

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.

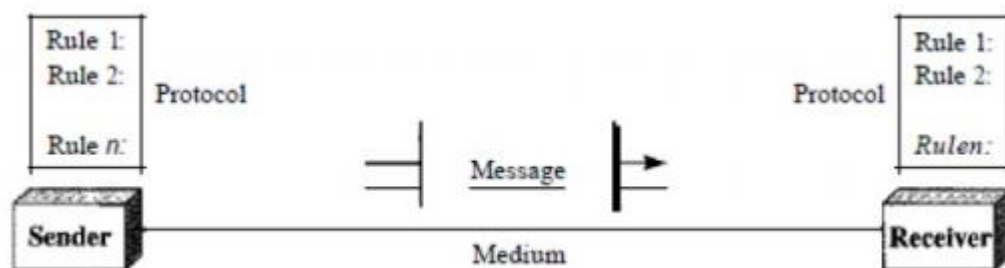


Fig: Data communication model

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.

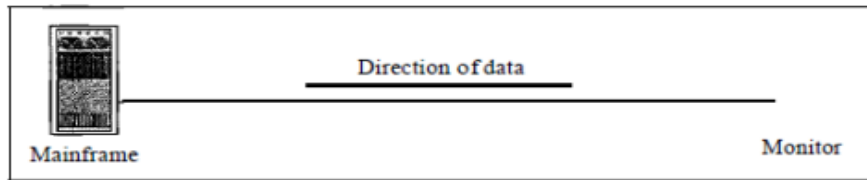
Data Transmission Modes

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

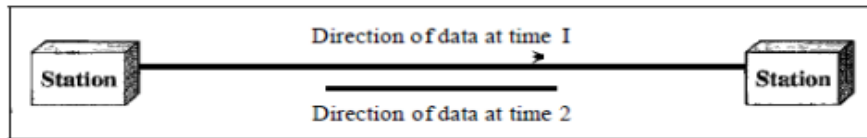
Simplex: In simplex mode, the communication is unidirectional, as on a one-way street. Only one of the two devices on a link can transmit; the other can only receive. Keyboards and traditional monitors are examples of simplex devices. The keyboard can only introduce input; the monitor can only accept 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.

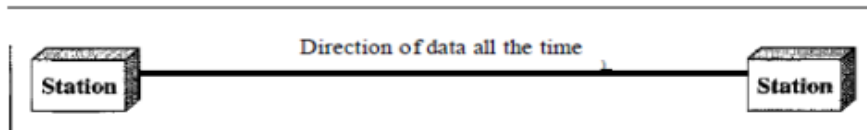
Full-Duplex: In full-duplex both stations can transmit and receive simultaneously. The full-duplex mode is like a two-way street with traffic flowing in both directions at the same time. In full-duplex mode, signals going in one direction share the capacity of the link: with signals going in the other 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.



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.

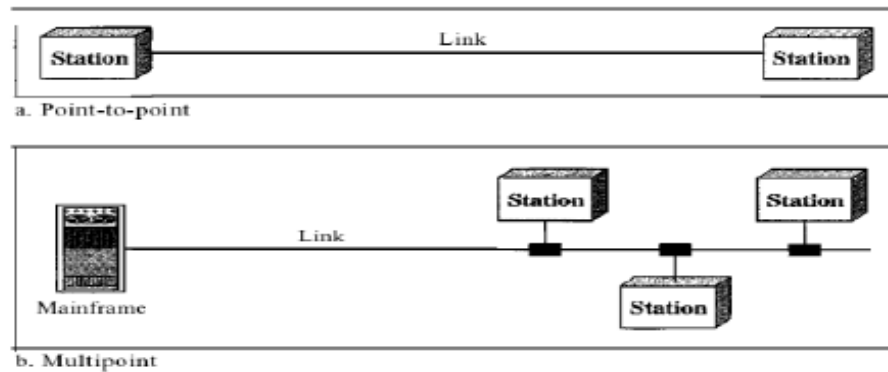
Security: Network security issues include protecting data from unauthorized access, protecting data from damage and development, and implementing policies and procedures for recovery from breaches and data losses.

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.

