Chapter 1

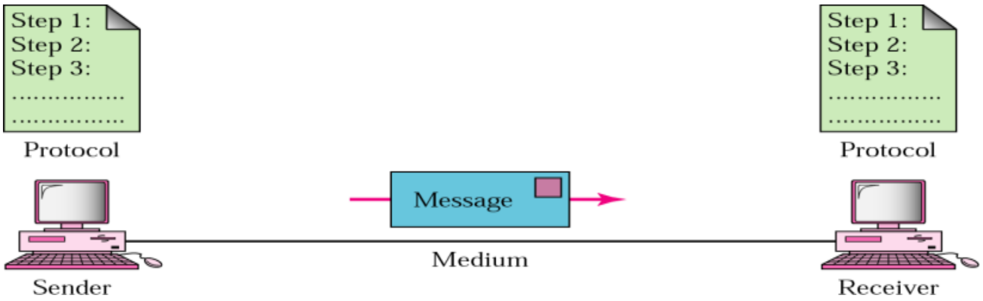
Introduction to Computer Networks

1. Data Communication

When we communicate, we are sharing information. This sharing can be local or remote. Between individuals, local communication usually occurs face to face, while remote communication takes place over distance.

1.1. Components

A data communications system has five components.



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

1.2. Data Representation

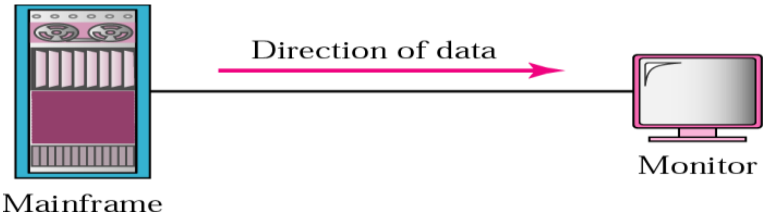
Information today comes in different forms such as text, numbers, images, audio, and video.

- **Text:** In data communications, text is represented as a bit pattern, a sequence of bits (0s or 1s). Different sets of bit patterns have been designed to represent text symbols. Each set is called a code, and the process of representing symbols is called coding. Today, the prevalent coding system is called Unicode, which uses 32 bits to represent a symbol or character used in any language in the world.
- **Numbers:** Numbers are also represented by bit patterns. However, a code such as ASCII is not used to represent numbers; the number is directly converted to a binary number to simplify mathematical operations.
- **Images:** Images are also represented by bit patterns. In its simplest form, an image is composed of a matrix of pixels (picture elements), where each pixel is a small dot. The size of the pixel depends on the resolution. For example, an image can be divided into 1000 pixels or 10,000 pixels. In the second case, there is a better representation of the image (better resolution), but more memory is needed to store the image. After an image is divided into pixels, each pixel is assigned a bit pattern. The size and the value of the pattern depend on the image. For an image made of only black-and- white dots (e.g., a chessboard), a 1-bit pattern is enough to represent a pixel. If an image is not made of pure white and pure black pixels, you can increase the size of the bit pattern to include gray scale. For example, to show four levels of gray scale, you can use 2-bit patterns. A black pixel can be represented by 00, a dark gray pixel by 01, a light gray pixel by 10, and a white pixel by 11. There are several methods to represent color images. One method is called RGB, so called because each color is made of a combination of three primary colors: red, green, and blue. The intensity of each color is measured, and a bit pattern is assigned to it. Another method is called YCM, in which a color is made of a combination of three other primary colors: yellow, cyan, and magenta.
- **Audio:** Audio refers to the recording or broadcasting of sound or music. Audio is by nature different from text, numbers, or images. It is continuous, not discrete. Even when we use a microphone to change voice or music to an electric signal, we create a continuous signal.
- **Video:** Video refers to the recording or broadcasting of a picture or movie. Video can either be produced as a continuous entity (e.g., by a TV camera), or it can be a combination of images, each a discrete entity, arranged to convey the idea of motion. Again we can change video to a digital or an analog signal.

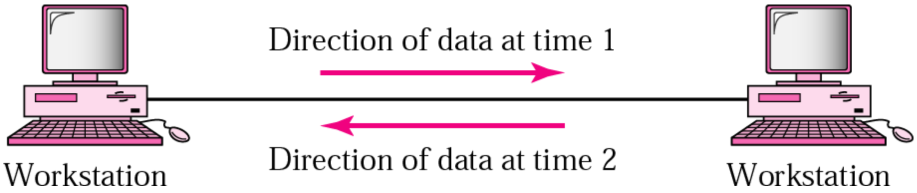
1.3. Data Flow

Communication between two devices can be simplex, half-duplex, or full-duplex. It refers to the direction in which data moves between two devices.

