer gets the entire message to the correct process on that computer.
- Segmentation and reassembly. A message is divided into transmittable segments, with each segment containing a sequence number. These numbers enable the transport layer to reassemble the message correctly upon arriving at the destination and to identify and replace packets that were lost in transmission.

- Connection control. The transport layer can be either connectionless or connection oriented. A connectionless transport layer treats each segment as an independent packet and delivers it to the transport layer at the destination machine. A connection-oriented transport layer makes a connection with the transport layer at the destination machine first before delivering the packets. After all the data are transferred, the connection is terminated.
- Flow control. Like the data link layer, the transport layer is responsible for flow control. However, flow control at this layer is performed end to end rather than across a single link.
- Error control. Like the data link layer, the transport layer is responsible for error control. However, error control at this layer is performed process-to-process rather than across a single link. The sending transport layer makes sure that the entire message arrives at the receiving transport layer without error (damage, loss, or duplication). Error correction is usually achieved through retransmission.

5. **Session Layer**

The services provided by the first three layers (physical, data link, and network) are not sufficient for some processes. The session layer is the network dialog controller. It establishes, maintains, and synchronizes the interaction among communicating systems.

Specific responsibilities of the session layer include the following:

- Dialog control. The session layer allows two systems to enter into a dialog. It allows the communication between two processes to take place in either half duplex (one way at a time) or full-duplex (two ways at a time) mode.
- Synchronization. The session layer allows a process to add checkpoints, or synchronization points, to a stream of data.

6. **Presentation Layer**

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

Specific responsibilities of the presentation layer include the following:

- Translation. The processes (running programs) in two systems are usually exchanging information in the form of character strings, numbers, and so on. The information must be changed to bit streams before being transmitted. Because different computers use different encoding systems, the presentation layer is responsible for interoperability between these different encoding methods. The presentation layer at the sender changes the information from its sender-dependent format into a common format. The presentation layer at the receiving machine changes the common format into its receiver-dependent format.
- Encryption. To carry sensitive information, a system must be able to ensure privacy. Encryption means that the sender transforms the original information to another form and sends the resulting message out over the network. Decryption reverses the original process to transform the message back to its original form.

- Compression. Data compression reduces the number of bits contained in the information. Data compression becomes particularly important in the transmission of multimedia such as text, audio, and video.

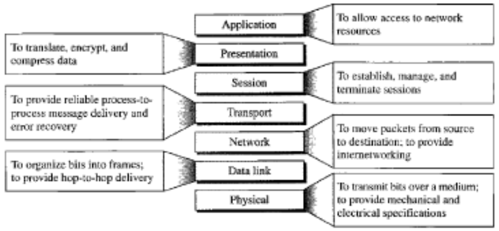
7. **Application Layer**

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

Specific services provided by the application layer include the following:

- Network virtual terminal. A network virtual terminal is a software version of a physical terminal, and it allows a user to log on to a remote host.
- File transfer, access, and management. This application allows a user to access files in a remote host (to make changes or read data), to retrieve files from a remote computer for use in the local computer, and to manage or control files in a remote computer locally.
- Mail services. This application provides the basis for e-mail forwarding and storage.
- Directory services. This application provides distributed database sources and access for global information about various objects and services.

Figure 2.15 *Summary of layers*


Application
To allow access to network resources
To translate, encrypt, and compress data
Presentation
Session
To establish, manage, and terminate sessions
To provide reliable process-to-process message delivery and error recovery
Transport
Network
To move packets from source to destination; to provide internetworking
To organize bits into frames; to provide hop-to-hop delivery
Data link
Physical
To transmit bits over a medium; to provide mechanical and electrical specifications

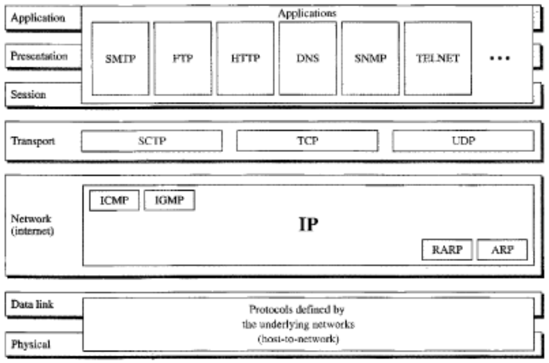
<u>TCP/IP Protocol Suite:</u>

The TCP/IP protocol suite was defined as having four layers: host-to-network, internet, transport, and application. However, when TCP/IP is compared to OSI, we can say that the host-to-network layer is equivalent to the combination of the physical and data link layers. The internet layer is equivalent to the network layer, and the application layer is roughly doing the job of the session, presentation, and application layers with the transport layer in TCP/IP taking care of part of the duties of the session layer.

TCP/IP is a hierarchical protocol made up of interactive modules, each of which provides a specific functionality; however, the modules are not necessarily interdependent. Whereas the OSI model specifies which functions belong to each of its layers, the layers of the TCP/IP protocol suite contain relatively independent protocols that can be mixed and matched depending on the needs of the system. The term hierarchical means that each upper-level protocol is supported by one or more lower-level protocols.

At the transport layer, TCP/IP defines three protocols: Transmission Control Protocol (TCP), User Datagram Protocol (UDP), and Stream Control Transmission Protocol (SCTP). At the network layer, the main protocol defined by TCP/IP is the Internetworking Protocol (IP); there are also some other protocols that support data movement in this layer.

Figure 2.16 *TCP/IP and OSI model*



1. Host-to-Network 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.

2. Internet Layer:

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.

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.

3. 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.

4. 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). 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.

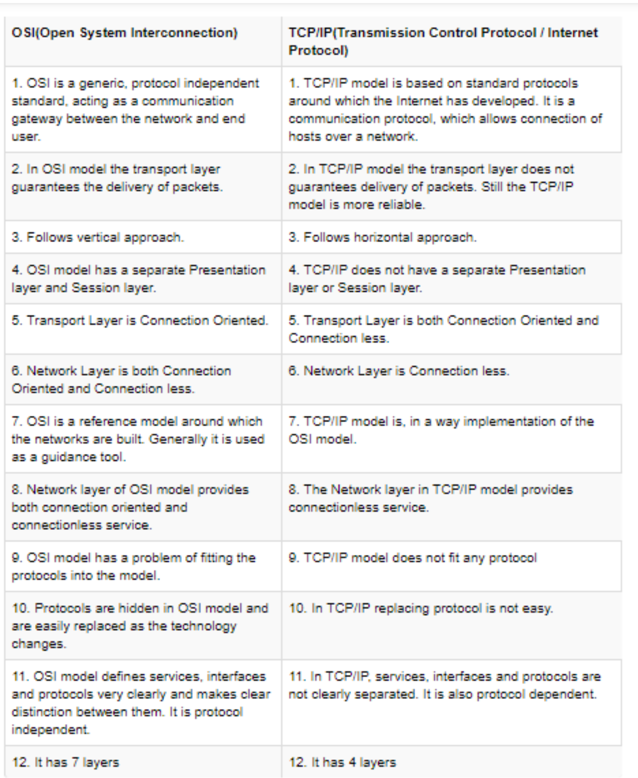
3. Protocols.

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.

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.



OSI(Open System Interconnection)	TCP/IP(Transmission Control Protocol / Internet Protocol)
1. OSI is a generic, protocol independent standard, acting as a communication gateway between the network and end user.	1. TCP/IP model is based on standard protocols around which the Internet has developed. It is a communication protocol, which allows connection of hosts over a network.
2. In OSI model the transport layer guarantees the delivery of packets.	2. In TCP/IP model the transport layer does not guarantees delivery of packets. Still the TCP/IP model is more reliable.
3. Follows vertical approach.	3. Follows horizontal approach.
4. OSI model has a separate Presentation layer and Session layer.	4. TCP/IP does not have a separate Presentation layer or Session layer.
5. Transport Layer is Connection Oriented.	5. Transport Layer is both Connection Oriented and Connection less.
6. Network Layer is both Connection Oriented and Connection less.	6. Network Layer is Connection less.
7. OSI is a reference model around which the networks are built. Generally it is used as a guidance tool.	7. TCP/IP model is, in a way implementation of the OSI model.
8. Network layer of OSI model provides both connection oriented and connectionless service.	8. The Network layer in TCP/IP model provides connectionless service.
9. OSI model has a problem of fitting the protocols into the model.	9. TCP/IP model does not fit any protocol
10. Protocols are hidden in OSI model and are easily replaced as the technology changes.	10. In TCP/IP replacing protocol is not easy.
11. OSI model defines services, interfaces and protocols very clearly and makes clear distinction between them. It is protocol independent.	11. In TCP/IP, services, interfaces and protocols are not clearly separated. It is also protocol dependent.
12. It has 7 layers	12. It has 4 layers

Critiques of OSI and TCP/IP Reference Model

Neither the OSI model and its protocols nor the TCP/IP model and its protocols are perfect. The criticism of the OSI model and its protocols can be summarized as:

- Bad Timing

- Bad Technology
- Bad Implementations
- Bad Politics

Bad timing

The competing TCP/IP protocols were already in widespread use by research universities by the time the OSI protocols appeared. While the billion-dollar wave of investment had not yet hit, the academic market was large enough that many vendors had begun cautiously offering TCP/IP products. When OSI came around, they did not want to support a second protocol stack until they were forced to, so there were no initial offerings. With every company waiting for every other company to go first, no company went first and OSI never happened.

Bad Technology

The second reason that OSI never caught on is that both the model and the protocols are flawed. The choice of seven layers was more political than technical, and two of the layers (session and presentation) are nearly empty, whereas two other ones (data link and network) are overfull.

The OSI model, along with its associated service definitions and protocols, is extraordinarily complex. They are also difficult to implement and inefficient in operation.

In addition to being incomprehensible, another problem with OSI is that some functions, such as addressing, flow control, and error control, reappear again and again in each layer.

Bad Implementations

Given the enormous complexity of the model and the protocols, it will come as no surprise that the initial implementations were huge, unwieldy, and slow. It did not take long for people to associate "OSI" with "poor quality". Although the products improved in the course of time, the image stuck.

In contrast, one of the first implementations of TCP/IP was part of Berkeley UNIX and was quite good. People began using it quickly, which led to a large user community, which led to improvements, which led to an even larger community. Here the spiral was upward instead of downward.

Bad Politics

On account of the initial implementation, many people, especially in academia, thought of TCP/IP as part of UNIX, and UNIX in the 1980s in academia was not unlike parenthood and apple pie.

OSI, on the other hand, was widely thought to be the creature of the European telecommunication ministries. The very idea of a bunch of government bureaucrats trying to shove a technically inferior standard down the throats of the poor researchers and programmers down in the trenches actually developing computer networks did not aid OSI's cause.

The TCP/IP model and protocols have their problems too. First, the model does not clearly distinguish the concepts of services, interfaces, and protocols. Good software engineering practice requires differentiating between the specification and the implementation, something that OSI does very carefully, but TCP/IP does not. Consequently, the TCP/IP model is not much of a guide for designing new networks using new technologies.

Second, the TCP/IP model is not at all general and is poorly suited to describing any protocol stack other than TCP/IP. Trying to use the TCP/IP model to describe Bluetooth, for example, is completely impossible.

Third, the link layer is not really a layer at all in the normal sense of the term as used in the context of layered protocols. It is an interface (between the network and data link layers). The distinction between an interface and layer is crucial and one should not be sloppy about it.

Fourth, the TCP/IP model does not distinguish between the physical and data link layers. These are completely different. The physical layer has to do with the transmission characteristics of copper wire, fiber optics, and wireless communication. The data link layer's job is to delimit the start and end of frames and get them from one side to the other with the desired degree of reliability. A proper model should include both as separate layers. The TCP/IP model does not do this

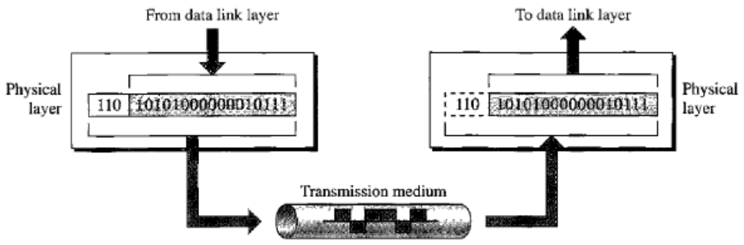
Finally, although the IP and TCP protocols were carefully thought out and well implemented, many of the other protocols were ad hoc, generally produced by a couple of graduate students hacking away until they got tired. The protocol implementations were then distributed free, which resulted in their becoming widely used, deeply entrenched, and thus hard to replace. Some of them are a bit of an embarrassment now.

Unit 2: The Physical Layer

Functions of Physical Layer:

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

Figure 2.5 *Physical layer*



The physical layer is responsible for movements of individual bits from one hop (node) to the next.

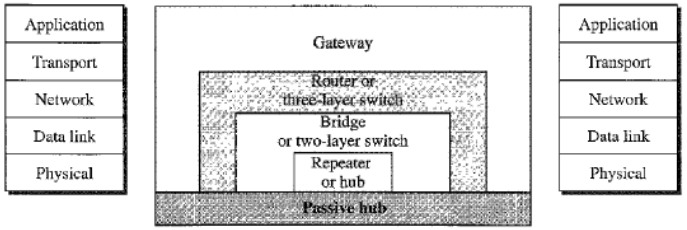
Specific responsibilities of phycial layer include:

- Physical characteristics of interfaces and medium
- Representation of bits
- Data rate
- Synchronization of bits
- Line configuration
- Physical topology
- Transmission mode

Network Devices (Connecting Devices):

We divide connecting devices into five different categories based on the layer in which they operate in a network, as shown in Figure below.

Figure 15.1 *Five categories of connecting devices*



The five categories contain devices which can be defined as:

1. Those which operate below the physical layer such as a passive hub.
2. Those which operate at the physical layer (a repeater or an active hub).
3. Those which operate at the physical and data link layers (a bridge or a two-layer switch).
4. Those which operate at the physical, data link, and network layers (a router or a three-layer switch).
5. Those which can operate at all five layers (a gateway).

Data and Signals:

Signal:

A signal is an electrical or electromagnetic current that is used for carrying data from one device or network to another. It is the key component behind data communication and networking. Signals can be periodic and nonperiodic. A periodic signal repeats the pattern over identical periods. A nonperiodic signal changes without repeating a pattern or cycle over time.

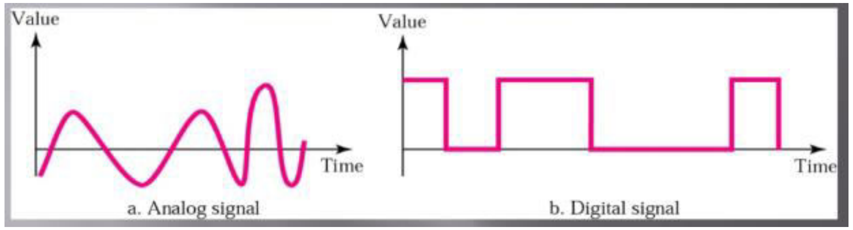
A signal can be either analog or digital.

Analog Signal:

Analog signal is a continuous wave that keeps on changing over a time period. In other words, an analog signal is a continuous wave denoted by a sine wave and may vary in signal strength (amplitude) or frequency (waves per unit time). Analog signals can be classified as simple or composite. A simple analog signal or sine wave cannot be further decomposed into simpler signals. A composite analog signal is composed of multiple sine waves.

Digital Signal:

Digital signals also carry information like analog signals but is somewhat is different from analog signals. Digital signal is noncontinuous, discrete time signal. Digital signal carries information or data in the binary form i.e. a digital signal represent information in the form of bits (0s and 1s). Digital signals are easier to transmit and are more reliable when compared to analog signals.



a. Analog signal

b. Digital signal

Key Differences Between Analog and Digital Signal:

- An analog signal represents a continuous wave that keeps changing over a time period. On the other hand, a digital signal represents a noncontinuous wave that carries information in a binary format and has discrete values.
- An analog signal is always represented by the continuous sine wave whereas, a digital signal is represented by square waves.
- While talking of analog signal, we describe the behavior of the wave in respect of amplitude, period or frequency, and phase of the wave. On the other hand, while talking of discrete signals we describe the behavior of the wave in respect of bit rate and bit interval.
- The range of an analog signal is not fixed whereas the range of the digital signal is finite and which can be 0 or 1.
- An analog signal is more prone to distortion in response to noise, but a digital signal has immunity in response to noise hence it rarely faces any distortion.
- An analog signal transmits data in the form of wave whereas, a digital signal transmits the data in the binary form i.e. in the form of bits.
- The best example of an analog signal is a human voice, and the best example of a digital signal is the transmission of data in a computer.