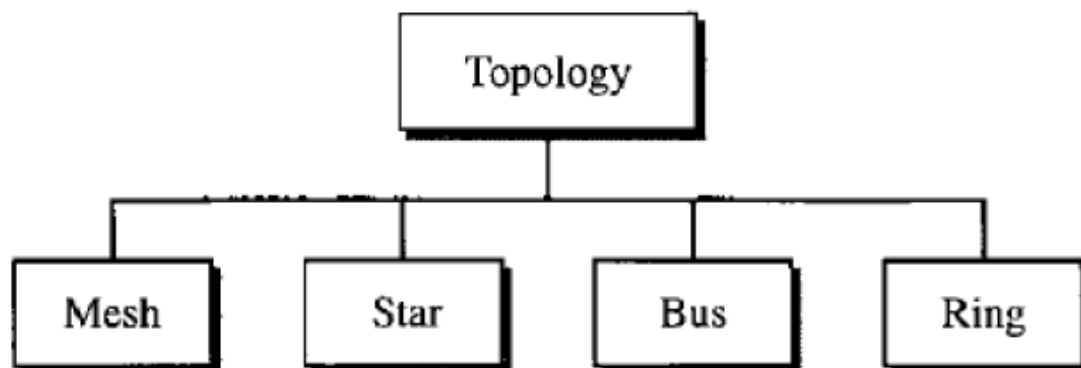
Point-to-Point: A point-to-point connection provides a dedicated link between two devices. The entire capacity of the link is reserved for transmission between those two devices. Most point-to-point connections use an actual length of wire or cable to connect the two ends, but other options, such as 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.



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 $n - 1$ 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

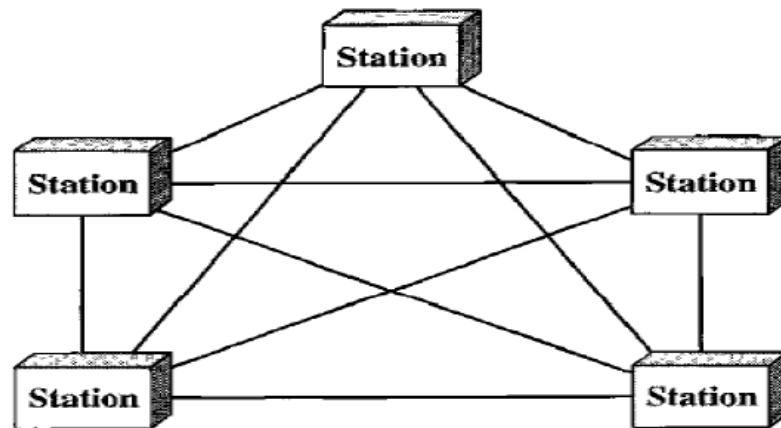
number of links by 2. In other words, we can say that in a mesh topology, we need $n(n-1)/2$ duplex-mode links. To accommodate that many links, every device on the network must have $n - 1$ input/output ports to be connected to the other $n - 1$ stations.

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

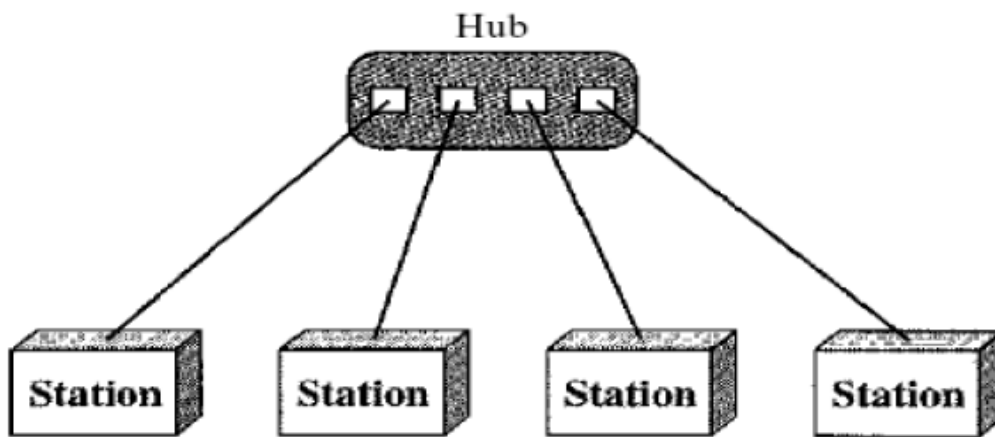
send data to another, it sends the data to the controller, which then relays the data to the other connected device.

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.