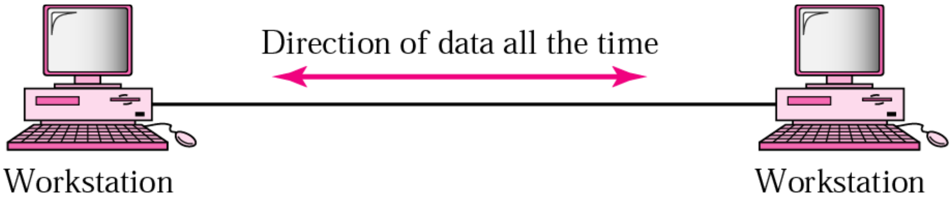
- **Simplex:** In simplex mode, the communication is unidirectional, as on a one-way street. Only one of the two devices on a link can transmit; the other can only receive. Keyboards and traditional monitors are examples of simplex devices.



- **Half-Duplex:** In half-duplex mode, each station can both transmit and receive, but not at the same time. When one device is sending, the other can only receive, and vice versa. The half-duplex mode is 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. Walkie-talkies and CB (citizens band) radios are both half-duplex systems.



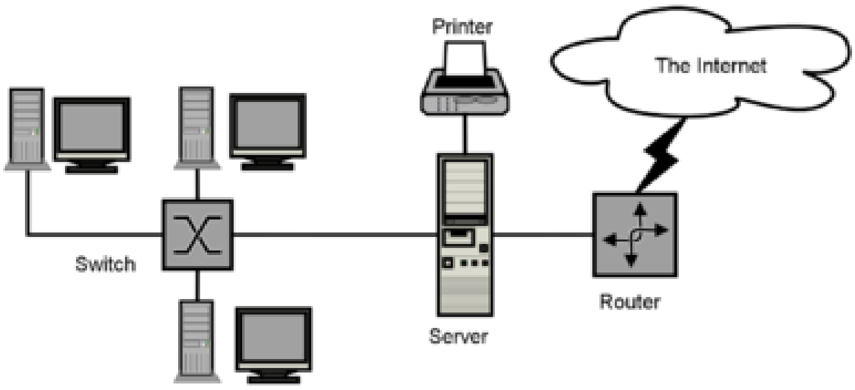
- **Full-Duplex:** In full-duplex mode, both stations can transmit and receive simultaneously. In full-duplex mode, signals going in one direction share the capacity of the link: with signals going in the other direction. One common example of full-duplex communication is the telephone network. When two people are communicating by a telephone line, both can talk and listen at the same time.



2. NETWORKS

2.1. Computer Network

A computer network is an interconnection among two or more computers or computing devices. Such interconnection allows computers to share data and resources among each other.



A sample network diagram

- **Notes:**
  - In a communication network, each device that is a part of a network and that can receive, create, store or send data to different network routes is called a **node**.
  - A special computer which provides services to other computer/devices in a network is called **server**.
  - A computer/device connected in a network sending request to server is called **client**.

2.2. Network Criteria

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

A) Performance

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.

➢ ***Parameters for Measuring Network Performance***

- **Bandwidth:** The maximum amount of data that can be transmitted over a network in a given amount of time, typically measured in bits per second **(bps)**.
- **Throughput:** Throughput is measured by tabulating the amount of data transferred between multiple locations during a specific period of time, usually resulting in the unit of bits per second **(bps).**

$$\text{Throughput} = \frac{\text{Total Data Transferred (bits or bytes)}}{\text{Time Taken (seconds)}}$$

While bandwidth lets you set the theoretical limit of data transfer, throughput measures the actual amount of data packets successfully sent to the destination via the network.

- **Latency:** In a network, during the process of data communication, latency (also known as **delay**) is defined as the total time taken for a complete message to arrive at the destination, starting with the time when the first bit of the message is sent out from the source and ending with the time when the last bit of the message is delivered at the destination. Latency is measured in milliseconds (**ms**).

Latency = Propagation Time + Transmission Time + Queuing Time + Processing Delay

- **Propagation Time**: It is the time required for a bit to travel from the source to the destination. For example, for an electric signal, propagation time is the time taken for the signal to travel through a wire.

Propagation time = Distance / Propagation speed

- **Transmission Time:** Transmission Time is a time based on how long it takes to send the signal down the transmission line.