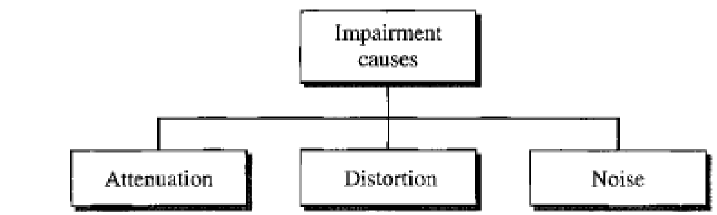
Analog Data:
The term analog data refers to information that is continuous; For example, an analog clock that has hour, minute, and second hands gives information in a continuous form; the movements of the hands are continuous. Analog data, such as the sounds made by a human voice, take on continuous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

Digital Data:
Digital data refers to information that has discrete states. For example, a digital clock that reports the hours and the minutes will change suddenly from 8:05 to 8:06. Digital data takes on discrete values. For example, data are stored in computer memory in the form of 0s and 1s. They can be converted to a digital signal or modulated into an analog signal for transmission across a medium.

Transmission Impairment:

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

Causes of impairment



Attenuation

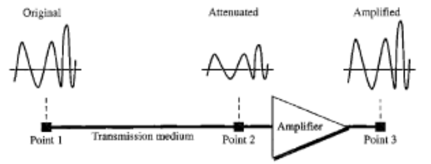
Attenuation means a loss of energy. When a signal, simple or composite, travels through a medium, it loses some of its energy in overcoming the resistance of the medium. That is why a wire carrying electric signals gets warm, if not hot, after a while. Some of the electrical energy in the signal is converted to heat. To compensate for this loss, amplifiers are used to amplify the signal. Attenuation is measured in terms of Decibels.

The decibel (dB) measures the relative strengths of two signals or one signal at two different points. Note that the decibel is negative if a signal is attenuated and positive if a signal is amplified.

$dB=10\log_{10} P2/P1$

Variables P1 and P2 are the powers of a signal at points 1 and 2, respectively.

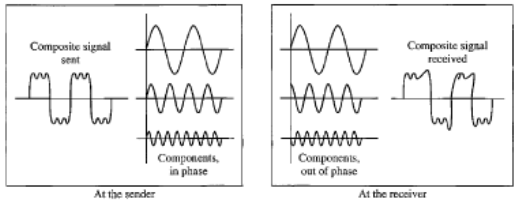
Figure 3.26 *Attenuation*



Distortion:

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

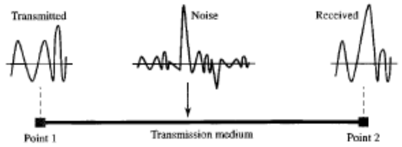
Figure 3.28 *Distortion*



Noise

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

Figure 3.29 *Noise*



Signal-to-Noise Ratio (SNR)

The signal-to-noise ratio is defined as

SNR= Average Signal power / Average Noise Power

SNR is actually the ratio of what is wanted (signal) to what is not wanted (noise). A high SNR means the signal is less corrupted by noise; a low SNR means the signal is more corrupted by noise. Because SNR is the ratio of two powers, it is often described in decibel units, SNR dB, defined as

$SNR_{dB} = 10\log_{10} SNR$

Data Rate Limits

Data rate can be defined as how fast can we send the data, in bits per second, over a channel. Maximum Data Rate (Channel Capacity) is the tight upper bound on the rate at which information can be reliably transmitted over a communication channel.

Data rate depends on three factors:

- The bandwidth available
- The level of the signals we use
- The quality of the channel (the level of noise)

There are two theoretical formulas to calculate the data rate:

- Nyquist for a noiseless channel
- Shannon for a noisy channel

Noiseless Channel: Nyquist Bit Rate

For a noiseless channel, the Nyquist bit rate formula defines the theoretical maximum bit rate

BitRate = 2 * Bandwidth * $\log_2$(L)

In the above equation, bandwidth is the bandwidth of the channel, L is the number of signal levels used to represent data, and BitRate is the bit rate in bits per second.

Bandwidth is a fixed quantity, so it cannot be changed. Hence, the data rate is directly proportional to the number of signal levels.

Increasing the levels of a signal may reduce the reliability of the system.

Noisy Channel: Shannon Capacity

In reality, we cannot have a noiseless channel; the channel is always noisy. Shannon capacity is used, to determine the theoretical highest data rate for a noisy channel:

Capacity = bandwidth * $\log_2$(1 + SNR)

In the above equation, bandwidth is the bandwidth of the channel, SNR is the signal-to-noise ratio, and capacity is the capacity of the channel in bits per second.

Bandwidth is a fixed quantity, so it cannot be changed. Hence, the channel capacity is directly proportional to the power of the signal, as SNR = (Power of signal) / (power of noise).

The signal-to-noise ratio (S/N) is usually expressed in decibels (dB) given by the formula:

SNR dB=10*$\log_{10}$ (SNR)

Performance

One important issue in networking is the performance of the network-how good it is? It can be referred as Quality of Service (QoS) that is an overall measurement of the network performance. There are four factors of determining network performance.

- Bandwidth
- Throughput

- Delay
- Jitter

Bandwidth

One characteristic that measures network performance is bandwidth. However, the term can be used in two different contexts with two different measuring values: bandwidth in hertz and bandwidth in bits per second.

- The first, bandwidth in hertz, refers to the range of frequencies in a composite signal or the range of frequencies that a channel can pass. For example, we can say the bandwidth of a subscriber telephone line is 4 kHz.
- The second, bandwidth in bits per second, refers to the speed of bit transmission in a channel or link. For example, one can say the bandwidth of a Fast Ethernet network (or the links in this network) is a maximum of 100 Mbps. This means that this network can send 100 Mbps.

Throughput

- The throughput is a measure of how fast we can actually send data through a network.
- Although, at first glance, bandwidth in bits per second and throughput seem the same, they are different.
- A link may have a bandwidth of B bps, but we can only send T bps through this link with T always less than B.
- In other words, the bandwidth is a potential measurement of a link; the throughput is an actual measurement of how fast we can send data.
- For example, we may have a link with a bandwidth of 1 Mbps, but the devices connected to the end of the link may handle only 200 kbps. This means that we cannot send more than 200 kbps through this link.

Delay (Latency)

- The latency or delay defines how long it takes for an entire message to completely arrive at the destination from the time the first bit is sent out from the source.
- We can say that latency is made of four components: propagation time, transmission time, queuing time and processing delay.

  Latency =propagation time +transmission time +queuing time + processing delay

- Propagation time measures the time required for a bit to travel from the source to the destination. The propagation time is calculated by dividing the distance by the propagation speed.

  Propagation Time= Distance / Propagation Speed

- In data communications we don't send just 1 bit, we send a message. The first bit may take a time equal to the propagation time to reach its destination; the last bit also may take the same amount of time. However, there is a time between the first bit leaving the sender and the last bit arriving at the receiver. The first bit leaves earlier and arrives earlier; the last bit leaves later and arrives later. The time required for transmission of a message depends on the size of the message and the bandwidth of the channel.

Transmission Time= Message Size / Bandwidth

- The third component in latency is the queuing time, the time needed for each intermediate or end device to hold the message before it can be processed. The queuing time is not a fixed factor; it changes with the load imposed on the network. When there is heavy traffic on the network, the queuing time increases. An intermediate device, such as a router, queues the arrived messages and processes them one by one. If there are many messages, each message will have to wait.

Jitter

- Another performance issue that is related to delay is jitter. We can roughly say that jitter is a problem if different packets of data encounter different delays and the application using the data at the receiver site is time-sensitive (audio and video data, for example).
- For Example, If the delay for the first packet is 20 ms, for the second is 45 ms, and for the third is 40 ms, then the real-time application that uses the packets endures jitter.

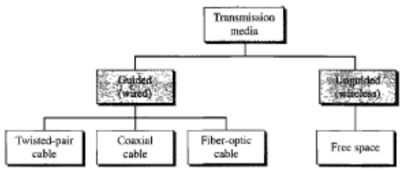
TRANSMISSION MEDIA:

A transmission medium can be broadly defined as anything that can carry information from a source to a destination. For example, the transmission medium for two people having a dinner conversation is the air. For a written message, the transmission medium might be a mail carrier, a truck, or an airplane.

In data communications, the definition of the information and the transmission medium is more specific. The transmission medium is usually free space, metallic cable, or fiber-optic cable. The information is usually a signal that is the result of a conversion of data from another form.

Classification of Transmission Media:

Figure 7.2 *Classes of transmission media*



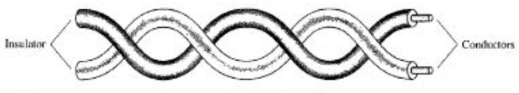
Guided Media

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

1. Twisted-Pair Cable

A twisted pair consists of two conductors (normally copper), each with its own plastic insulation, twisted together, as shown in Figure below.

Figure 7.3 *Twisted-pair cable*



One of the wires is used to carry signals to the receiver, and the other is used only as a ground reference. The receiver uses the difference between the two. In addition to the signal sent by the sender on one of the wires, interference (noise) and crosstalk may affect both wires and create unwanted signals. If the two wires are parallel, the effect of these unwanted signals is not the same in both wires because they are at different locations relative to the noise or crosstalk sources (e,g., one is closer and the other is farther). This results in a difference at the receiver.

By twisting the pairs, a balance is maintained. For example, suppose in one twist, one wire is closer to the noise source and the other is farther; in the next twist, the reverse is true. Twisting makes it probable that both wires are equally affected by external influences (noise or crosstalk). This means that the receiver, which calculates the difference between the two, receives no unwanted signals. The unwanted signals are mostly canceled out. From the above discussion, it is clear that the number of twists per unit of length (e.g., inch) has some effect on the quality of the cable.

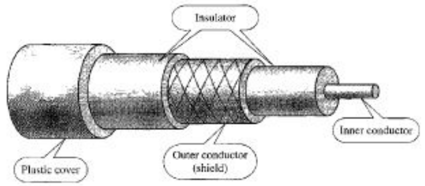
Applications

Twisted-pair cables are used in telephone lines to provide voice and data channels. The local loop-the line that connects subscribers to the central telephone office-commonly consists of unshielded twisted-pair cables. The DSL lines that are used by the telephone companies to provide high-data-rate connections also use the high-bandwidth capability of unshielded twisted-pair cables. Local-area networks also use twisted-pair cables.

2. Coaxial Cable

Coaxial cable (or coax) carries signals of higher frequency ranges than those in twisted pair cable, in part because the two media are constructed quite differently. Instead of having two wires, coax has a central core conductor of solid or stranded wire (usually copper) enclosed in an insulating sheath, which is, in turn, encased in an outer conductor of metal foil, braid, or a combination of the two. The outer metallic wrapping serves both as a shield against noise and as the second conductor, which completes the circuit. This outer conductor is also enclosed in an insulating sheath, and the whole cable is protected by a plastic cover (see Figure below).

Figure 7.7 *Coaxial cable*



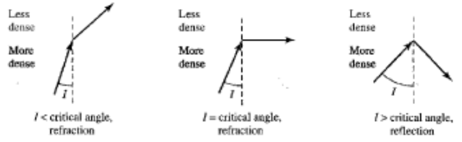
Applications

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

3. Fiber Optic Cable:

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

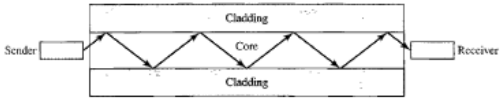
Figure 7.10 *Bending of light ray*


Less dense
More dense
I
I < critical angle, refraction
Less dense
More dense
I
I = critical angle, refraction
Less dense
More dense
I
I > critical angle, reflection

As the figure shows, if the angle of incidence I (the angle the ray makes with the line perpendicular to the interface between the two substances) is less than the critical angle, the ray refracts and moves closer to the surface. If the angle of incidence is equal to the critical angle, the light bends along the interface. If the angle is greater than the critical angle, the ray reflects (makes a turn) and travels again in the denser substance. Note that the critical angle is a property of the substance, and its value differs from one substance to another.

Optical fibers use reflection to guide light through a channel. A glass or plastic core is surrounded by a cladding of less dense glass or plastic. The difference in density of the two materials must be such that a beam of light moving through the core is reflected off the cladding instead of being refracted into it. See Figure below.

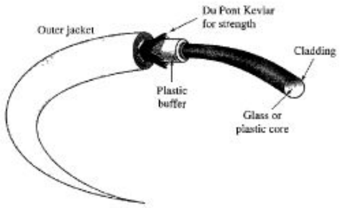
Figure 7.11 *Optical fiber*



Cable Composition

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

Figure 7.14 *Fiber construction*



Applications

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

Advantages and Disadvantages of Optical Fiber

Advantages

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

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

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

c. Immunity to electromagnetic interference. Electromagnetic noise cannot affect fiber-optic cables.

d. Resistance to corrosive materials. Glass is more resistant to corrosive materials than copper.

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

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

Disadvantages

There are some disadvantages in the use of optical fiber.

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

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

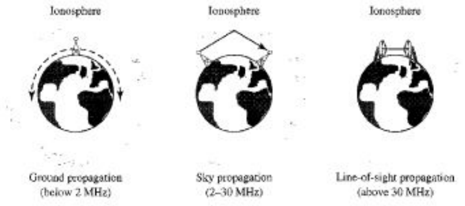
c. Cost. The cable and the interfaces are relatively more expensive than those of other guided media. If the demand for bandwidth is not high, often the use of optical fiber cannot be justified.

UNGUIDED MEDIA: WIRELESS

Unguided media transport electromagnetic waves without using a physical conductor. This type of communication is often referred to as wireless communication. Signals are normally broadcast through free space and thus are available to anyone who has a device capable of receiving them.

Unguided signals can travel from the source to destination in several ways: ground propagation, sky propagation, and line-of-sight propagation, as shown in Figure below.

Figure 7.18 *Propagation methods*


Ionosphere
Ionosphere
Ionosphere
Ground propagation (below 2 MHz)
Sky propagation (2–30 MHz)
Line-of-sight propagation (above 30 MHz)

In ground propagation, radio waves travel through the lowest portion of the atmosphere, hugging the earth. These low-frequency signals emanate in all directions from the transmitting antenna and follow the curvature of the planet. Distance depends on the amount of power in the signal: The greater the power, the greater the distance. In sky propagation, higher-frequency radio waves radiate upward into the ionosphere where they are reflected back to earth. This type of transmission allows for greater distances with lower output power. In line-or-sight propagation, very high-frequency signals are transmitted in straight lines directly from antenna to antenna. Antennas must be directional, facing each other, and either tall enough or close enough together not to be affected by the curvature of the earth. Line-of-sight propagation is tricky because radio transmissions cannot be completely focused.

1. Radio Waves

Waves ranging in frequencies between 3 kHz and 1 GHz are called radio waves. Radio waves, for the most part, are omnidirectional. When an antenna transmits radio waves, they are propagated in all directions. This means that the sending and receiving antennas do not have to be aligned. A sending antenna sends waves that can be received by any receiving antenna. The omnidirectional property has a disadvantage, too. The radio waves transmitted by one antenna are susceptible to interference by another antenna that may send signals using the same frequency or band. Radio waves, particularly those waves that propagate in the sky mode, can travel long distances. This makes radio waves a good candidate for long-distance broadcasting such as AM radio. Radio waves, particularly those of low and medium frequencies, can penetrate walls. This characteristic can be both an advantage and a disadvantage. It is an advantage because, for example, an AM radio can receive signals inside a building. It is a disadvantage because we cannot isolate a communication to just inside or outside a building. The radio wave band is relatively narrow, just under 1 GHz, compared to the microwave band. When this band is divided into sub bands, the sub bands are also narrow, leading to a low data rate for digital communications.

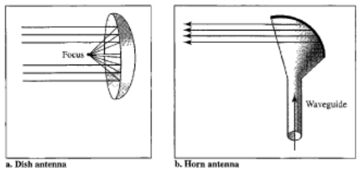
Figure 7.20 *Omnidirectional antenna*

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

2. Microwaves

Electromagnetic waves having frequencies between 1 and 300 GHz are called microwaves. Microwaves are unidirectional. When an antenna transmits microwave waves, they can be narrowly focused. This means that the sending and receiving antennas need to be aligned. The unidirectional property has an obvious advantage. A pair of antennas can be aligned without interfering with another pair of aligned antennas.

Figure 7.21 *Unidirectional antennas*



Microwaves need unidirectional antennas that send out signals in one direction. Two types of antennas are used for microwave communications: the parabolic dish and the horn (see Figure). A parabolic dish antenna is based on the geometry of a parabola: Every line parallel to the line of symmetry (line of sight) reflects off the curve at angles such that all the lines intersect in a common point called the focus. Outgoing transmissions are broadcast through a horn aimed at the dish. The microwaves hit the dish and are deflected outward in a reversal of the receipt path.

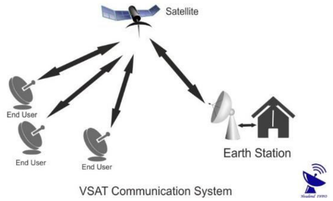
3. Infrared

Infrared waves, with frequencies from 300 GHz to 400 THz (wavelengths from 1 mm to 770 nm), can be used for short-range communication. Infrared waves, having high frequencies, cannot penetrate walls. This advantageous characteristic prevents interference between one system and another; a short-range communication system in one room cannot be affected by another system in the next room. When we use our infrared remote control, we do not interfere with the use of the remote by our neighbors. However, this same characteristic makes infrared signals useless for long-range communication. In addition, we cannot use infrared waves outside a building because the sun's rays contain infrared waves that can interfere with the communication.