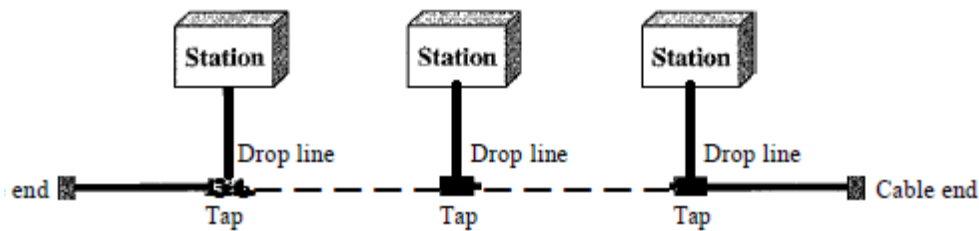
Advantages:

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.

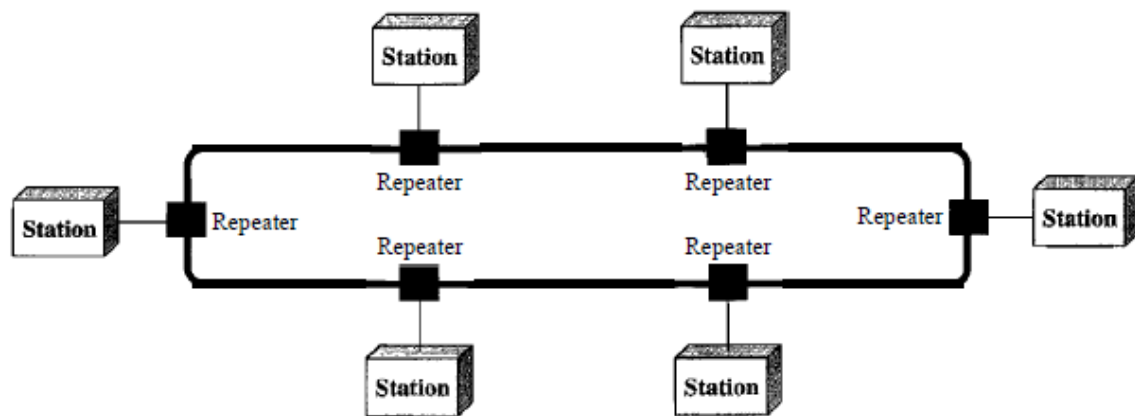
Advantages:

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*

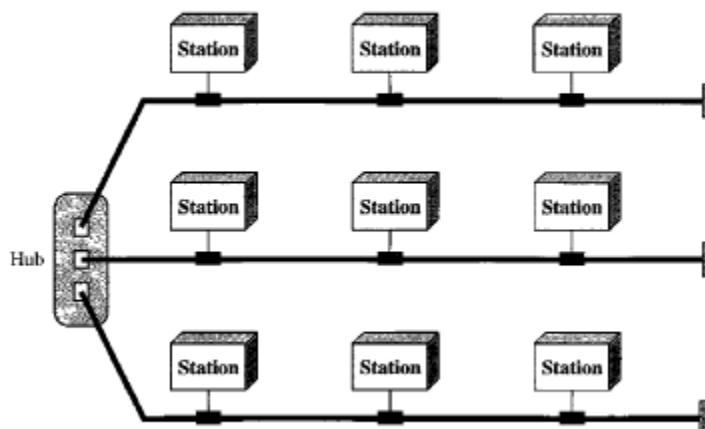


5. Hybrid Topology

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

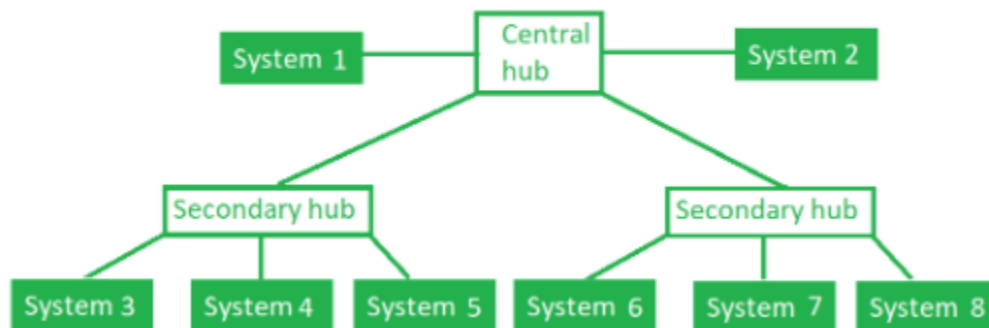
The main advantage of hybrid structures is the degree of flexibility they provide, as there are few limitations on the network structure itself that a hybrid setup can't accommodate.