Transmission time = Message size / Bandwidth

- **Queuing Time:** Time a packet spends waiting in queues.
- **Processing Delay:** Time for a router or switch to process the packet.
- **Jitter:** The variation in packet delay over time. It is critical for real-time applications like voice and video streaming.

B) Reliability

In addition to accuracy of delivery, network reliability is measured by the frequency of failure, the time it takes a link to recover from a failure, and the network's robustness in a catastrophe.

C) 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.

2.3. Advantages of Computer Network

- **Resource Sharing:** Data, Hardware resources (Modem, Hard Disk, DVD Drive, Scanner, etc.) and Software resources (Application Software, Anti-Virus tools, etc.) can be easily shared on computer networks by connecting these devices to one computer (server).
- **Cost saving:** Sharing of resources in computer networking leads to cost saving.

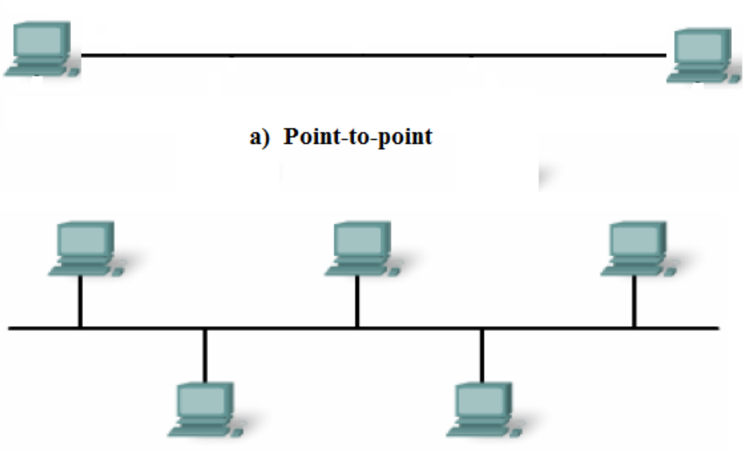
- **Improved Communication:** A computer network enables fast, reliable and easy communication among its users. We can easily communicate with anyone through email, video conferencing or chatting through networking.
- **Time saving:** It takes negligible time to send and receive messages, audio, video and images on a computer network. We can easily watch live videos and can talk live to anyone sitting in some other corner of the world on the computer network. This leads to time saving.
- **Increased storage:** On a computer network, same data is replicated on multiple computers to ensure the availability of data in case of some computer getting faulty. Mostly the data is kept on servers and is shared with legitimate users. This ensures data security and reliability.

2.4. Physical Structures

2.4.1. Type of Connection

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. 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

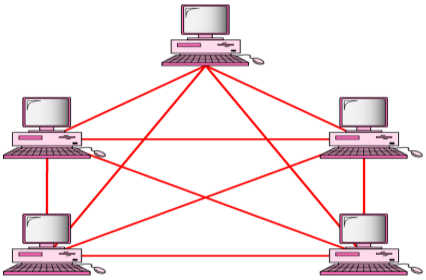
b) Multipoint

2.4.2. Physical Topology

The term physical topology refers to the way in which a network is laid out physically. 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.

A) Mesh topology

In a mesh topology, every device has a dedicated point-to-point link to every other device.



- **Mesh topology advantages**

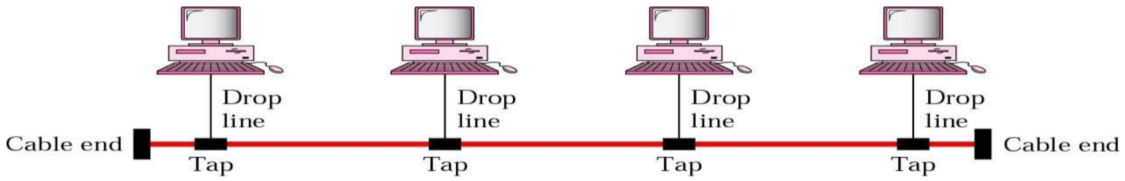
- It offers highest **reliability, privacy and security**. Every message travels along a dedicated line, only the intended recipient sees it.
- **Traffic problems are eliminated.** because there is a dedicated link between any two devices which guarantees that each connection can carry its own data load.
- Mesh topology is **robust** because the network does not crash if one link becomes unusable.
- **Fault identification and fault isolation is easy**. This enables the network manager to discover the precise location of fault and aids in finding its cause and solution.