The infrared band, almost 400 THz, has an excellent potential for data transmission. Such a wide bandwidth can be used to transmit digital data with a very high data rate.

VSAT:

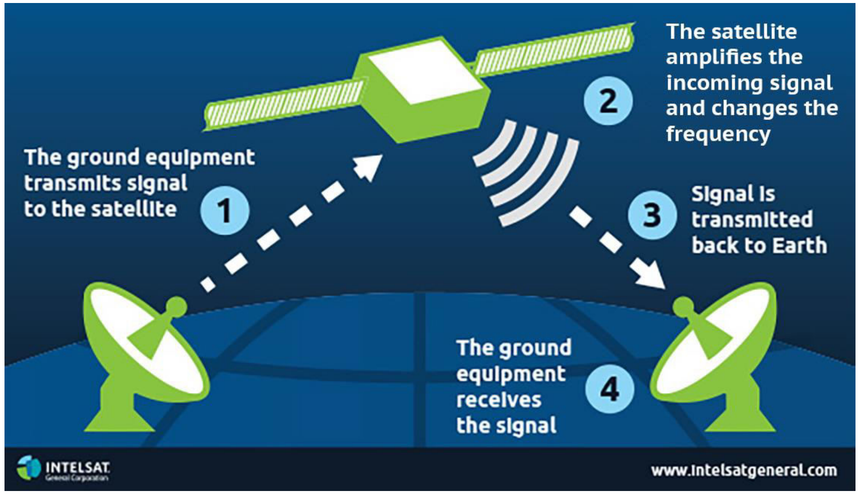
A very small aperture terminal (VSAT) is a small telecommunication earth station that receives and transmits real-time data via satellite. VSAT is a satellite communications system that serves home and business users. A VSAT end user needs a box that interfaces between the user's computer and an outside antenna with a transceiver. The tranceiver receives or sends a signal to a satellite transponder in the sky. The satellite sends and receives signals from an earth station computer that acts as a hub for the system. For one end user to communicate with another, each transmission has to first go to the hub station which retransmits it via the satellite to the other end user's VSAT.



VSAT is designed to serve both businesses and individuals and involves the use of specific technology and devices that are designed to facilitate effective telecommunications and Internet connectivity. When the system is comprised of multiple users, in order to establish communications with one another the data must be transmitted to the station-based PC which sends the signal to the sky satellite. The satellite sky transponder then forwards the data transmission to the end user’s VSAT antenna and finally to the end user’s device. VSAT can be used by both home users who sign up with a primary VSAT service and by private organizations and companies that lease or operate their own VSAT infrastructure. A main advantage of VSAT is it provides companies with complete control over their own communications infrastructure without having to depend upon third party sources.

Satellite:

A communications satellite is an artificial satellite that relays and amplifies radio telecommunications signals via a transponder; it creates a communication channel between a source transmitter and a receiver at different locations on Earth. Communications satellites are used for television, telephone, radio, internet, and military applications. The purpose of communications satellites is to relay the signal around the curve of the Earth allowing communication between widely separated geographical points.


The ground equipment transmits signal to the satellite
1
The satellite amplifies the incoming signal and changes the frequency
2
Signal is transmitted back to Earth
3
The ground equipment receives the signal
4
INTELSAT General Corporation
www.intelsatgeneral.com

Applications: Television, Internet, Military

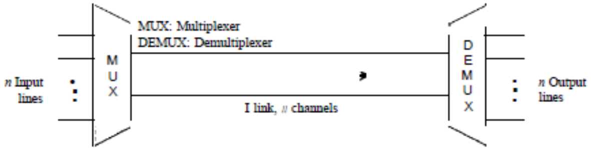
Bandwidth Utilization: Multiplexing and Spreading

- Bandwidth utilization is the wise use of available bandwidth to achieve specific goals.
- There are two broad categories of bandwidth utilization: multiplexing and spreading.
    - In multiplexing, our goal is efficiency; we combine several channels into one.
    - In spreading, our goals are privacy and antijamming; we expand the bandwidth of a channel.

Multiplexing

- Whenever the bandwidth of a medium linking two devices is greater than the bandwidth needs of the devices, the link can be shared.
- Multiplexing is the set of techniques that allows the simultaneous transmission of multiple signals across a single data link.
- As data and telecommunications use increases, so does traffic.
- We can accommodate this increase by continuing to add individual links each time a new channel is needed; or we can install higher-bandwidth links and use each to carry multiple signals.
- If the bandwidth of a link is greater than the bandwidth needs of the devices connected to it, the bandwidth is wasted.
- An efficient system maximizes the utilization of all resources; bandwidth is one of the most precious resources we have in data communications.
- In a multiplexed system, n lines share the bandwidth of one link.

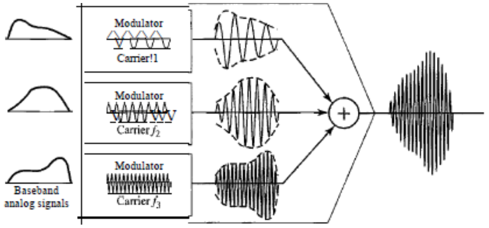
- MUX combines the streams into a single stream at the sender side whereas the DEMUX separates the streams back into its component transmissions.

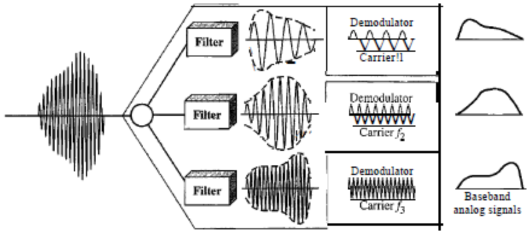


- There are three basic multiplexing techniques: frequency-division multiplexing, wavelength-division multiplexing, and time-division multiplexing.
- The first two are techniques designed for analog signals, the third, for digital signals.

Frequency-Division Multiplexing

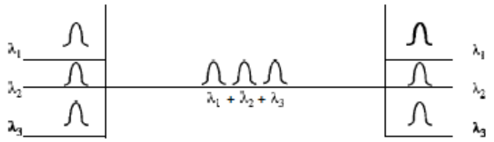
- Frequency-division multiplexing (FDM) is an analog technique that can be applied when the bandwidth of a link (in hertz) is greater than the combined bandwidths of the signals to be transmitted.
- In FDM, signals generated by each sending device modulate different carrier frequencies. These modulated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

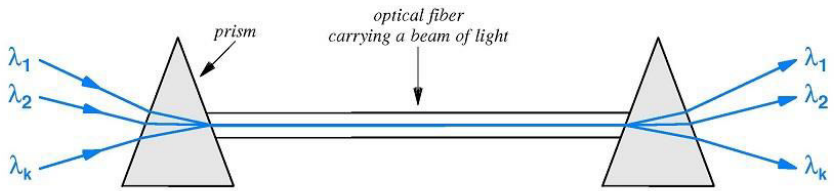




Wavelength-Division Multiplexing

- Wavelength-division multiplexing (WDM) is designed to use the high-data-rate capability of fiber-optic cable.
- The optical fiber data rate is higher than the data rate of metallic transmission cable. Using a fiber-optic cable for one single line wastes the available bandwidth.
- Multiplexing allows us to combine several lines into one.
- WDM is conceptually the same as FDM, except that the multiplexing and demultiplexing involve optical signals transmitted through fiber-optic channels.
- The idea is the same: We are combining different signals of different frequencies.
- The difference is that the frequencies are very high.



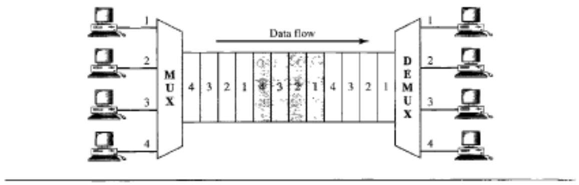


- Although WDM technology is very complex, the basic idea is very simple.
- We want to combine multiple light sources into one single light at the multiplexer and do the reverse at the demultiplexer. The combining and splitting of light sources are easily handled by a prism.
- A prism bends a beam of light based on the angle of incidence and the frequency.
- Using this technique, a multiplexer can be made to combine several input beams of light, each containing a narrow band of frequencies, into one output beam of a wider band of frequencies.
- A demultiplexer can also be made to reverse the process.

Time-Division Multiplexing (TDM)

- TDM is a digital process that allows several connections to share the high bandwidth of a line. Instead of sharing a portion of the bandwidth as in FDM, time is shared.
- Each connection occupies a portion of time in the link.
- Note that the same link is used as in FDM; here, however, the link is shown sectioned by time rather than by frequency.
- In the figure below, portions of signals 1,2,3, and 4 occupy the link sequentially.

Figure 6.12 *TDM*

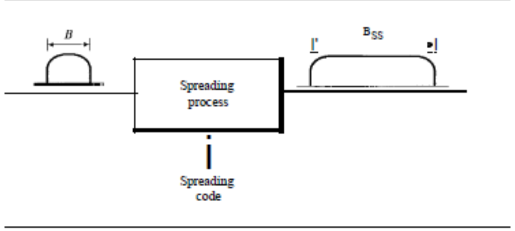


- Note that in Figure above, we are concerned with only multiplexing, not switching. This means that all the data in a message from source 1 always go to one specific destination, be it 1, 2, 3, or 4. The delivery is fixed and unvarying.

Spread Spectrum:

- Multiplexing combines signals from several sources to achieve bandwidth efficiency; the available bandwidth of a link is divided between the sources.
- In spread spectrum, we also combine signals from different sources to fit into a larger bandwidth, but our goals are somewhat different.
- Spread spectrum is designed to be used in wireless applications (LANs and WANs).
- In these types of applications, we have some concerns that outweigh bandwidth efficiency.
- In wireless applications, all stations use air (or a vacuum) as the medium for communication.
- Stations must be able to share this medium without interception by an eavesdropper and without being subject to jamming from a malicious intruder (in military operations, for example).
- To achieve these goals, spread spectrum techniques add redundancy; they spread the original spectrum needed for each station.
- If the required bandwidth for each station is B, spread spectrum expands it to B_{SS} such that B_{SS} » B.
- The expanded bandwidth allows the source to wrap its message in a protective envelope for a more secure transmission.

Spread spectrum



- Spread spectrum achieves its goals through two principles:

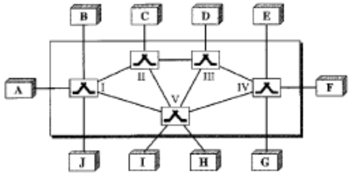
- The bandwidth allocated to each station needs to be, by far, larger than what is needed. This allows redundancy.
- The expanding of the original bandwidth B to the bandwidth B_{SS} must be done by a process that is independent of the original signal. In other words, the spreading process occurs after the signal is created by the source.

Switching:

A network is a set of connected devices. Whenever we have multiple devices, we have the problem of how to connect them to make one-to-one communication possible. One solution is to make a point-to-point connection between each pair of devices (Mesh topology) or between a central device and every other device (a star topology). These methods are not applicable for very large networks. Other topologies employing multipoint connections are also not efficient due to the distances between devices and the total number of devices increase beyond the capacities of the media and equipment.

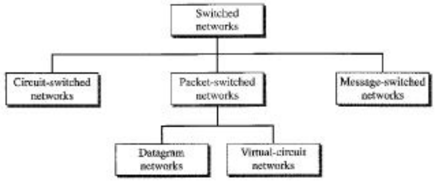
A better solution is **switching**. A switched network consists of a series of interlinked nodes, called switches. Switches are devices capable of creating temporary connections between two or more devices linked to the switch.

Switched network



We can divide today's network into three broad categories: circuit-switched networks, packet-switched networks and message-switched networks.

8.2 *Taxonomy of switched networks*

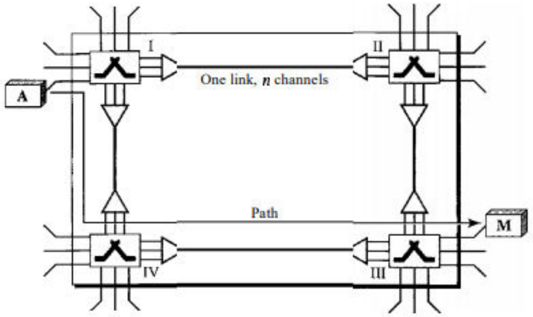


Circuit Switched Networks:

A circuit-switched network consists of a set of switches connected by physical links. A connection between two stations is a dedicated path made of one or more links. However, each connection uses only one dedicated channel on each link. Each link is normally divided into n channels by using FDM or TDM.

Figure below shows a trivial circuit-switched network with four switches and four links. Each link is divided into n (n is 3 in the figure) channels by using FDM or TDM.

A trivial circuit-switched network



We have explicitly shown the multiplexing symbols to emphasize the division of the link into channels even though multiplexing can be implicitly included in the switch fabric.

The end systems, such as computers or telephones, are directly connected to a switch. We have shown only two end systems for simplicity. When end system A needs to communicate with end system M, system A needs to request a connection to M that must be accepted by all switches as well as by M itself. This is called the setup phase; a circuit (channel) is reserved on each link, and the combination of circuits or channels defines the dedicated path. After the dedicated path made of connected circuits (channels) is established, data transfer can take place. After all data have been transferred, the circuits are torn down.

We need to emphasize several points here:

- Circuit switching takes place at the physical layer.
- Before starting communication, the stations must make a reservation for the resources to be used during the communication. These resources, such as channels (bandwidth in FDM and time slots in TDM), switch buffers, switch processing time, and switch input/output ports, must remain dedicated during the entire duration of data transfer until the teardown phase.
- Data transferred between the two stations are not packetized (physical layer transfer of the signal). The data are a continuous flow sent by the source station and received by the destination station, although there may be periods of silence.

- There is no addressing involved during data transfer. The switches route the data based on their occupied band (FDM) or time slot (TDM). Of course, there is end-to-end addressing used during the setup phase.

Three Phases:

The actual communication in a circuit-switched network requires three phases: connection setup, data transfer, and connection teardown.

Setup Phase

Before the two parties (or multiple parties in a conference call) can communicate, a dedicated circuit (combination of channels in links) needs to be established. The end systems are normally connected through dedicated lines to the switches, so connection setup means creating dedicated channels between the switches.

Data Transfer Phase

After the establishment of the dedicated circuit (channels), the two parties can transfer data.

Teardown Phase

When one of the parties needs to disconnect, a signal is sent to each switch to release the resources.

Packet Switched Networks:

A packet switched network (PSN) is a type of computer communications network that groups and sends data in the form of small packets. It enables the sending of data or network packets between a source and destination node over a network channel that is shared between multiple users and/or applications. A packet switched is also known as a connectionless network, as it does not create a permanent connection between a source and destination node. Packet-switched describes the type of network in which packets are routed through a network based on the destination address contained within each packet. Breaking communication down into packets allows the same data path to be shared among many users in the network.

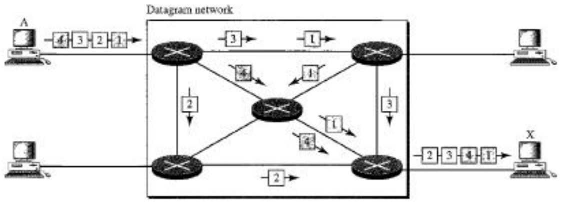
In packet switching, there is no resource allocation for a packet. This means that there is no reserved bandwidth on the links, and there is no scheduled processing time for each packet. Resources are allocated on demand. The allocation is done on a first-come, first serve basis. When a switch receives a packet, no matter what is the source or destination, the packet must wait if there are other packets being processed.

Packet switching may be classified into connectionless packet switching, also known as datagram switching, and connection-oriented packet switching, also known as virtual circuit switching.

Datagram Approach:

In a datagram network, each packet is treated independently of all others. Even if a packet is part of a multipacket transmission, the network treats it as though it existed alone. Packets in this approach are referred to as datagrams. Datagram switching is normally done at the network layer.

Figure 8.7 *A datagram network with four switches (routers)*

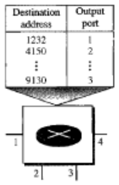


In this example, all four datagrams belong to the same message, but may travel different paths to reach their destination. This is so because the links may be involved in carrying packets from other sources and do not have the necessary bandwidth available to carry all the packets from A to X. This approach can cause the datagrams of a transmission to arrive at their destination out of order with different delays between the packets. Packets may also be lost or dropped because of a lack of resources.

The datagram networks are sometimes referred to as connectionless networks. The term connectionless here means that the switch does not keep information about the connection state. There are no setup and or teardown phases. Each packet is treated the same by a switch regardless of its source or destination.

Since there are no setup or teardown phases, each switch has a routing table which is based on the destination address. The routing tables are dynamic and are updated periodically. The destination address and the corresponding forwarding output ports are recorded in the tables. This is different from circuit-switched network in which each entry is created when the setup phase is completed and deleted when the teardown phase is over.