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.

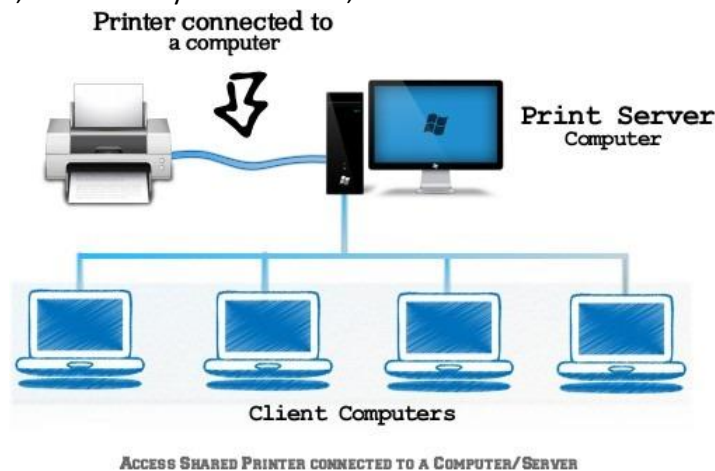
Disadvantages:

- 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.



- **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 like 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

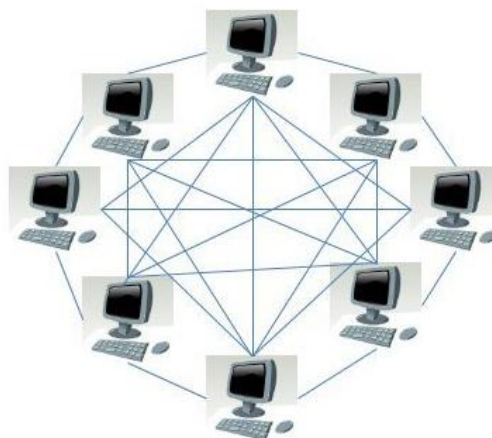
Network architecture is the design of a computer network. It is a framework for the specification of a network's physical components and their functional organization and configuration, its operational principles and procedures, as well as communication protocols used.

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:

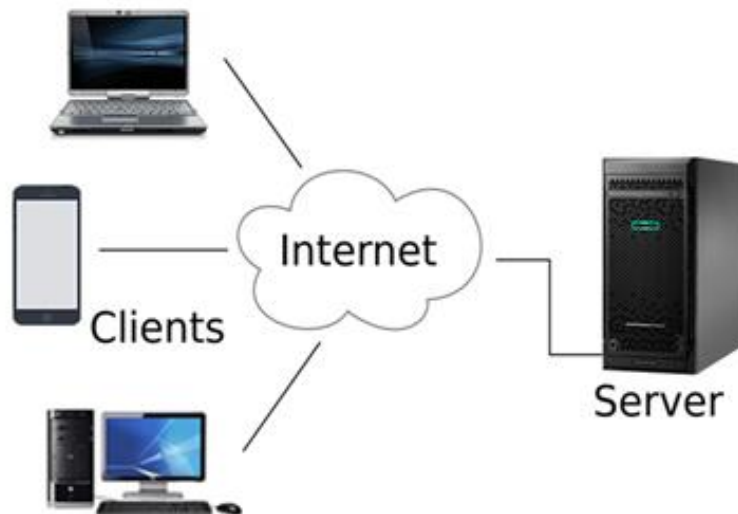
- It is less costly as it does not contain any dedicated server.
- If one computer stops working but, other computers will continue working.
- It is easy to set up and maintain as each computer manages itself.

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.

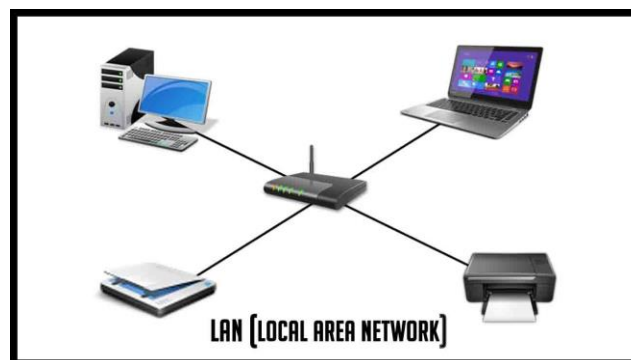


Fig: Local Area Network

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.