- **Mesh topology disadvantages**

- Main disadvantage is the amount of cabling required to interconnect the devices and the number of I/O ports needed at each device.
- Installation of new devices and reconnection is difficult.
- Mesh topology is expensive due to large number of hardware requirement (cables/ports).

B) Bus Topology

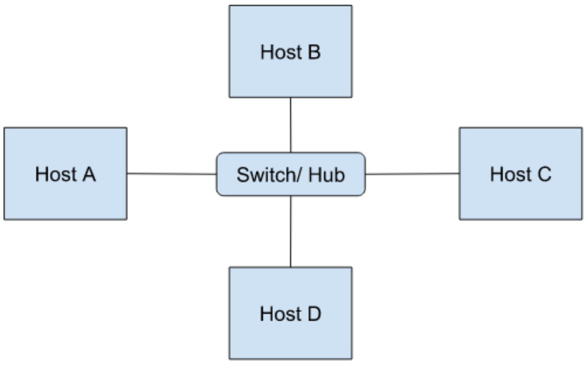
In a bus topology, one long cable acts as a **backbone** to link all the devices in a network. Nodes are connected to the bus cable by **drop lines and taps**. A drop line is a connection running between the device and the main cable. A tap is a connector that either splices into the main cable or punctures the sheathing of a cable to create a contact with the metallic core.



- **Bus topology advantages**
  - Easy to install.
  - Main cable can be laid out to attach new node when required.
  - Less cabling required as compare to mesh.
- **Bus topology disadvantages**
  - As the signal travels along the backbone, it becomes weaker. Hence there is limit on the number of devices that a bus can support.
  - It is difficult to identify and isolate fault (determine cause of a problem).
  - Signal reflection/loss of signal at taps cause degradation in quality of the signal.
  - A fault or break in the bus cable stops all transmission.

C) Star Topology

In a star topology, each device has a **dedicated point-to-point link** only to a central controller. 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.

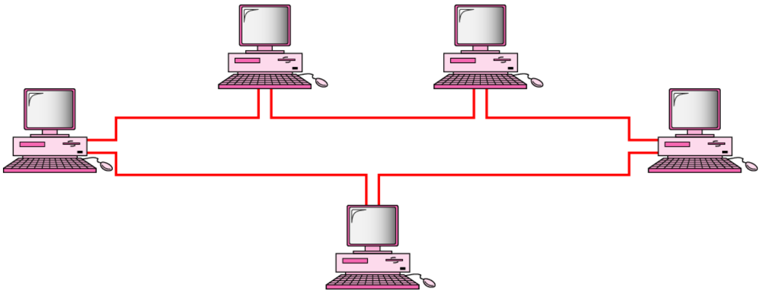


- **Star topology advantages**
  - Since each device requires only one link to the controller, it is less expensive than mesh topologie.
  - Easy to install and reconfigure.
  - It is robust because even if one link fails, the other remain active.
  - Fault identification and isolation is easy.

- **Star topology disadvantages**
  - Performance of network depends entirely on the hub. If the hub fails entire network stops.
  - Speed of communication depends upon number of devices connected to the hub.
  - Cabling cost is more than some other topologies since cables must be pulled from all computers to the central hub.

D) Ring Topology

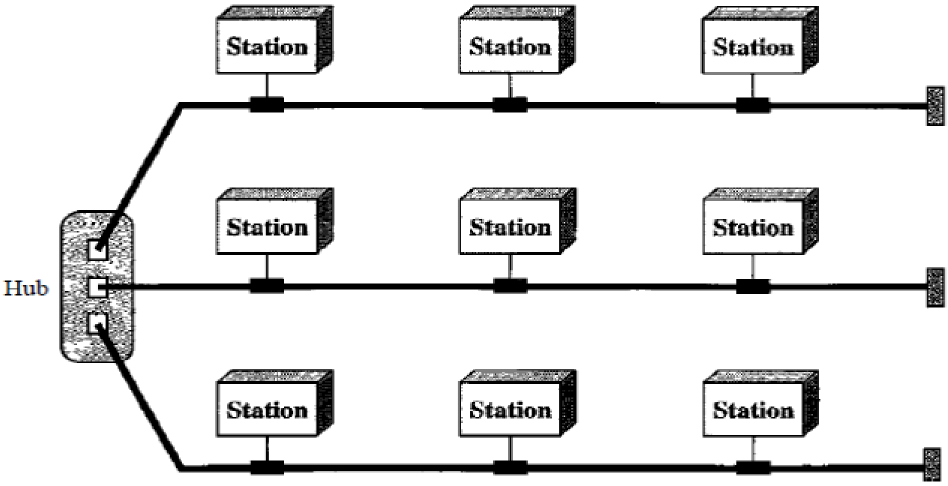
In a ring topology, each device has a **dedicated point-to-point** connection with only the two devices on either side of it. A signal is passed along the ring in one direction, from device to device, until it reaches its destination. Each device in the ring incorporates a **repeater**. When a device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