Routing table in a datagram network



Destination address	Output port
1232	1
4150	2
⋮	⋮
9130	3

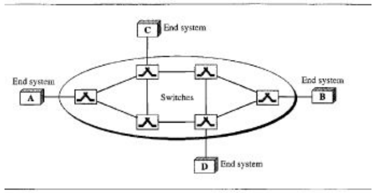
Virtual Circuit Networks:

A virtual circuit network is a cross between a circuit-switched network and a datagram network. It has some characteristics of both.

- As in a circuit-switched network, there are setup and teardown phases in addition to the data transfer phase.

- Resources can be allocated during the setup phase, as in a circuit-switched network, or on demand, as in a datagram network.
- As in a datagram network, data are packetized and each packet carries an address in the header. However, the address is of the next hop to be reached towards the destination.
- As in a circuit-switched network, all packets follow the same path established during the connection.
- A virtual-circuit network is normally implemented in the data link layer, while a circuit-switched network is implemented in the physical layer and a datagram network is implemented in the network layer.

10 Virtual-circuit network



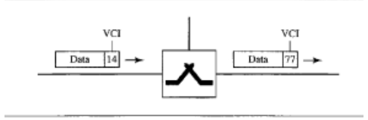
Addressing:

In a virtual-circuit network, two types of addressing are involved: global and local.

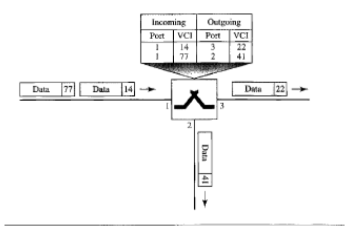
Global addressing: A source or a destination needs to have a global address-an address that can be unique in the scope of the network or internationally.

Local Addressing: The address that is actually used for data transfer is called the virtual-circuit identifier (VCI). A VCI, unlike a global address, is a smaller number that has only one switch scope; it is used by a frame between two switches.

Virtual-circuit identifier



Switch and tables in a virtual-circuit network



Three phases:

As in circuit-switched network, a source and destination need to go through three phases in a virtual circuit network: setup, data transfer, and teardown.

In the setup phase, the source and destination use their global addresses to help switches make table entries for the connection.

In the teardown phase, the source and destination inform the switches to delete the corresponding entry.

Data transfer occurs between these two phases.

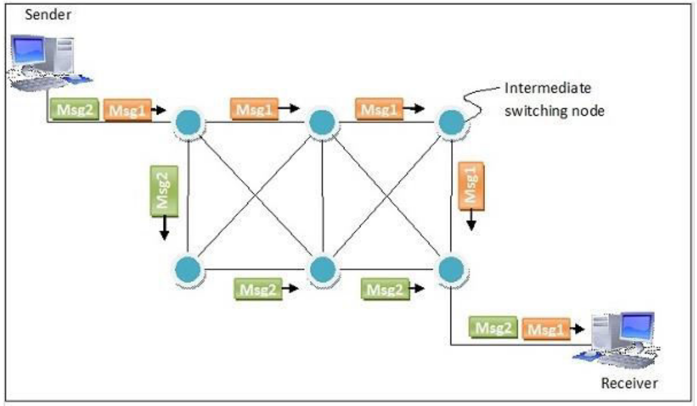
Message Switched Networks:

Message switching is a network switching technique in which data is routed in its entirety from the source node to the destination node, one hop at a time. During message routing, every intermediate switch in the network stores the whole message. If the entire network's resources are engaged or the network becomes blocked, the message-switched network stores and delays the message until ample resources become available for effective transmission of the message.

Before the advancements in packet switching, message switching acted as an efficient substitute for circuit switching.

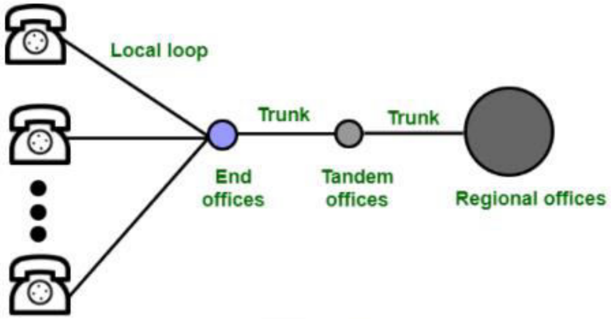
In message switching, the source and destination nodes are not directly connected. Instead, the intermediary nodes (mainly switches) are responsible for transferring the message from one node to the next. Thus, every intermediary node inside the network needs to store every message prior to retransferring the messages one-by-one as adequate resources become available. If the resources are not available, the messages are stored indefinitely. This characteristic is known as store and forward.

The following diagram represents routing of two separate messages from the same source to same destination via different routes, using message switching.



Telephone Network:

- Telephone Network is used to provide voice communication which uses Circuit Switching.
- Originally, the entire network was referred to as a plain old telephone system (POTS) which used analog signals.
- With the advancement of technology, i.e., in the computer era, there comes a feature to carry data in addition to voice. Today's network is both analogous and digital.
- The telephone network is made of three major components: local loops, trunks, and switching offices.
- The telephone network has several levels of switching offices such as end offices, tandem offices, and regional offices.



A telephone system

Local Loops

- One component of the telephone network is the local loop, a twisted-pair cable that connects the subscriber telephone to the nearest end office or local central office.
- The local loop, when used for voice, has a bandwidth of 4000 Hz (4 kHz). It is interesting to examine the telephone number associated with each local loop.
- The first three digits of a local telephone number define the office, and the next four digits define the local loop number.

Trunks

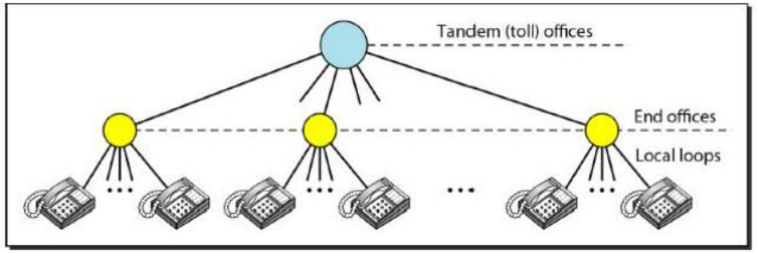
- Trunks are transmission media that handle the communication between offices.
- A trunk normally handles hundreds or thousands of connections through multiplexing.
- Transmission is usually through optical fibers or satellite links.

Switching Offices

- To avoid having a permanent physical link between any two subscribers, the telephone company has switches located in a switching office.
- A switch connects several local loops or trunks and allows a connection between different subscribers.

LATAs

- The Local-Access Transport Areas (LATAs) are the local telephone networks covering small or large metropolitan area.
- A small state or province may have one single LATA; a large state/province may have several LATAs. A LATA boundary may overlap the boundary of a state; part of a LATA can be in one state, part in another state.



- Communication inside a LATA is handled by end switches and tandem switches.
- A call that can be completed by using only end offices is considered toll-free. A call that has to go through a tandem office is charged.

Intra-LATA Services

The services offered by the common carriers (telephone companies) inside a LATA are called intra-LATA services. The carrier that handles these services is called a local exchange carrier (LEC).

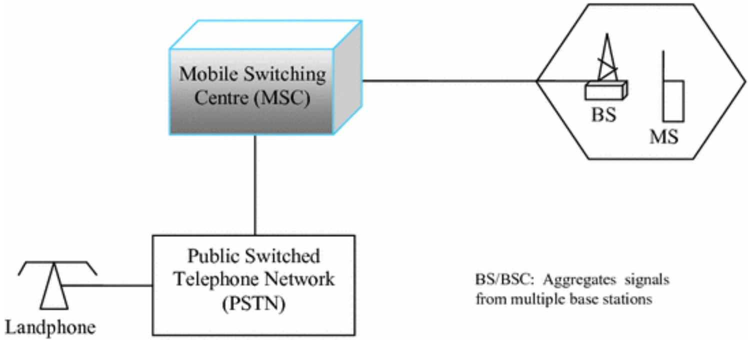
Inter-LATA Services

The services between LATAs are handled by interexchange carriers (IXCs). These carriers, sometimes called long-distance companies, provide communication services between two customers in different LATAs.

Mobile Networks:

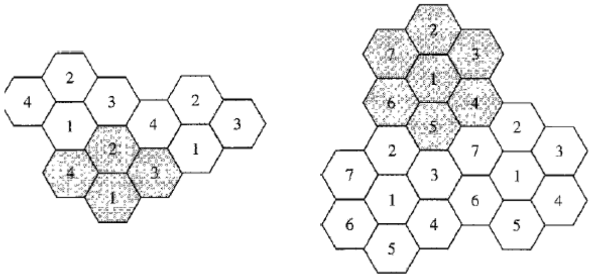
- Mobile Networks or Cellular networks are high-speed, high-capacity voice and data communication networks with enhanced multimedia and seamless roaming capabilities for supporting cellular devices (wireless end devices).
- With the increase in popularity of cellular devices, these networks are used for more than just entertainment and phone calls.
- Cellular telephony is designed to provide communications between two moving units, called mobile stations (MSs), or between one mobile unit and one stationary unit, often called a land unit.
- A service provider must be able to locate and track a caller, assign a channel to the call, and transfer the channel from base station to base station as the caller moves out of range.
- To make this tracking possible, each cellular service area is divided into small regions called cells.

- Each cell contains an antenna and is controlled by a solar or AC powered network station, called the base station (BS).
- Each base station, in turn, is controlled by a switching office, called a mobile switching center (MSC).
- The MSC coordinates communication between all the base stations and the telephone central office.
- It is a computerized center that is responsible for connecting calls, recording call information, and billing.


Mobile Switching Centre (MSC)
BS
MS
Public Switched Telephone Network (PSTN)
Landphone
BS/BSC: Aggregates signals from multiple base stations

- In general, neighboring cells cannot use the same set of frequencies for communication because it may create interference for the users located near the cell boundaries.
- However, the set of frequencies available is limited, and frequencies need to be reused.
- A frequency reuse pattern is a configuration of N cells, N being the reuse factor, in which each cell uses a unique set of frequencies.
- When the pattern is repeated, the frequencies can be reused. There are several different patterns.

Frequency reuse patterns



a. Reuse factor of 4

b. Reuse factor of 7

Transmitting

- To place a call from a mobile station, the caller enters a code of 7 or 10 digits (a phone number) and presses the send button.
- The mobile station then scans the band, seeking a setup channel with a strong signal, and sends the data (phone number) to the closest base station using that channel.
- The base station relays the data to the MSC. The MSC sends the data on to the telephone central office.
- If the called party is available, a connection is made and the result is relayed back to the MSC.
- At this point, the MSC assigns an unused voice channel to the call, and a connection is established.
- The mobile station automatically adjusts its tuning to the new channel, and communication can begin.

Receiving

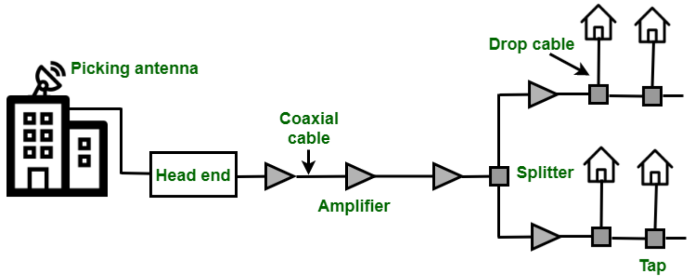
- When a mobile phone is called, the telephone central office sends the number to the MSC.
- The MSC searches for the location of the mobile station by sending query signals to each cell in a process called paging.
- Once the mobile station is found, the MSC transmits a ringing signal and, when the mobile station answers, assigns a voice channel to the call, allowing voice communication to begin.

Cable Networks:

- The cable TV network started as a video service provider, but it has moved to the business of Internet access.

Traditional Cable Networks:

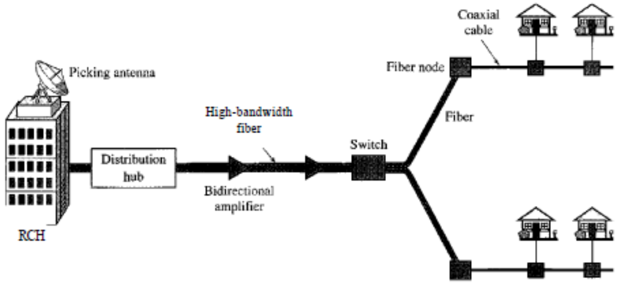
- Cable TV started to distribute broadcast video signals to locations with poor or no reception in the late 1940s.
- It was called community antenna TV (CATV) because an antenna at the top of a tall hill or building received the signals from the TV stations and distributed them, via coaxial cables, to the community.



- The cable TV office, called the head end, receives video signals from broadcasting stations and feeds the signals into coaxial cables.
- The signals became weaker and weaker with distance, so amplifiers were installed through the network to renew the signals. There could be up to 35 amplifiers between the head end and the subscriber premises.
- At the other end, splitters split the cable, and taps and drop cables make the connections to the subscriber premises.
- The traditional cable TV system used coaxial cable end to end.
- Due to attenuation of the signals and the use of a large number of amplifiers, communication in the traditional network was unidirectional (one-way).
- Video signals were transmitted downstream, from the head end to the subscriber premises.

Hybrid Fiber-Coaxial Network:

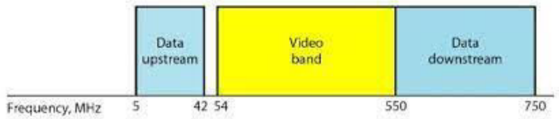
- Hybrid Fiber-Coaxial Network is that the second generation of the cable network. which is a combination of fiber-optic and coaxial cable is used in this type of network.
- The transmission mode is used is fiber node i.e., fiber mode. The schematic diagram of the HFC network is as follows:



- There are nearly 400, 000 subscribers served by Regional Cable Head (RCH). The RCHs feed the distribution hubs, each of which serves up to 40,000 subscribers. The distribution hub plays an important role in the new infrastructure.
- Modulation and demodulation of the signal are done through the distribution hubs after these signals are sent to the fiber nodes through fiber-optic cables.
- The fiber node split the analog signal so that the same signal is sent to each coaxial cable. Approx. 1000 subscribers are served by coaxial cable.
- The use of fiber-optic cable reduces the need for amplifiers down to eight or less.
- One reason for moving from traditional to hybrid infrastructure is to make the cable network bidirectional (two-way).

Cable TV for Data Transfer:

- Cable companies are now competing with telephone companies for the residential customer who wants high-speed data transfer.
- DSL technology provides high-data-rate connections for residential subscribers over the local loop. However, DSL uses the existing unshielded twisted-pair cable, which is very susceptible to interference. This imposes an upper limit on the data rate.
- Another solution is the use of the cable TV network.
- Even in an HFC system, the last part of the network, from the fiber node to the subscriber premises, is still a coaxial cable.
- This coaxial cable has a bandwidth that ranges from 5 to 750 MHz (approximately).
- To provide Internet access, the cable company has divided this bandwidth into three bands: video, downstream data, and upstream data.



Downstream Video Band:

- The downstream video band occupies frequencies from 54 to 550 MHZ. Since each TV channel occupies 6 MHz, this can accommodate more than 80 channels.

Downstream Data Band:

- The downstream data (from the Internet to the subscriber premises) occupies the upper band, from 550 to 750 MHZ. This band is also divided into 6-MHz channels.

Upstream Data Band:

- The upstream data (from the subscriber premises to the Internet) occupies the lower band, from 5 to 42 MHZ. This band is also divided into 6-MHz channels.

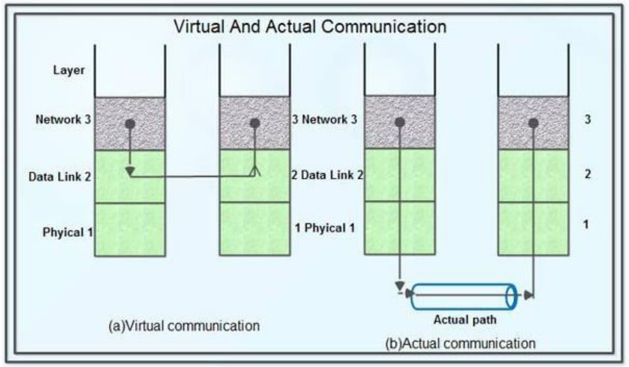
Unit 3: The Data Link Layer

Functions of Data Link Layer:

- The data link layer transforms the physical layer, a raw transmission facility, to a link responsible for node-to-node (hop-to-hop) communication.
- Specific responsibilities of the data link layer include framing, addressing, flow control, error control, and media access control.
- The data link layer divides the stream of bits received from the network layer into manageable data units called frames. The data link layer adds a header to the frame to define the addresses of the sender and receiver of the frame.
- If the rate at which the data are absorbed by the receiver is less than the rate at which data are produced in the sender, the data link layer imposes a flow control mechanism to avoid overwhelming the receiver.
- The data link layer also adds reliability to the physical layer by adding mechanisms to detect and retransmit damaged, duplicate, or lost frames.
- When two or more devices are connected to the same link, data link layer protocols are necessary to determine which device has control over the link at any given time.