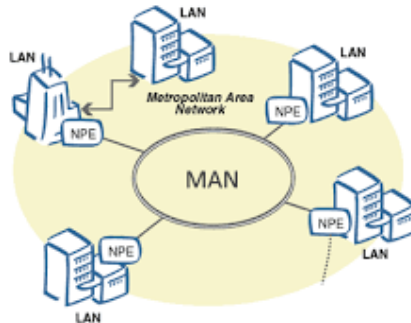


Fig: Metropolitan Area Network

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.

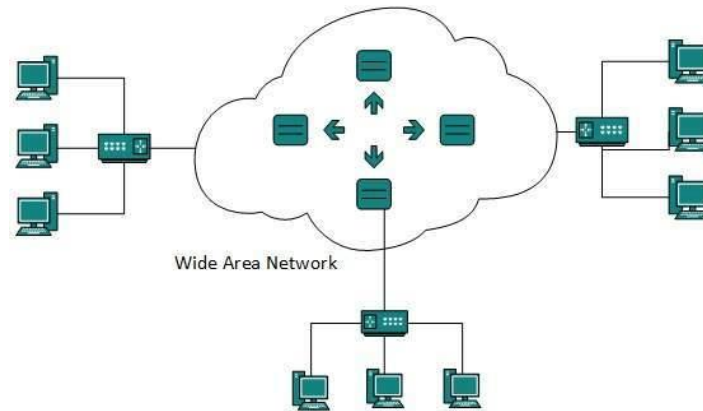


Fig: Wide Area Network

Characteristics

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

Advantages

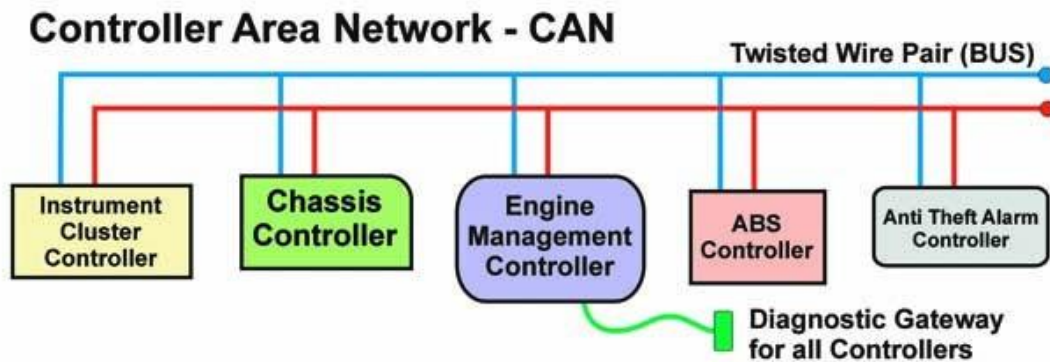
- Covers large geographical area
- Shared software and resources with connecting workstations
- Information can be exchanged to anyone else worldwide in the network

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.