- **Ring topology Advantages**
  - Since each device is linked only to its immediate neighbors, the ring network is **easy to install or reconfigure**.
  - **Fault isolation is easy** since the faulty device can be bypassed by altering the connections.
  - Only **one device can send data at a time**. The device which wants to send data must **capture a special frame** called token.
- **Ring topology disadvantages**
  - A **break in the ring** (such as disabled station) can disable the entire network.
  - The major constraint is the **maximum ring length and the number of devices connected.**
  - Since the traffic is unidirectional, **data transfer is slow**.
  - **Failure** of one computer on the ring can affect the whole network.

E) Hybrid Topology

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

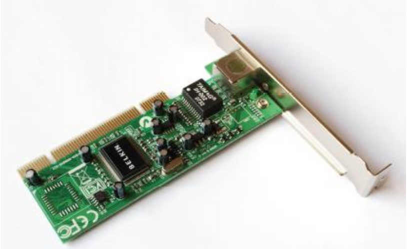


A hybrid topology: a star backbone with three bus networks

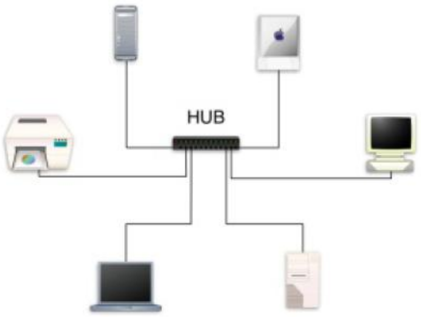
2.5. Network Devices

Other than the transmission media many other devices are required to form computer networks. Some of these devices are:

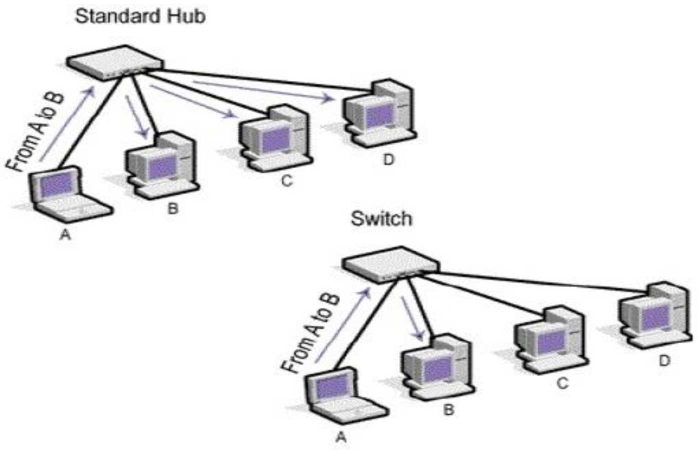
- **NIC (Network Interface Card):**An NIC (Network Interface Card) is a device that enables a computer to connect to a network and communicate. Any computer which has to be a part of a computer network must have an NIC installed in it.



- **MODEM:** A modem (Modulator - Demodulator) is a peripheral device that enables a computer to transmit data over, telephone or cable lines. It converts the digital data from the sender computer into analog form to be able to send it over telephone lines. At the receiving end modem converts the data from analog form to digital form and stores into receiving computer.
- **HUB:** A Hub is an electronic device that connects several nodes to form a network and redirects the received information to all the connected nodes in broadcast mode. The computer(s) for which the information is intended receive(s) this information and accept(s) it. Other computers on the network simply reject this information.



- **SWITCH:** A Switch is an intelligent device that connects several nodes to form a network and redirects the received information only to the intended node(s).
  The difference between the switch and the hub is that Hub broadcasts the received information to all the nodes. Switch does not broadcast instead sends the information selectively only to those computers for which it is intended. This makes a switch more efficient than a hub.



- **Repeater:** A Repeater is a device that is used to regenerate a signal which is on its way through a communication channel. A repeater regenerates the received signal and retransmits it to its destination.
- **Gateway:** A Gateway is a device, which is used to connect different types of networks.