Services Provided to Network Layer:

- Data link layer provides several services to the network layer. The one of the major services provided is the transferring the data from network layer on the source machine to the network layer on destination machine.
- On source machine data link layer receives the data from network layer and on destination machine pass on this data to the network layer as shown in Figure. The path shown in fig (a) is the virtual path.
- But the actual path is Network layer -> Data link layer -> Physical layer on source machine, then to physical media and thereafter physical layer -> Data link layer -> Network layer on destination machine.

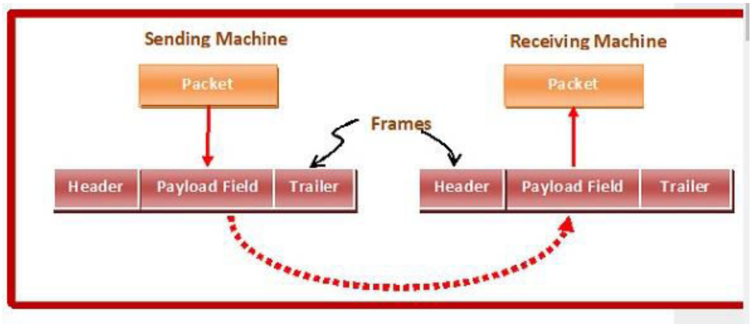


The three major types of services offered by data link layer are:

1. Unacknowledged connectionless service.
2. Acknowledged connectionless service.
3. Acknowledged connection-oriented service.

Framing:

- In the physical layer, data transmission involves synchronized transmission of bits from the source to the destination. The data link layer packs these bits into frames.
- Data-link layer takes the packets from the Network Layer and encapsulates them into frames.
- If the frame size becomes too large, then the packet may be divided into small sized frames.
- Smaller sized frames make flow control and error control more efficient. Then, it sends each frame bit-by-bit on the hardware.
- At receiver's end, data link layer picks up signals from hardware and assembles them into frames.



Parts of a Frame

A frame has the following parts −

Frame Header − It contains the source and the destination addresses of the frame.

Payload field − It contains the message to be delivered.

Trailer − It contains the error detection and error correction bits.

Flag − It marks the beginning and end of the frame.

Types of Framing

Framing can be of two types, fixed sized framing and variable sized framing.

Fixed-sized Framing

Here the size of the frame is fixed and so the frame length acts as delimiter of the frame. Consequently, it does not require additional boundary bits to identify the start and end of the frame.

Example − ATM cells.

Variable – Sized Framing

Here, the size of each frame to be transmitted may be different. So additional mechanisms are kept to mark the end of one frame and the beginning of the next frame.

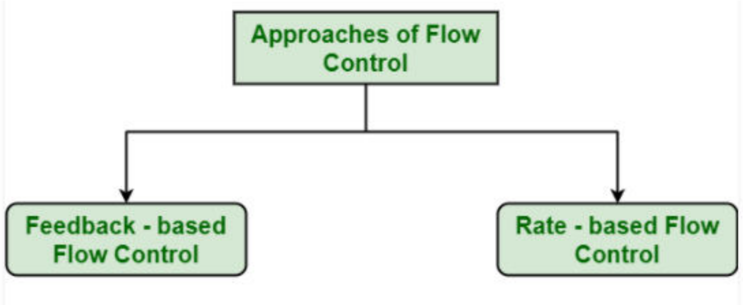
It is used in local area networks.

Flow Control:

- Data link layer protocols are mainly responsible for flow control. When a data frame is sent from one host to another over a single medium, it is required that the sender and receiver should work at the same speed.
- That is, sender sends at a speed on which the receiver can process and accept the data. If sender is sending too fast, the receiver may be overloaded and data may be lost.
- Flow control is basically technique that gives permission to two of stations that are working and processing at different speeds to just communicate with one another.
- Flow control in Data Link Layer simply restricts and coordinates number of frames or amount of data sender can send just before it waits for an acknowledgment from receiver.
- Flow control is actually set of procedures that explains sender about how much data or frames it can transfer or transmit before data overwhelms receiver.

Approaches to Flow Control:

Flow Control is classified into two categories:


Approaches of Flow Control
Feedback - based Flow Control
Rate - based Flow Control

Feedback-based Flow Control:

In these protocols, the sender sends frames after it has received acknowledgments from the user. This is used in the data link layer.

Rate-based Flow Control:

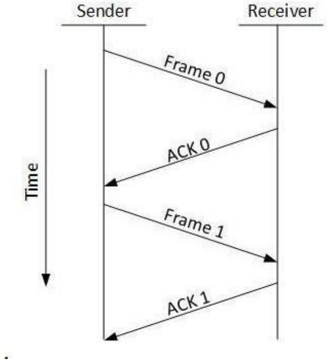
These protocols have built in mechanisms to restrict the rate of transmission of data without requiring acknowledgment from the receiver. This is used in the transport layer.

Two types of mechanisms can be deployed to control the flow based on the feedback:

- A simple stop and wait Protocol
- Sliding Window Protocol

Simplex Stop and Wait

- This flow control mechanism forces the sender after transmitting a data frame to stop and wait until the acknowledgement of the data-frame sent is received.
- The sender sends the next frame only when it has received a positive acknowledgement from the receiver that it is available for further data processing.
- Data transmission is one directional, but must have bidirectional line.



Sliding Window

- In this flow control mechanism, both sender and receiver agree on the number of data-frames after which the acknowledgement should be sent.
- As we learnt, stop and wait flow control mechanism wastes resources, this protocol tries to make use of underlying resources as much as possible.
- In this protocol, multiple frames can be sent by a sender at a time before receiving an acknowledgment from the receiver.
- This protocol improves the efficiency of stop and wait protocol by allowing multiple frames to be transmitted before receiving an acknowledgment.
- The working principle of this protocol can be described as follows –
- Both the sender and the receiver have finite sized buffers called windows. The sender and the receiver agree upon the number of frames to be sent based upon the buffer size.
- The sender sends multiple frames in a sequence, without waiting for acknowledgment. When its sending window is filled, it waits for acknowledgment. On receiving acknowledgment, it advances the window and transmits the next frames, according to the number of acknowledgments received.

Error Control:

- Error control in data link layer is the process of detecting and correcting data frames that have been corrupted or lost during transmission.
- In case of lost or corrupted frames, the receiver does not receive the correct data-frame and sender is ignorant about the loss.
- Data link layer follows a technique to detect transit errors and take necessary actions, which is retransmission of frames whenever error is detected or frame is lost. The process is called Automatic Repeat Request (ARQ).

<u>Phases in Error Control</u>

The error control mechanism in data link layer involves the following phases –

- Detection of Error: Transmission error, if any, is detected by either the sender or the receiver.
- Acknowledgment: acknowledgment may be positive or negative.
    - Positive ACK − On receiving a correct frame, the receiver sends a positive acknowledge.
    - Negative ACK − On receiving a damaged frame or a duplicate frame, the receiver sends a negative acknowledgment back to the sender.
- Retransmission: The sender maintains a clock and sets a timeout period. If an acknowledgment of a data-frame previously transmitted does not arrive before the timeout, or a negative acknowledgment is received, the sender retransmits the frame.

<u>Error Control Techniques</u>

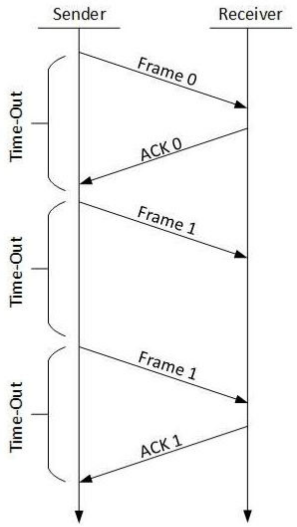
Error Control Techniques in data link layer are:

- Stop and Wait ARQ
- Sliding Window ARQ
    - Go-Back-N ARQ
    - Selective Repeat ARQ

Stop and Wait ARQ:

The following transition may occur in Stop-and-Wait ARQ:

- The sender maintains a timeout counter.
- When a frame is sent, the sender starts the timeout counter.
- If acknowledgement of frame comes in time, the sender transmits the next frame in queue.
- If acknowledgement does not come in time, the sender assumes that either the frame or its acknowledgement is lost in transit. Sender retransmits the frame and starts the timeout counter.
- If a negative acknowledgement is received, the sender retransmits the frame.



Sliding Window ARQ:

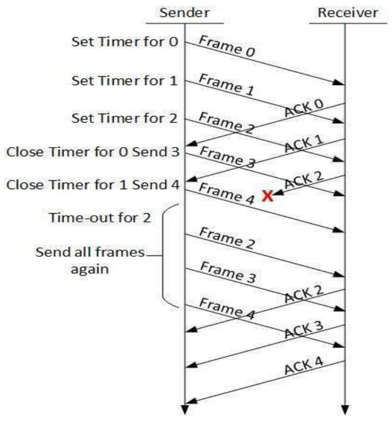
This technique is generally used for continuous transmission error control. It is further categorized into two categories as given below:

<u>Go-Back-N ARQ:</u>

In this protocol, we can send several frames before receiving acknowledgements; we keep a copy of these frames until the acknowledgements arrive. Stop and wait mechanism does not utilize the resources at their best. When the acknowledgement is received, the sender sits idle and does nothing. In Go-Back-N method, both sender and receiver maintain a window.

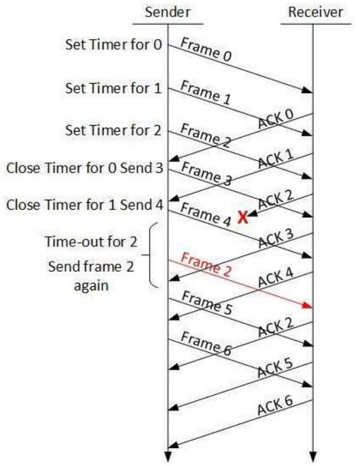
- The sending-window size enables the sender to send multiple frames without receiving the acknowledgement of the previous ones.
- The receiving-window enables the receiver to receive multiple frames and acknowledge them. The receiver keeps track of incoming frame’s sequence number.

When the sender sends all the frames in window, it checks up to what sequence number it has received positive acknowledgement. If all frames are positively acknowledged, the sender sends next set of frames. If sender finds that it has received NACK or has not receive any ACK for a particular frame, it retransmits all the frames after which it does not receive any positive ACK.



Selective Repeat ARQ:

- In Go-back-N ARQ, it is assumed that the receiver does not have any buffer space for its window size and has to process each frame as it comes.
- This enforces the sender to retransmit all the frames which are not acknowledged.
- In Selective-Repeat, the receiver while keeping track of sequence numbers, buffers the frames in memory and sends NACK for only frame which is missing or damaged.
- The sender in this case, sends only packet for which NACK is received.



Error Detection and Correction Techniques:

- Networks must be able to transfer data from one device to another with acceptable accuracy.
- For most applications, a system must guarantee that the data received are identical to the data transmitted.
- Any time data are transmitted from one node to the next, they can become corrupted in passage. Many factors can alter one or more bits of a message.
- Some applications require a mechanism for detecting and correcting errors.
- Some applications can tolerate a small level of error. For example, random errors in audio or video transmissions may be tolerable, but when we transfer text, we expect a very high level of accuracy.

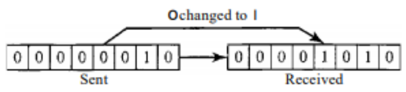
Types of Errors:

- Whenever bits flow from one point to another, they are subject to unpredictable changes because of interference. This interference can change the shape of the signal.
- In a single-bit error, a 0 is changed to a 1 or a 1 to a 0.
- In a burst error, multiple bits are changed.

Single-Bit Error:

- The term single-bit error means that only 1 bit of a given data unit (such as a byte, character, or packet) is changed from 1 to 0 or from 0 to 1.

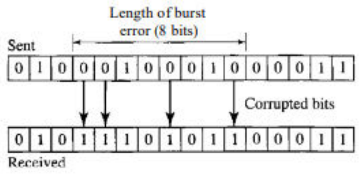
Single-bit error



Burst Error:

- The term burst error means that 2 or more bits in the data unit have changed from 1 to 0 or from 0 to 1.

Burst error oflength 8



- A burst error is more likely to occur than a single-bit error.

- The duration of noise is normally longer than the duration of 1 bit, which means that when noise affects data, it affects a set of bits.
- The number of bits affected depends on the data rate and duration of noise.

Redundancy

- The central concept in detecting or correcting errors is redundancy. To be able to detect or correct errors, we need to send some extra bits with our data.
- These redundant bits are added by the sender and removed by the receiver. Their presence allows the receiver to detect or correct corrupted bits.

Detection Versus Correction

- The correction of errors is more difficult than the detection.
- In error detection, we are looking only to see if any error has occurred.
- The answer is a simple yes or no. We are not even interested in the number of errors. A single-bit error is the same for us as a burst error.
- In error correction, we need to know the exact number of bits that are corrupted and more importantly, their location in the message.
- The number of the errors and the size of the message are important factors.
- If we need to correct one single error in an 8-bit data unit, we need to consider eight possible error locations; if we need to correct two errors in a data unit of the same size, we need to consider 28 possibilities.

Forward Error Correction Versus Retransmission

- There are two main methods of error correction. Forward error correction is the process in which the receiver tries to guess the message by using redundant bits. This is possible if the number of errors is small.
- Correction by retransmission is a technique in which the receiver detects the occurrence of an error and asks the sender to resend the message. Resending is repeated until a message arrives that the receiver believes is error-free.

Error Detecting Codes:

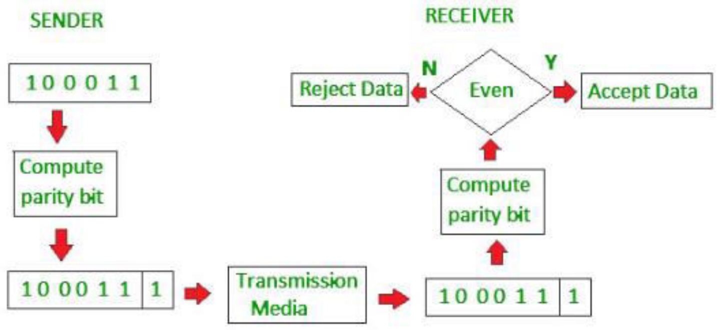
- Whenever a message is transmitted, it may get scrambled by noise or data may get corrupted.
- To avoid this, we use error-detecting codes which are additional data added to a given digital message to help us detect if any error has occurred during transmission of the message.

Some popular techniques for error detection are:

- Parity
- Checksum
- Cyclic redundancy check

Parity check

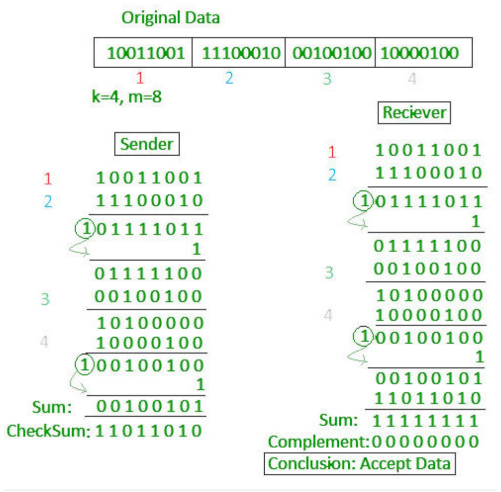
- The most common and least expensive mechanism for error- detection is the parity check. The parity check is done by adding an extra bit, called parity bit to the data to make a number of 1s either even in case of even parity or odd in case of odd parity. While creating a frame, the sender counts the number of 1s in it and adds the parity bit in the following way:
- In case of even parity: If a number of 1s is even then parity bit value is 0. If the number of 1s is odd then parity bit value is 1.
- In case of odd parity: If a number of 1s is odd then parity bit value is 0. If a number of 1s is even then parity bit value is 1.
- On receiving a frame, the receiver counts the number of 1s in it. In case of even parity check, if the count of 1s is even, the frame is accepted, otherwise, it is rejected. A similar rule is adopted for odd parity check.
- The parity check is suitable for single bit error detection only.



Checksum

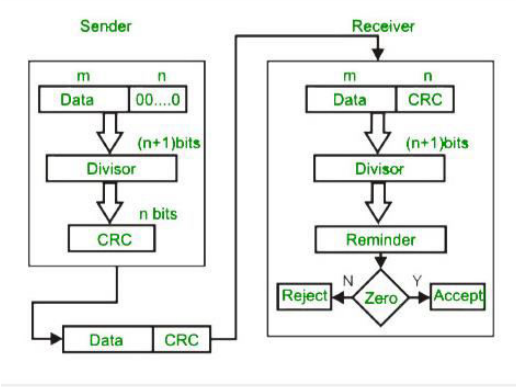
In this error detection scheme, the following procedure is applied:

- Data is divided into fixed sized frames or segments. (k segments each of m bits)
- The sender adds the segments using 1's complement arithmetic to get the sum. It then complements the sum to get the checksum and sends it along with the data frames.
- The receiver adds the incoming segments along with the checksum using 1's complement arithmetic to get the sum and then complements it.
- If the result is zero, the received frames are accepted; otherwise, they are discarded.