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

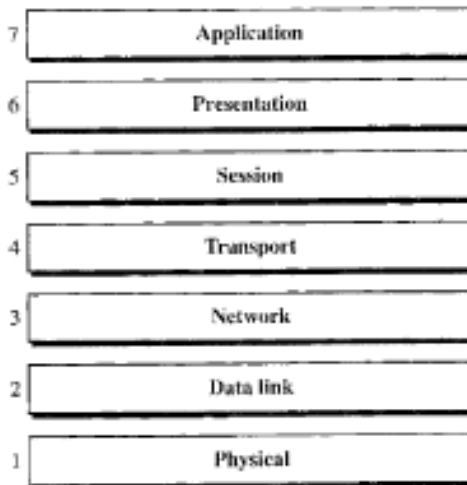
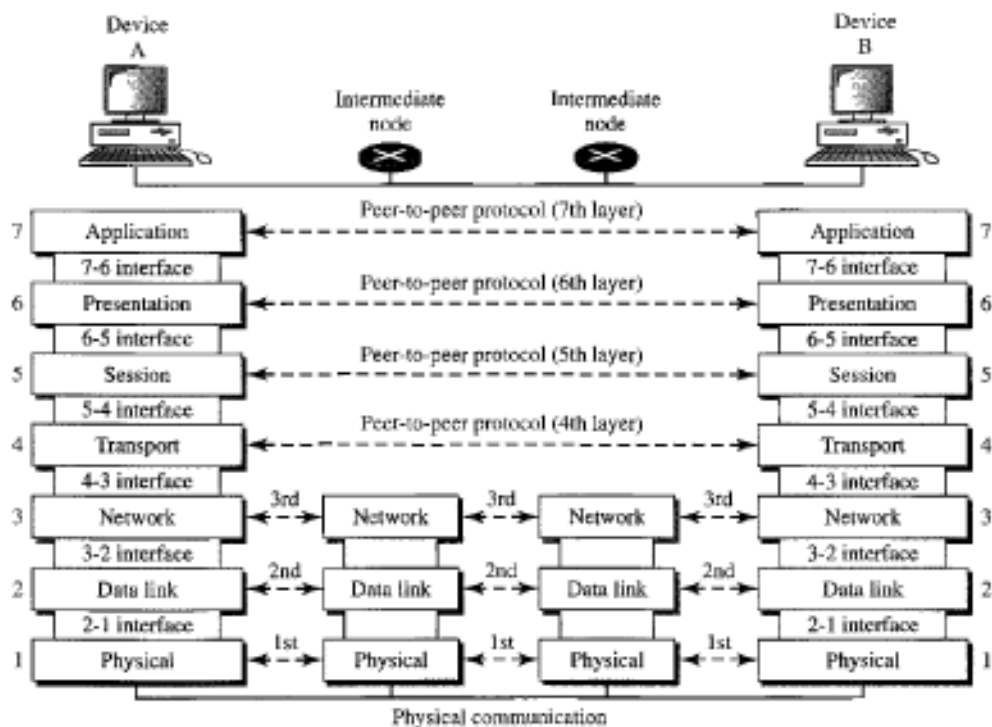


Figure 2.3 *The interaction between layers in the OSI model*



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.

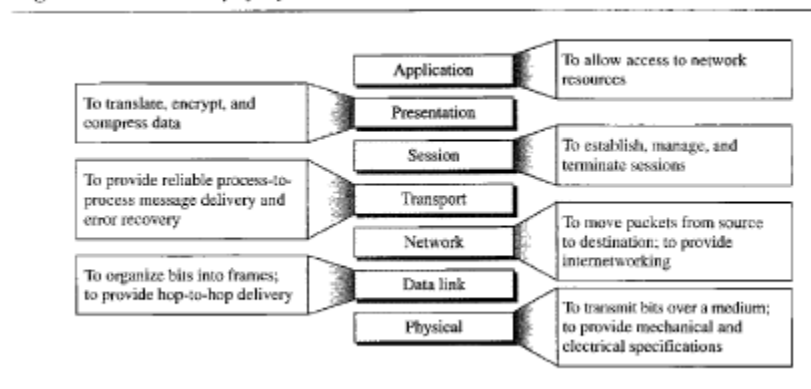
- **Compression.** Data compression reduces the number of bits contained in the information. Data compression becomes particularly important in the transmission of multimedia such as text, audio, and video.

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



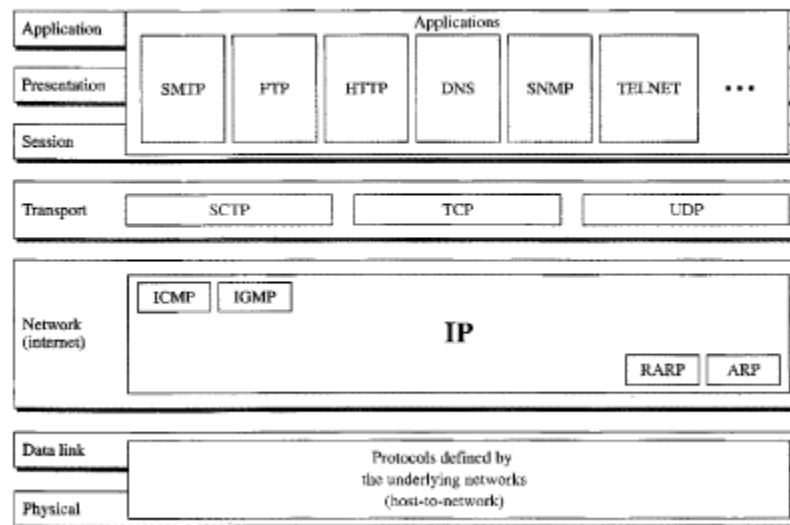
TCP/IP Protocol Suite:

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 them travel independently to the destination (potentially on a different network). They may even arrive in a different order than they were sent, in which case it is the job of higher layers to rearrange them, if in-order delivery is desired.

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 guarantee 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.