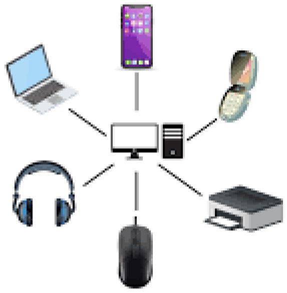
2.6. Categories of Networks

Computer network can be classified **based on the geographical area** they cover, i.e. the area over which the network is spread. Computer networks are broadly categorized as:

- PAN (Personal Area Networks)
- LAN (Local Area Networks)
- MAN (Metropolitan Area Networks)
- WAN (Wide Area Networks)

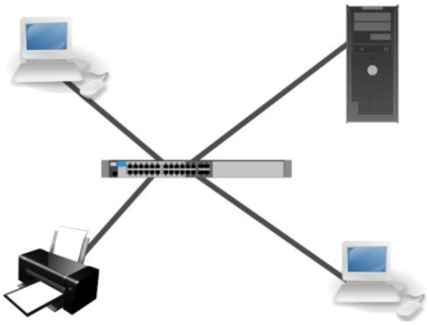
2.6.1. *Personal Area Network (PAN)*

The interconnection of devices within the range of an individual person, typically within a range of 10 meters. For example, a wireless network connecting a computer with its keyboard, mouse or printer is a PAN. Also, a PDA (Personal Digital Assistant) that controls the user's hearing aid or pacemaker fits in this category. Another example of PAN is a Bluetooth. Typically, this kind of network could also be interconnected without wires to the Internet or other networks.



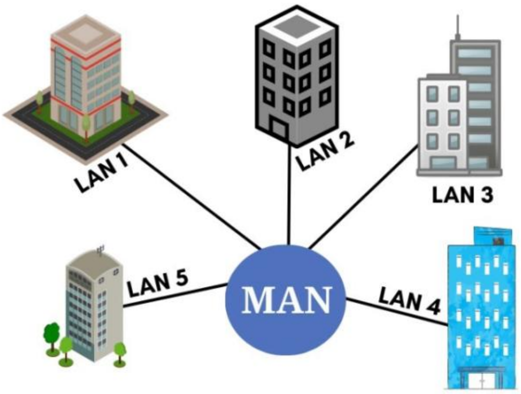
2.6.2. *Local Area Network (LAN)*

Privately-owned networks covering a small geographic area, like a home, office, building or group of buildings (e.g. campus). They are widely used to connect computers in company offices and factories to share resources (e.g., printers) and exchange information. 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.



2.6.3. *Metropolitan Area Network (MAN)*

Covers a larger geographical area than is a LAN, ranging from several blocks of buildings to entire cities. MANs can also depend on communications channels of moderate-to-high data rates. A MAN might be owned and operated by a single organization, but it usually will be used by many individuals and organizations. MANs might also be owned and operated as public utilities. They will often provide means for internetworking of LANs. Metropolitan Area Networks can span up to 50km, devices used are modem and wire/cable.



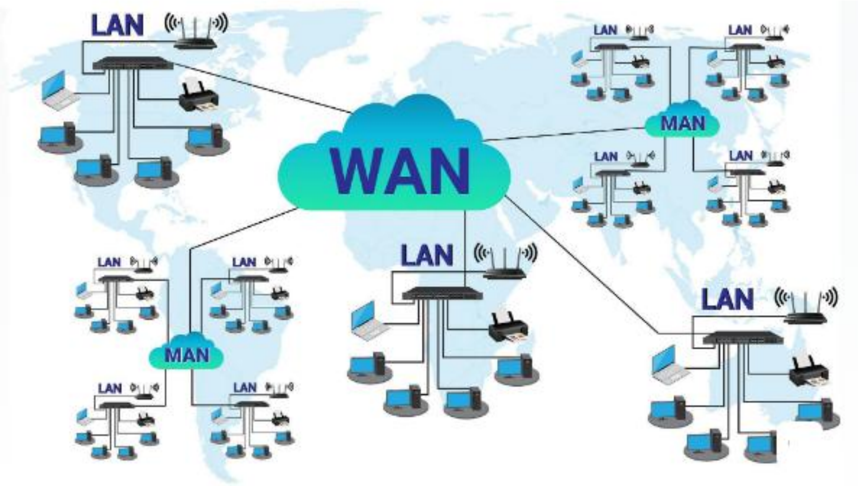
2.6.4. *Wide Area Network (WAN)*

A WAN (Wide Area Network) is a type of computer network that covers a large geographic area, typically connecting multiple smaller networks like LANs or MANs. WANs allow devices in different locations, such as cities, countries, or continents, to communicate and share resources.