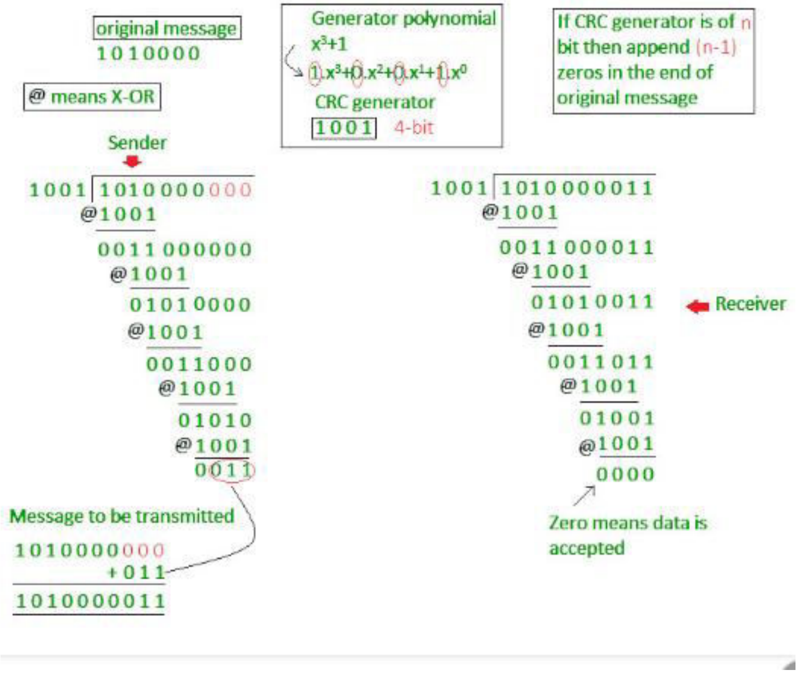


Cyclic Redundancy Check:

A cyclic redundancy check (CRC) is an error-detecting code commonly used in digital networks and storage devices to detect accidental changes to raw data. 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.



- At the sender side, the data unit to be transmitted is divided by a predetermined divisor (binary number) in order to obtain the remainder. This remainder is called CRC.
- The CRC has one bit less than the divisor. It means that if CRC is of n bits, divisor is of n+ 1 bit.
- The sender appends this CRC to the end of data unit such that the resulting data unit becomes exactly divisible by predetermined divisor i.e., remainder becomes zero.
- At the destination, the incoming data unit i.e., data + CRC is divided by the same number (predetermined binary divisor).
- If the remainder after division is zero then there is no error in the data unit & receiver accepts it.
- If remainder after division is not zero, it indicates that the data unit has been damaged in transit and therefore it is rejected.
- This technique is more powerful than the parity check and checksum error detection.



High Level Data Link Control (HDLC):

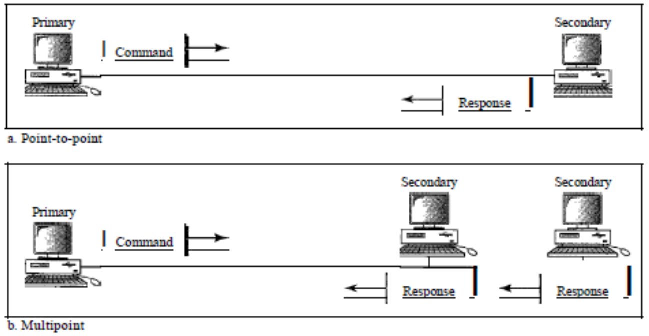
High-level Data Link Control (HDLC) is a group of communication protocols of the data link layer for transmitting data between network points or nodes. Since it is a data link protocol, data is organized into frames. A frame is transmitted via the network to the destination that verifies its successful arrival. It is a bit - oriented protocol that is applicable for both point - to - point and multipoint communications.

Configurations and Transfer Modes

HDLC provides two common transfer modes that can be used in different configurations: normal response mode (NRM) and asynchronous balanced mode (ABM).

Normal Response Mode

- In normal response mode (NRM), the station configuration is unbalanced. We have one primary station and multiple secondary stations.
- A primary station can send commands; a secondary station can only respond. The NRM is used for both point-to-point and multiple-point links.

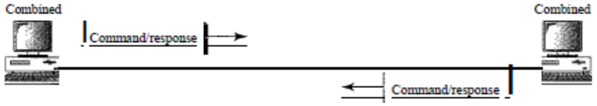


a. Point-to-point

b. Multipoint

Asynchronous Balanced Mode

- In asynchronous balanced mode (ABM), the configuration is balanced.
- The link is point-to-point, and each station can function as a primary and a secondary (acting as peers). This is the common mode today.

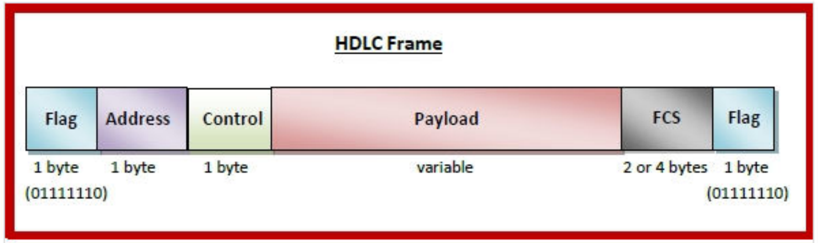


HDLC Frame

HDLC is a bit - oriented protocol where each frame contains up to six fields. The structure varies according to the type of frame. The fields of a HDLC frame are −

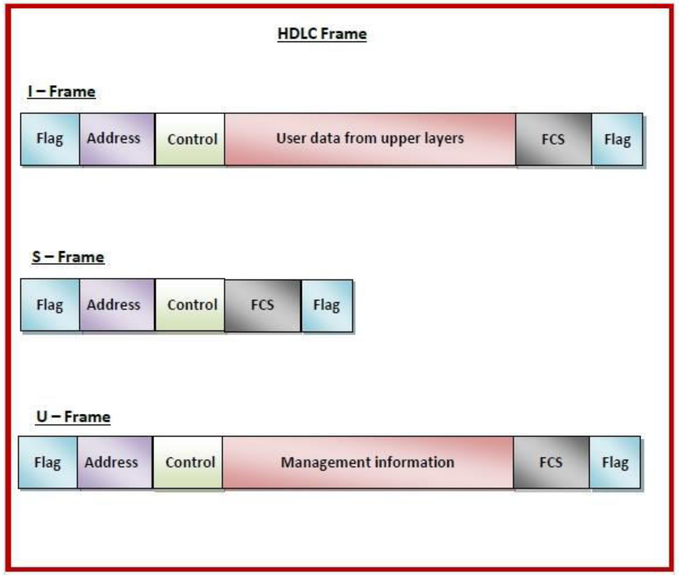
- Flag − It is an 8-bit sequence that marks the beginning and the end of the frame. The bit pattern of the flag is 01111110.

- Address − It contains the address of the receiver. If the frame is sent by the primary station, it contains the address(es) of the secondary station(s). If it is sent by the secondary station, it contains the address of the primary station. The address field may be from 1 byte to several bytes.
- Control − It is 1- or 2-bytes containing flow and error control information.
- Payload − This carries the data from the network layer. Its length may vary from one network to another.
- FCS − It is a 2 byte or 4 bytes frame check sequence for error detection. The standard code used is CRC (cyclic redundancy code)



Types of HDLC Frames

There are three types of HDLC frames. The type of frame is determined by the control field of the frame:



- **I-frame** – I-frames or Information frames carry user data from the network layer. They also include flow and error control information that is piggybacked on user data. The first bit of control field of I-frame is 0.
- **S-frame** – S-frames or Supervisory frames do not contain information field. They are used for flow and error control when piggybacking is not required. The first two bits of control field of S-frame is 10.
- **U-frame** – U-frames or Un-numbered frames are used for myriad miscellaneous functions, like link management. It may contain an information field, if required. The first two bits of control field of U-frame is 11.

Point-to-Point Protocol (PPP):

- Although HDLC is a general protocol that can be used for both point-to-point and multipoint configurations, one of the most common protocols for point-to-point access is the Point-to-Point Protocol (PPP).
- Today, millions of Internet users who need to connect their home computers to the server of an Internet service provider use PPP.
- The majority of these users have a traditional modem; they are connected to the Internet through a telephone line, which provides the services of the physical layer.
- But to control and manage the transfer of data, there is a need for a point-to-point protocol at the data link layer. PPP is by far the most common.

PPP provides several services:

- PPP defines the format of the frame to be exchanged between devices.
- PPP defines how two devices can negotiate the establishment of the link and the exchange of data.
- PPP defines how network layer data are encapsulated in the data link frame.
- PPP defines how two devices can authenticate each other.
- PPP provides multiple network layer services supporting a variety of network layer protocols.
- PPP provides connections over multiple links.
- PPP provides network address configuration. This is particularly useful when a home user needs a temporary network address to connect to the Internet.

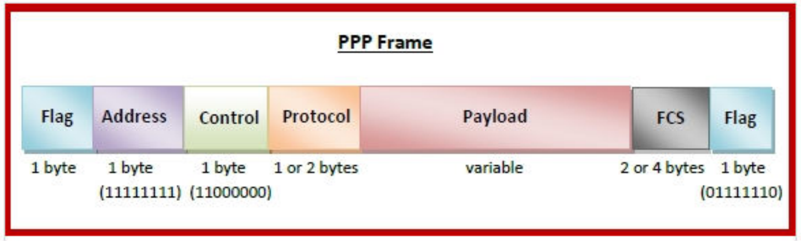
On the other hand, to keep PPP simple, several services are missing:

- PPP does not provide flow control. A sender can send several frames one after another with no concern about overwhelming the receiver.
- PPP has a very simple mechanism for error control. A CRC field is used to detect errors. If the frame is corrupted, it is silently discarded; the upper-layer protocol needs to take care of the problem. Lack of error control and sequence numbering may cause a packet to be received out of order.
- PPP does not provide a sophisticated addressing mechanism to handle frames in a multipoint configuration.

PPP Frame

PPP is a byte - oriented protocol where each field of the frame is composed of one or more bytes. The fields of a PPP frame are:

- Flag – 1 byte that marks the beginning and the end of the frame. The bit pattern of the flag is 01111110.
- Address – 1 byte which is set to 11111111 in case of broadcast.
- Control – 1 byte set to a constant value of 11000000.
- Protocol – 1 or 2 bytes that define the type of data contained in the payload field.
- Payload – This carries the data from the network layer. The maximum length of the payload field is 1500 bytes. However, this may be negotiated between the endpoints of communication.
- FCS – It is a 2 byte or 4 bytes frame check sequence for error detection. The standard code used is CRC (cyclic redundancy code)



Byte Stuffing in PPP Frame:

- Byte stuffing is used is PPP payload field whenever the flag sequence appears in the message, so that the receiver does not consider it as the end of the frame.
- The escape byte, 01111101, is stuffed before every byte that contains the same byte as the flag byte or the escape byte.
- The receiver on receiving the message removes the escape byte before passing it onto the network layer.

Channel Allocation Problem:

- When there are more than one user who desire to access a shared network channel, an algorithm is deployed for channel allocation among the competing users.
- The network channel may be a single cable or optical fiber connecting multiple nodes, or a portion of the wireless spectrum.
- Channel allocation algorithms allocate the wired channels and bandwidths to the users, who may be base stations, access points or terminal equipment.

Channel Allocation Schemes

Channel Allocation may be done using two schemes:

- Static Channel Allocation
- Dynamic Channel Allocation

Static Channel Allocation

- In static channel allocation scheme, a fixed portion of the frequency channel is allotted to each user.
- For N competing users, the bandwidth is divided into N channels using frequency division multiplexing (FDM), and each portion is assigned to one user.
- This scheme is also referred as fixed channel allocation or fixed channel assignment.
- In this allocation scheme, there is no interference between the users since each user is assigned a fixed channel.
- However, it is not suitable in case of a large number of users with variable bandwidth requirements.

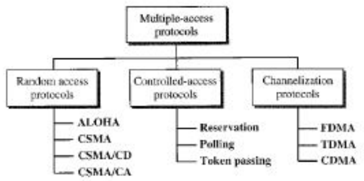
Dynamic Channel Allocation

- In dynamic channel allocation scheme, frequency bands are not permanently assigned to the users.
- Instead, channels are allotted to users dynamically as needed, from a central pool.
- The allocation is done considering a number of parameters so that transmission interference is minimized.
- This allocation scheme optimizes bandwidth usage and results is faster transmissions.
- Dynamic channel allocation is further divided into centralized and distributed allocation.

Multiple Access Protocols:

- When nodes or stations are connected and use a common link, called a multipoint or broadcast link, we need a multiple access protocol to coordinate access to the link.
- Many formal protocols have been devised to handle access to a shared link. We categorize them into three groups.

Taxonomy of multiple-access protocols discussed in this chapter



Random access:

- In random access or contention methods, no station is superior to another station and none is assigned the control over another.
- No station permits or does not permit another station to send.
- At each instance, a station that has to send uses a procedure defined by the protocol to make a decision on whether or not to send.
- This decision depends on the state of the medium (idle or busy). In other words, each station can transmit when it desires on testing of the state of the medium.

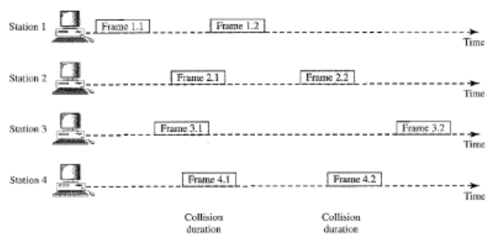
ALOHA:

- ALOHA is the earliest random-access method developed for wireless LAN but can be used on any shared medium.
- In this, multiple stations can transmit data at the same time and can hence lead to collision. The data from the two stations collide.

Pure ALOHA:

- The idea behind this protocol is that each station sends a frame whenever it has a frame to send.
- However, since there is only one channel to share, there is possibility of collision between frames from different stations.
- Even if one bit of a frame coexists on the channel with one bit from another frame, there is a collision and both will be destroyed.

Figure 12.3 *Frames in a pure ALOHA network*



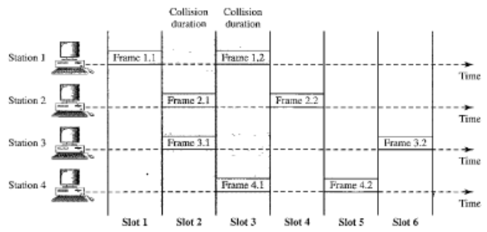
- The pure ALOHA protocol relies on acknowledgements from the receiver. When a station sends a frame, it expects the receiver to send an acknowledgement.
- If the acknowledgement does not arrive after the time out period, the station assumes that the frame (or the acknowledgement) has been destroyed and resends the frame.
- A collision involves two or more stations. If all these stations try to resend their frames after the time out period passes, each station waits a random amount of time before resending its frame.
- The randomness will help avoid more collisions, called back-off time.

- Since different stations may wait for different amount of time, the probability of further collision decreases.

Slotted ALOHA:

- In pure ALOHA, there is no rule that defines when the station can send.
- A station may send soon after another station has started or soon before another station has finished. So, still the collision may occur.
- Slotted ALOHA is similar to pure ALOHA, except that we divide time into slots and sending of data is allowed only at the beginning of these slots.
- If a station misses out the allowed time, it must wait for the next slot. This reduces the probability of collision.

Figure 12.6 *Frames in a slotted ALOHA network*



- Allowing a station to send only at the beginning of the time slot means that the station sending in the previous slot has finished sending its frame already.
- However, there is still possibility of collision if two stations try to send at the beginning of the same time slot.

Carrier Sense Multiple Access (CSMA):

- The chance of collision can be reduced if a station senses the medium before trying to use it.
- Carrier Sense Multiple Access ensures fewer collisions as the station is required to first sense the medium (for idle or busy) before transmitting data.
- If it is idle then it sends data, otherwise it waits till the channel becomes idle. (Listen before talk)
- However, there is still chance of collision in CSMA due to propagation delay. For example, if station A wants to send data, it will first sense the medium.
- If it finds the channel idle, it will start sending data. However, by the time the first bit of data is transmitted (delayed due to propagation delay) from station A, if station B requests to send data and senses the medium it will also find it idle and will also send data.
- This will result in collision of data from station A and B.

CSMA access modes-

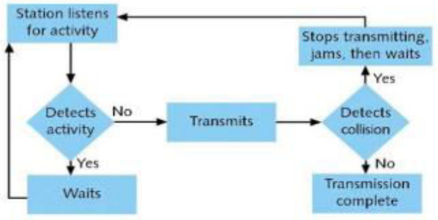
- 1-persistent: The node senses the channel, if idle it sends the data, otherwise it continuously keeps on checking the medium for being idle and transmits unconditionally (with 1 probability) as soon as the channel gets idle.