- *Key features of WAN:*

- **Geographic Scope**: Covers large areas, potentially intercontinental.
- **Connectivity**: Uses public or private transmission lines, such as leased lines, satellites, or fiber optics.
- **Speed and Bandwidth**: Typically offers lower speeds compared to LANs due to long-distance data transmission, but modern WANs can achieve high speeds using advanced technologies.

- **Examples:**
  - The **Internet** is the largest WAN, connecting millions of devices globally.
  - A company's **corporate WAN** connects offices in different countries, allowing seamless data access and communication.



3. The Internet

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

3.1. Brief History

An internet (note the lowercase letter i) is two or more networks that can communicate with each other. The most notable internet is called the Internet (uppercase letter I), a collaboration of more than hundreds of thousands of interconnected networks. Private individuals as well as various organizations such as government agencies, schools, research facilities, corporations, and libraries in more than 100 countries use the Internet. Millions of people are users. Yet this extraordinary communication system only came into being in 1969.

In the mid-1960s, mainframe computers in research organizations were standalone devices. Computers from different manufacturers were unable to communicate with one another. The Advanced Research Projects Agency (ARPA) in the Department of Defense (DoD) was interested in finding a way to connect computers so that the researchers they funded could share their findings, thereby reducing costs and eliminating duplication of effort.

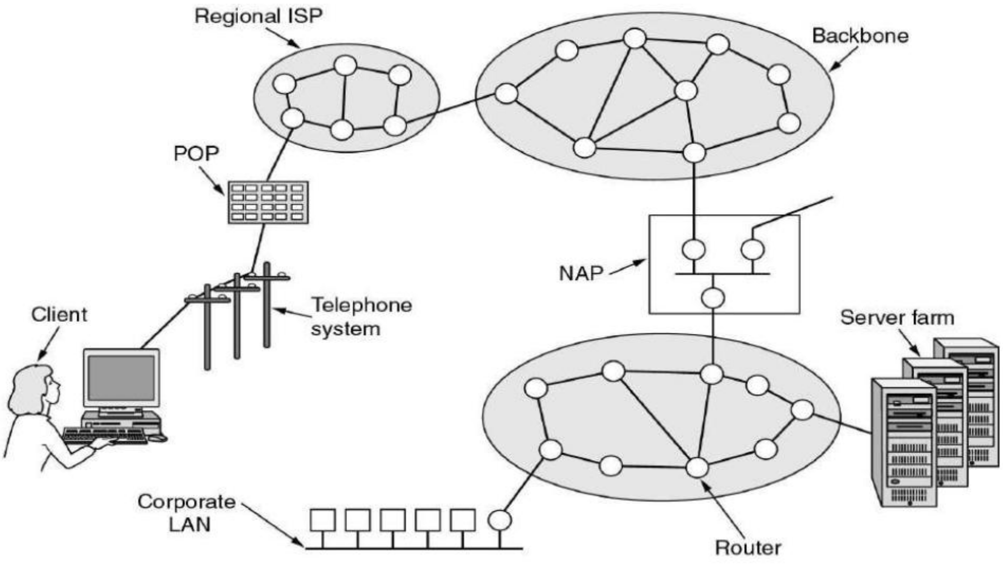
In 1967, at an Association for Computing Machinery (ACM) meeting, ARPA presented its ideas for ARPANET, a small network of connected computers. The idea was that each host computer (not necessarily from the same manufacturer) would be attached to a specialized computer, called an *inteiface message processor* (IMP). The IMPs, in tum, would be connected to one another. Each IMP had to be able to communicate with other IMPs as well as with its own attached host. By 1969, ARPANET was a reality. Four nodes, at the University of California at Los Angeles (UCLA), the University of California at Santa Barbara (UCSB), Stanford ResearchInstitute (SRI), and the University of Utah, were connected via the IMPs to form a network. Software called the *Network Control Protocol* (NCP) provided communication between the hosts.

In 1972, Vint Cerf and Bob Kahn, both of whom were part of the core ARPANET group, collaborated on what they called the *Internetting Projec1.* Cerf and Kahn's landmark 1973 paper outlined the protocols to achieve end-to-end delivery of packets. This paper on Transmission Control Protocol (TCP) included concepts such as encapsulation, the datagram, and the functions of a gateway.

Shortly thereafter, authorities made a decision to split TCP into two protocols: Transmission Control Protocol (TCP) and Internetworking Protocol (lP). IP would handle datagram routing while TCP would be responsible for higher-level functions such as segmentation, reassembly, and error detection. The internetworking protocol became known as TCP/IP.