- Non-Persistent: The node senses the channel, if idle it sends the data, otherwise it checks the medium after a random amount of time (not continuously) and transmits when found idle.
- P-persistent: The node senses the medium, if idle it sends the data with p probability. If the data is not transmitted ((1-p) probability) then it waits for some time and checks the medium again, now if it is found idle then it sends with p probability. This repeat continues until the frame is sent. It is used in Wi-Fi and packet radio systems.
- O-persistent: Superiority of nodes is decided beforehand and transmission occurs in that order. If the medium is idle, node waits for its time slot to send data.

Carrier sense multiple access with collision detection (CSMA/CD):

- The CSMA method does not specify the procedure following a collision.
- In Carrier sense multiple access with collision detection method, a station monitors the medium after it sends a frame to see if the transmission was successful. If so, the transmission is completed. However, if there is a collision, the frame is sent again.

The basic idea behind CSMA/CD is that a station needs to be able to receive while transmitting, to detect a collision. When there is no collision, the station receives one signal; its own signal. When there is a collision, the station receives two signals: its own signal and the signal transmitted by a second station. To distinguish between these two cases, the received signals in these two cases must be significantly different. In other words, the signal from the second station needs to add a significant amount of energy to the one created by the first station.

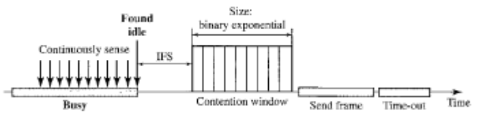


Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA):

- The process of collision detection involves sender receiving acknowledgement signals.
- If there is just one signal (its own), then the data is successfully sent but if there are two signals (its own and the one with which it has collided), then it means a collision has occurred.
- To distinguish between these two cases, collision must have a lot of impact on received signal. The second signal adds significant amount of energy to the first signal.
- However, this applies only to the wired networks since the received signal has almost the same energy as the sent signal.
- In wireless networks, much of the sent energy is lost in transmission. The received signal has very little energy.

- Therefore, a collision may add only 5 to 10 percent additional energy. This is not useful for effective collision detection.
- We need to avoid collisions on wireless networks because they cannot be detected. Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) was invented for this network.
- In contrast to the Carrier Sense Multiple Access/Collision Detect (CSMA/CD) protocol, which handles transmissions only after a collision has taken place, CSMA/CA works to avoid collisions prior to their occurrence.
- Collisions are avoided through the use of CSMA/CA's three strategies as shown in figure below.

Figure 12.16 *Timing in CSMA/CA*



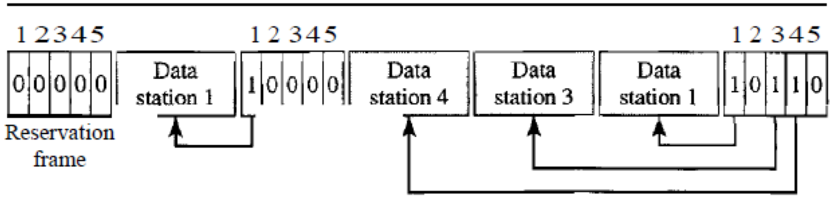
- Interframe space – Station waits for medium to become idle and if found idle it does not immediately send data (to avoid collision due to propagation delay) rather it waits for a period of time called Interframe space or IFS. After this time, it again checks the medium for being idle. IFS can also be used to define the priority of a station or a frame. Higher the IFS lower is the priority.
- Contention Window –It is the amount of time divided into slots. A station which is ready to send frames chooses random number of slots as wait time.
- Acknowledgement – The positive acknowledgements and time-out timer can help guarantee a successful transmission of the frame.

Controlled Access:

- In controlled access, the stations consult one another to find which station has the right to send. A station cannot send unless it has been authorized by other stations.
- We discuss three popular controlled-access methods: Reservation, Polling & Token Passing

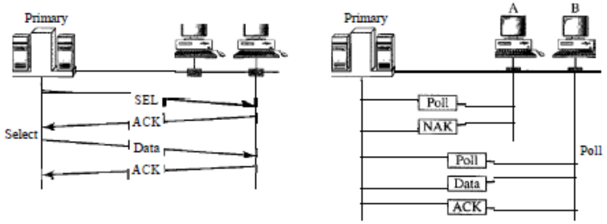
Reservation

- In the reservation method, a station needs to make a reservation before sending data.
- Time is divided into intervals. In each interval, a reservation frame precedes the data frames sent in that interval.
- If there are N stations in the system, there are exactly N reservation minislots in the reservation frame. Each minislot belongs to a station.
- When a station needs to send a data frame, it makes a reservation in its own minislot.
- The stations that have made reservations can send their data frames after the reservation frame.
- Figure below shows a situation with five stations and a five-minislot reservation frame.
- In the first interval, only stations 1, 3, and 4 have made reservations. In the second interval, only station 1 has made a reservation.



Polling:

- Polling works with topologies in which one device is designated as a primary station and the other devices are secondary stations.
- All data exchanges must be made through the primary device even when the ultimate destination is a secondary device.
- The primary device controls the link; the secondary devices follow its instructions. It is up to the primary device to determine which device is allowed to use the channel at a given time.
- The primary device, therefore, is always the initiator of a session.
- If the primary wants to receive data, it asks the secondaries if they have anything to send; this is called poll function.
- If the primary wants to send data, it tells the secondary to get ready to receive; this is called select function.



Select

- The select function is used whenever the primary device has something to send. Remember that the primary controls the link.
- If the primary is neither sending nor receiving data, it knows the link is available. If it has something to send, the primary device sends it.
- What it does not know, however, is whether the target device is prepared to receive. So, the primary must alert the secondary to the upcoming transmission and wait for an acknowledgment of the secondary's ready status.
- Before sending data, the primary creates and transmits a select (SEL) frame, one field of which includes the address of the intended secondary.

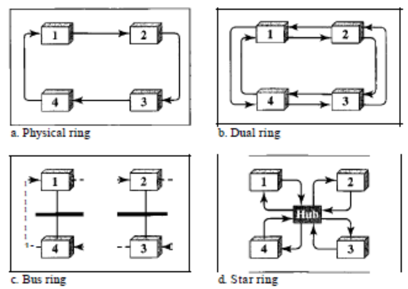
Poll

- The poll function is used by the primary device to solicit transmissions from the secondary devices.
- When the primary is ready to receive data, it must ask (poll) each device in turn if it has anything to send.
- When the first secondary is approached, it responds either with a NAK frame if it has nothing to send or with data (in the form of a data frame) if it does.
- If the response is negative (a NAK frame), then the primary polls the next secondary in the same manner until it finds one with data to send.
- When the response is positive (a data frame), the primary reads the frame and returns an acknowledgment (ACK frame), verifying its receipt.

Token Passing

- In the token-passing method, the stations in a network are organized in a logical ring. In other words, for each station, there is a predecessor and a successor.
- The predecessor is the station which is logically before the station in the ring; the successor is the station which is after the station in the ring.
- The current station is the one that is accessing the channel now. The right to this access has been passed from the predecessor to the current station.
- The right will be passed to the successor when the current station has no more data to send.
- But how is the right to access the channel passed from one station to another? In this method, a special packet called a token circulates through the ring.
- The possession of the token gives the station the right to access the channel and send its data.
- When a station has some data to send, it waits until it receives the token from its predecessor.
- It then holds the token and sends its data. When the station has no more data to send, it releases the token, passing it to the next logical station in the ring.
- The station cannot send data until it receives the token again in the next round.
- In this process, when a station receives the token and has no data to send, it just passes the data to the next station.

Logical ring and physical topology in token-passing access method



a. Physical ring

b. Dual ring

c. Bus ring

d. Star ring

Channelization:

- Channelization is a multiple-access method in which the available bandwidth of a link is shared in time, frequency, or through code, between different stations.
- There are three channelization protocols: FDMA, TDMA, and CDMA.

Frequency-Division Multiple Access (FDMA)

- In frequency-division multiple access (FDMA), the available bandwidth is divided into frequency bands. Each station is allocated a band to send its data.
- In other words, each band is reserved for a specific station, and it belongs to the station all the time.
- Each station also uses a bandpass filter to confine the transmitter frequencies. To prevent station interferences, the allocated bands are separated from one another by small guard bands.
- FDMA specifies a predetermined frequency band for the entire period of communication. This means that stream data (a continuous flow of data that may not be packetized) can easily be used with FDMA.
- We need to emphasize that although FDMA and FDM (Frequency Division Multiplexing) conceptually seem similar, there are differences between them.
- FDM is a physical layer technique that combines the loads from low-bandwidth channels and transmits them by using a high-bandwidth channel. FDMA, on the other hand, is an access method in the data link layer.

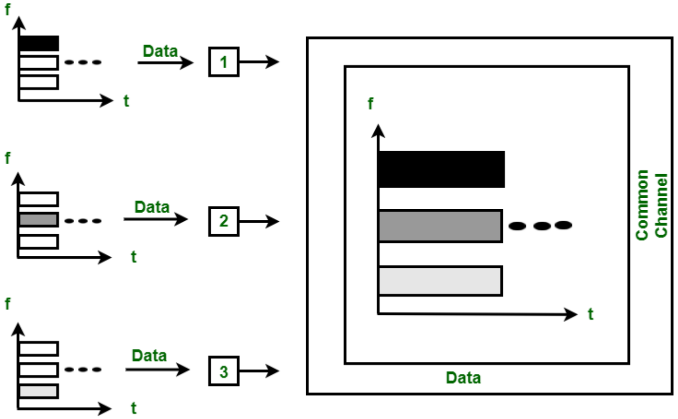


Fig: Frequency-Division Multiple Access

Time-Division Multiple Access (TDMA)

- In time-division multiple access (TDMA), the stations share the bandwidth of the channel in time.
- Each station is allocated a time slot during which it can send data. Each station transmits its data in is assigned time slot.
- The main problem with TDMA lies in achieving synchronization between the different stations. Each station needs to know the beginning of its slot and the location of its slot.
- This may be difficult because of propagation delays introduced in the system if the stations are spread over a large area. To compensate for the delays, we can insert guard times.
- Synchronization is normally accomplished by having some synchronization bits (normally referred to as preamble bits) at the beginning of each slot.
- We also need to emphasize that although TDMA and TDM conceptually seem the same, there are differences between them.
- TDM is a physical layer technique that combines the data from slower channels and transmits them by using a faster channel.
- The process uses a physical multiplexer that interleaves data units from each channel.
- TDMA, on the other hand, is an access method in the data link layer. The data link layer in each station tells its physical layer to use the allocated time slot. There is no physical multiplexer at the physical layer.

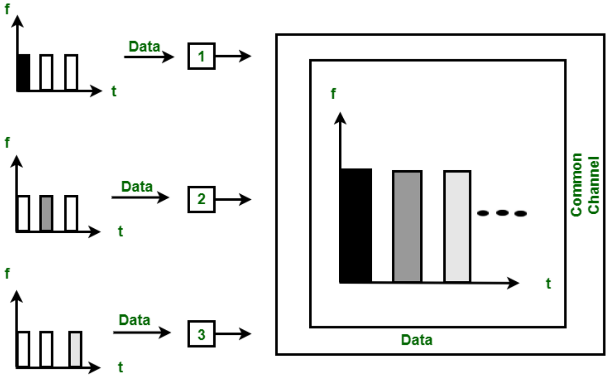


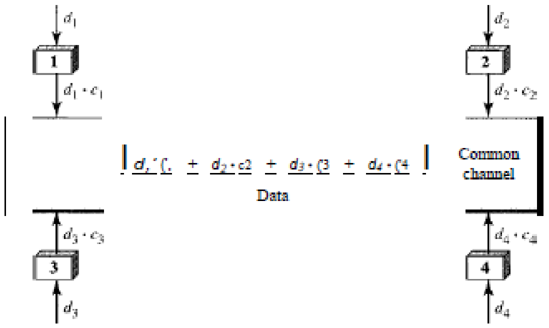
Fig: Time Division Multiple Access

Code Division Multiple Access (CDMA):

- Code-division multiple access (CDMA) was conceived several decades ago. Recent advances in electronic technology have finally made its implementation possible.

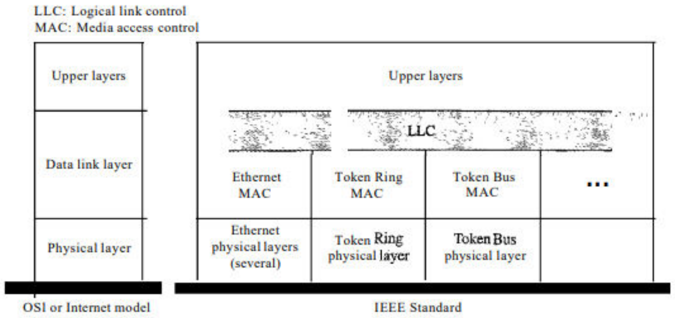
- CDMA differs from FDMA because only one channel occupies the entire bandwidth of the link. It differs from TDMA because all stations can send data simultaneously; there is no timesharing.
- CDMA simply means communication with different codes. For example, in a large room with many people, two people can talk in English if nobody else understands English. Another two people can talk in Chinese if they are the only ones who understand Chinese, and so on. In other words, the common channel, the space of the room in this case, can easily allow communication between several couples, but in different languages (codes).
- Let us assume we have four stations 1, 2, 3, and 4 connected to the same channel. The data from station 1 are d1, from station 2 are d2, and so on. The code assigned to the first station is c1, to the second is c2, and so on.
- Station 1 multiplies its data by its code to get d1. c1. Station 2 multiplies its data by its code to get d2.c2 and so on.
- The data that go on the channel are the sum of all these terms, as shown in the box. Any station that wants to receive data from one of the other three multiplies the data on the channel by the code of the sender.
- For example, suppose stations 1 and 2 are talking to each other. Station 2 wants to hear what station 1 is saying. It multiplies the data on the channel by c1 the code of station 1.

Simple idea ofcommunication with code



Wired LAN: IEEE Standards

In 1985, the Computer Society of the IEEE (Institute of Electrical and Electronics Engineers) started a project, called Project 802, to set standards to enable intercommunication among equipment from a variety of manufacturers. Project 802 does not seek to replace any part of the OSI or the Internet model. Instead, it is a way of specifying functions of the physical layer and the data link layer of major LAN protocols.

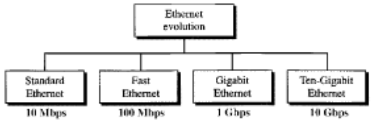


IEEE 802 is comprised of standards with separate working groups that regulate different communication networks, including IEEE 802.1 for bridging (bottom sublayer), 802.2 for Logical link (upper sublayer), 802.3 for Ethernet, 802.5 for token ring, 802.11 for Wi-Fi, 802.15 for Wireless Personal area networks, 802.15.1 for Bluetooth, 802.16 for Wireless Metropolitan Area Networks etc.

Ethernet:

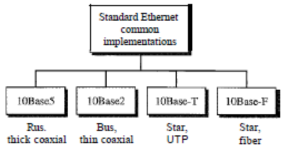
The original Ethernet was created in 1976 and since then, it has gone through four generations. Ethernet is most widely used LAN Technology, which is defined under IEEE standards 802.3. The reason behind its wide usability is Ethernet is easy to understand, implement, maintain and allows low-cost network 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.

Ethernet evolution through four generations

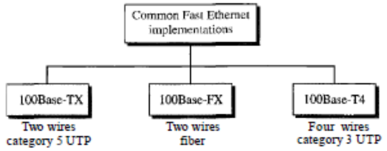


Ethernet is a family of computer networking technologies commonly used in local area networks (LAN), metropolitan area networks (MAN) and wide area networks (WAN). Systems using Ethernet communication divide data streams into packets, which are known as frames. Frames include source and destination address information, as well as mechanisms used to detect errors in transmitted data and retransmission requests. An Ethernet cable is the physical, encased wiring over which the data travels. Compared to wireless LAN technology, Ethernet is typically less vulnerable to disruptions. It can also offer a greater degree of network security and control than wireless technology, as devices must connect using physical cabling, making it difficult for outsiders to access network data or hijack bandwidth for unsanctioned devices.

Categories ofStandard Ethernet



Fast Ethernet implementations



Gigabit Ethernet implementations

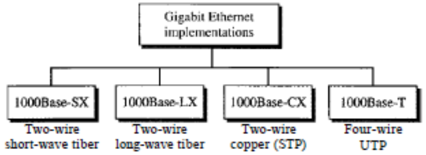


Table 13.4 *Summary ofTen-Gigabit Ethernet implementations*

Characteristics	*lOGBase-S*	*lOGBase-L*	*lOGBase-E*
Media	Short-wave S50-nm multimode	Long-wave 131O-nm single mode	Extended 1550-mm single mode
Maximum length	300m	1Okm	40km

Fiber Distributed Data Interface (FDDI):

Fiber Distributed Data Interface (FDDI) is a standard for transmission of data in local area network (LAN) over fiber optic cables. It is applicable in large LANs that can extend up to 200 kilometers in diameter.

<u>Features</u>

- FDDI uses optical fiber as its physical medium.
- It operates in the physical and medium access control (MAC layer) of the Open Systems Interconnection (OSI) network model.
- It provides high data rate of 100 Mbps and can support thousands of users.
- It is used in LANs up to 200 kilometers for long distance voice and multimedia communication.
- It uses ring-based token passing mechanism and is derived from IEEE 802.4 token bus standard.
- It contains two token rings, a primary ring for data and token transmission and a secondary ring that provides backup if the primary ring fails.
- FDDI technology can also be used as a backbone for a wide area network (WAN).

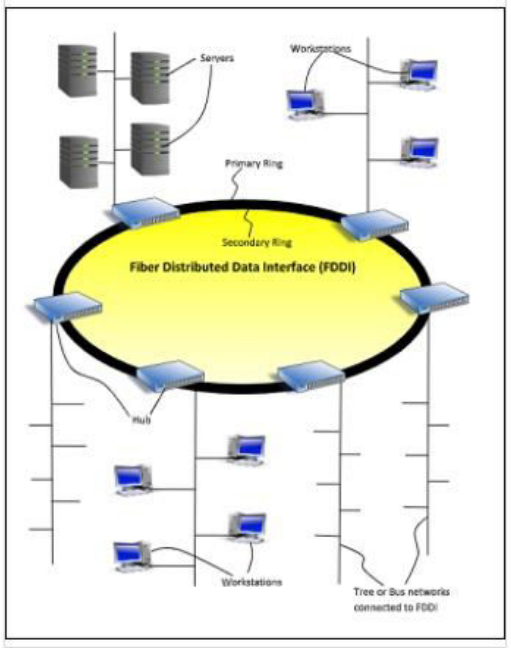

Servers
Workstations
Primary Ring
Secondary Ring
Fiber Distributed Data Interface (FDDI)
Hub
Workstations
Tree or Bus networks connected to FDDI

Fig: FDDI Implementation

Wireless LANs: IEEE 802.11x

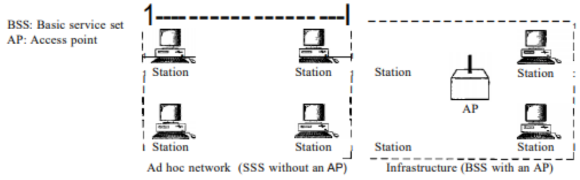
IEEE has defined the specifications for a wireless LAN, called IEEE 802.11, which covers the physical and data link layers. IEEE 802.11, commonly known as **Wi-Fi**, specifies an over-the-air interface between a wireless client and a base station or between two wireless clients.

The standard defines two kinds of services: the basic service set (BSS) and the extended service set (ESS).

Basic Service Set

IEEE 802.11 defines the basic service set (BSS) as the building block of a wireless LAN. A basic service set is made up of stationary or mobile wireless stations and an optional central base station, known as the access point (AP). Figure below shows two sets in this standard.

Figure 14.1 *Basic service sets (BSSs)*

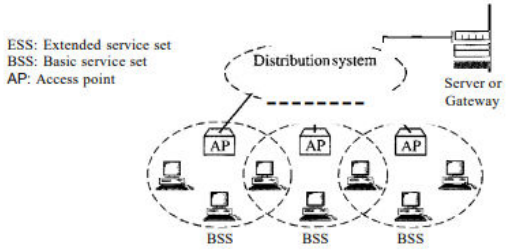


The BSS without an AP is a stand-alone network and cannot send data to other BSSs. It is called an ad hoc architecture. In this architecture, stations can form a network without the need of an AP; they can locate one another and agree to be part of a BSS. A BSS with an AP is sometimes referred to as an infrastructure network.

Extended Service Set

An extended service set (ESS) is made up of two or more BSSs with APs. In this case, the BSSs are connected through a distribution system, which is usually a wired LAN. The distribution system connects the APs in the BSSs. IEEE 802.11 does not restrict the distribution system; it can be any IEEE LAN such as an Ethernet. Note that the extended service set uses two types of stations: mobile and stationary. The mobile stations are normal stations inside a BSS. The stationary stations are AP stations that are part of a wired LAN.

Extended service sets (ESSs)



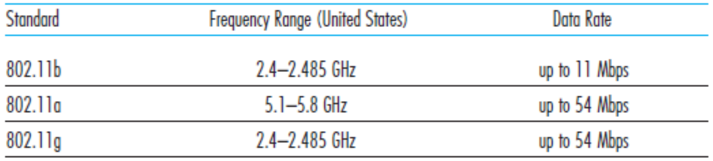
When BSSs are connected, the stations within reach of one another can communicate without the use of an AP. However, communication between two stations in two different BSSs usually occurs via two APs. The idea is similar to communication in a cellular network if we consider each BSS to be a cell and each AP to be a base station. Note that a mobile station can belong to more than one BSS at the same time.

Wi-Fi:

The IEEE 802.11 wireless LAN, also known as Wi-Fi, is the name of a popular wireless networking technology that uses radio waves to provide wireless high-speed Internet and network connections. Wi-

Fi networks have no physical wired connection between sender and receiver, by using radio frequency (RF) technology (a frequency within the electromagnetic spectrum associated with radio wave propagation). When an RF current is supplied to an antenna, an electromagnetic field is created that then is able to propagate through space.

There are several 802.11 standards for wireless LAN technology, including 802.11b, 802.11a, and 802.11g. Table below summarizes the main characteristics of these standards. 802.11g is by far the most popular technology.



Standard	Frequency Range (United States)	Data Rate
802.11b	2.4–2.485 GHz	up to 11 Mbps
802.11a	5.1–5.8 GHz	up to 54 Mbps
802.11g	2.4–2.485 GHz	up to 54 Mbps

Wi-Fi uses multiple parts of the IEEE 802 protocol family and is designed to seamlessly interwork with its wired sister protocol Ethernet. Devices that can use Wi-Fi technologies include desktops and laptops, smartphones and tablets, smart TVs, printers, digital audio players, digital cameras, cars and drones. Compatible devices can connect to each other over Wi-Fi through a wireless access point as well as to connected Ethernet devices and may use it to access the Internet. Such an access point (or hotspot) has a range of about 20 meters (66 feet) indoors and a greater range outdoors. Hotspot coverage can be as small as a single room with walls that block radio waves, or as large as many square kilometers achieved by using multiple overlapping access points.

Wi-Fi is potentially more vulnerable to attack than wired networks because anyone within range of a network with a wireless network interface controller can attempt access. Wi-Fi Protected Access (WPA) is a family of technologies created to protect information moving across Wi-Fi networks and includes solutions for personal and enterprise networks.

Bluetooth:

Bluetooth is a short-range wireless communication technology that allows devices such as mobile phones, computers, and peripherals to transmit data or voice wirelessly over a short distance. The purpose of Bluetooth is to replace the cables that normally connect devices, while still keeping the communications between them secure. It creates a 10-meter radius wireless network, called a personal area network (PAN) or piconet, which can network between two and eight devices. Bluetooth uses less power and costs less to implement than Wi-Fi. Its lower power also makes it far less prone to suffering from or causing interference with other wireless devices in the same 2.4GHz radio band.

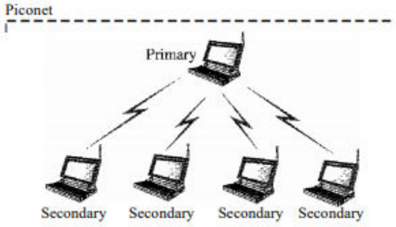
There are some downsides to Bluetooth. The first is that it can be a drain on battery power for mobile wireless devices like smartphones, though as the technology (and battery technology) has improved, this problem is less significant than it used to be. Also, the range is fairly limited, usually extending only about 30 feet, and as with all wireless technologies, obstacles such as walls, floors, or ceilings can reduce this

range further. The pairing process may also be difficult, often depending on the devices involved, the manufacturers, and other factors that all can result in frustration when attempting to connect.

Bluetooth defines two types of networks: piconet and scatternet.

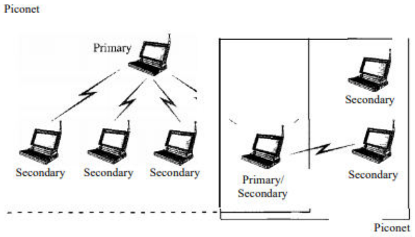
Piconets

A Bluetooth network is called a piconet, or a small net. A piconet can have up to eight stations, one of which is called the primary; t the rest are called secondaries. All the secondary stations synchronize their clocks and hopping sequence with the primary. Note that a piconet can have only one primary station. The communication between the primary and the secondary can be one-to-one or one-to-many. Figure below shows a piconet.



Scatternet

Piconets can be combined to form what is called a scatternet. A secondary station in one piconet can be the primary in another piconet. This station can receive messages from the primary in the first piconet (as a secondary) and, acting as a primary, deliver them to secondaries in the second piconet. A station can be a member of two piconets. Figure below illustrates a scatternet.



Token Ring:

Token ring is the IEEE 802.5 standard for a token-passing ring in Communication networks. A ring consists of a collection of ring interfaces connected by point-to-point lines i.e. ring interface of one station is connected to the ring interfaces of its left station as well as right station. Internally, signals travel around the Communication network from one station to the next in a ring. These point-to-point links can be created with twisted pair, coaxial cable or fiber optics. Each bit arriving at an interface is copied into a 1-

bit buffer. In this buffer the bit is checked and may be modified and is then copied out to the ring again. This copying of bit in the buffer introduces a 1-bit delay at each interface.

Token Ring is a LAN protocol defined in the IEEE 802.5 where all stations are connected in a ring and each station can directly hear transmissions only from its immediate neighbor. Permission to transmit is granted by a message (token) that circulates around the ring. A token is a special bit pattern (3 bytes long). There is only one token in the network. Token-passing networks move a small frame, called a token, around the network. Possession of the token grants the right to transmit. If a node receiving the token in order to transmit data, it seizes the token, alters 1 bit of the token (which turns the token into a start-of-frame sequence), appends the information that it wants to transmit, and sends this information to the next station on the ring. Since only one station can possess the token and transmit data at any given time, there are no collisions.

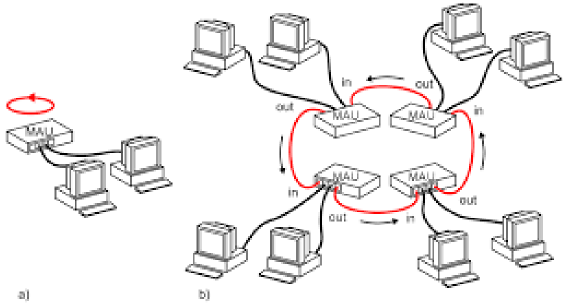


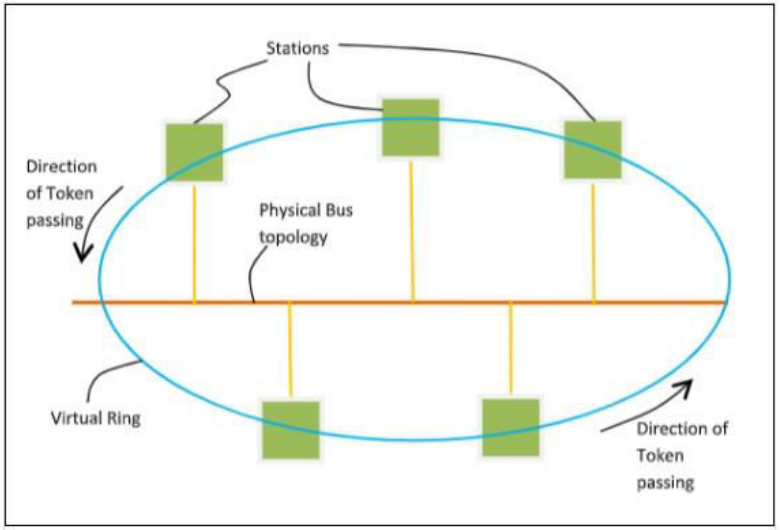
Fig: Two examples of Token Ring networks a) Using a single MAU b) Using several MAUs connected to each other, MAU (Media Access Unit)

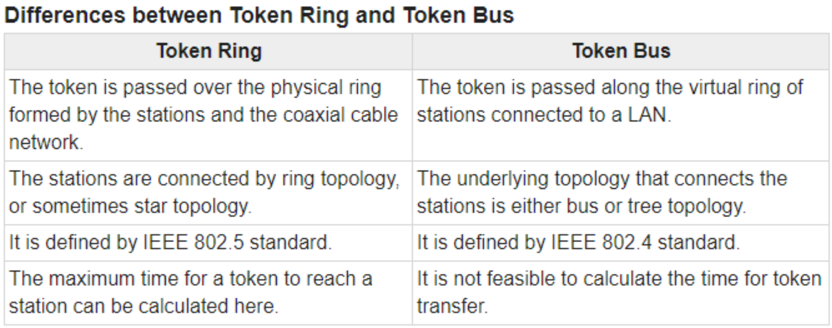
Token Bus:

Token Bus (IEEE 802.4) is a standard for implementing token ring over virtual ring in LANs. The physical media has a bus or a tree topology and uses coaxial cables. A virtual ring is created with the nodes/stations and the token is passed from one node to the next in a sequence along this virtual ring. Each node knows the address of its preceding station and its succeeding station. A station can only transmit data when it has the token. The working principle of token bus is similar to Token Ring.

Token Passing Mechanism in Token Bus

A token is a small message that circulates among the stations of a computer network providing permission to the stations for transmission. If a station has data to transmit when it receives a token, it sends the data and then passes the token to the next station; otherwise, it simply passes the token to the next station. This is depicted in the following diagram:





Differences between Token Ring and Token Bus

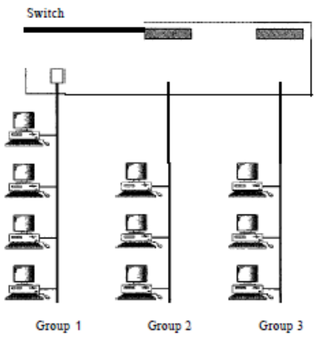
Token Ring	Token Bus
The token is passed over the physical ring formed by the stations and the coaxial cable network.	The token is passed along the virtual ring of stations connected to a LAN.
The stations are connected by ring topology, or sometimes star topology.	The underlying topology that connects the stations is either bus or tree topology.
It is defined by IEEE 802.5 standard.	It is defined by IEEE 802.4 standard.
The maximum time for a token to reach a station can be calculated here.	It is not feasible to calculate the time for token transfer.

Virtual LANs:

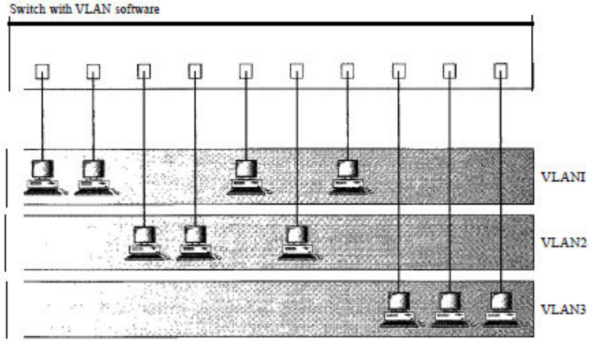
- A station is considered part of a LAN if it physically belongs to that LAN. The criterion of membership is geographic.
- What happens if we need a virtual connection between two stations belonging to two different physical LANs?
- We can roughly define a virtual local area network (VLAN) as a local area network configured by software, not by physical wiring.

- Figure below shows a switched LAN in an engineering firm in which 10 stations are grouped into three LANs that are connected by a switch.

A switch connecting three LANs



- The first four engineers work together as the first group, the next three engineers work together as the second group, and the last three engineers work together as the third group. The LAN is configured to allow this arrangement.
- But what would happen if the administrators needed to move two engineers from the first group to the third group, to speed up the project being done by the third group?
- The LAN configuration would need to be changed. The network technician must rewire.
- The problem is repeated if, in another week, the two engineers move back to their previous group.
- In a switched LAN, changes in the work group mean physical changes in the network configuration.
- Figure below shows the same switched LAN divided into VLANs. The whole idea of VLAN technology is to divide a LAN into logical, instead of physical, segments.



- A LAN can be divided into several logical LANs called VLANs. Each VLAN is a work group in the organization.

- If a person moves from one group to another, there is no need to change the physical configuration. The group membership in VLANs is defined by software, not hardware.
- Any station can be logically moved to another VLAN. All members belonging to a VLAN can receive broadcast messages sent to that particular VLAN.
- This means if a station moves from VLAN 1 to VLAN 2, it receives broadcast messages sent to VLAN 2, but no longer receives broadcast messages sent to VLAN 1.
- It is obvious that the problem in our previous example can easily be solved by using VLANs. Moving engineers from one group to another through software is easier than changing the configuration of the physical network.