3.2. The Internet Today

The Internet has come a long way since the 1960s. The Internet today is not a simple hierarchical structure. It is made up of many wide- and local-area networks joined by connecting devices and switching stations. It is difficult to give an accurate representation of the Internet because it is continually changing-new networks are being added, existing networks are adding addresses, and networks of defunct companies are being removed. Today most end users who want Internet connection use the services of Internet service providers (ISPs). There are international service providers, national service providers, regional service providers, and local service providers. The Internet today is run by private companies, not the government.



4. Protocols and standards

4.1. 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.

4.2. 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.

5. Network models

A network model is a conceptual structure which allows better understanding of how networking tasks are performed.

Two important network models:

1. OSI Reference Model
2. TCP/IP Reference Model

5.1. The OSI Reference Model

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

The OSI model has seven layers. The principles that were applied to arrive at the seven layers can be briefly summarized as follows:

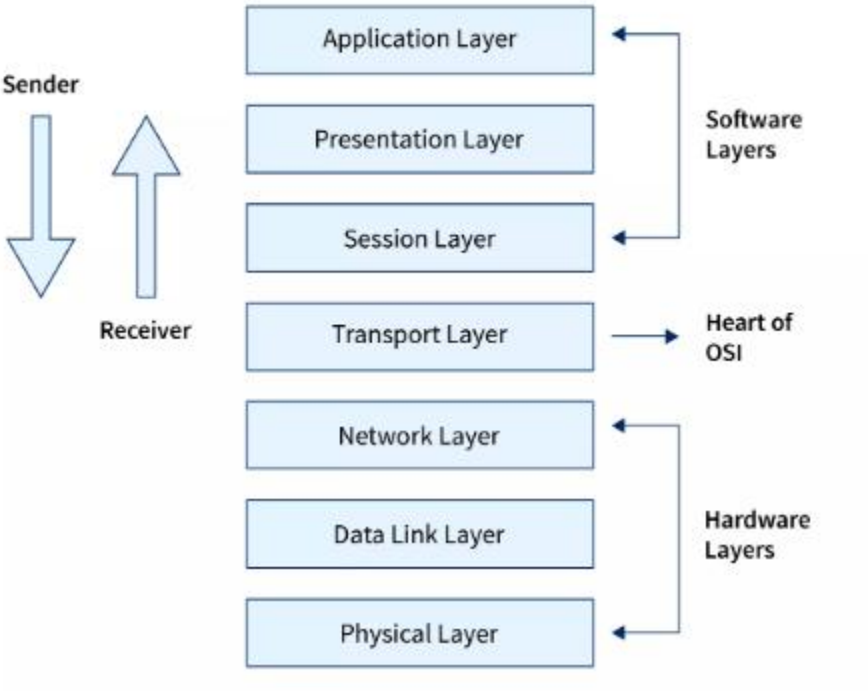
1. A layer should be created where a different abstraction is needed.

2. Each layer should perform a well-defined function.

3. The function of each layer should be chosen with an eye toward defining internationally standardized protocols.

4. The layer boundaries should be chosen to minimize the information flow across the interfaces.

5. The number of layers should be large enough that distinct functions need not be thrown together in the same layer out of necessity and small enough that the architecture does not become unwieldy.



The OSI reference model

- **The Physical Layer**

  It is connected with transmitting raw bits over a communication channel from one node to the next. It deals with physical devices required for data communications. The physical layer deals with the following issues:

  - *Mechanical:* Consider the physical properties of the medium and interfaces between devices like connectors, cables, etc.
  - *Signal representation:* How to represent data bits on the transmission medium encoding mechanisms.
  - *Timing:* Number of bits sent per second i.e. Transmission rate, bit duration.
  - *Synchronization:* Sequence of events, synchronization between sender and receiver, connection establishment, etc.
  - *Line configuration:* Connection of devices to communication links - point to point, multipoint, etc.
  - *Physical topology:* How to connect the devices - mesh, star, ring etc.

- **The Data Link Layer**

  This layer is responsible for error-free transmission of frames from one hop to the next. Its main task is to transform raw data bits into a frame that is free of transmission errors. It accomplishes this task by breaking the data into frames and takes care of identification of frames.

The main functions of the Data link layer are:

- *Framing:* The stream of bits is divided into logical units called frames.
- *Error control:* Provides mechanisms to detect or correct errors.
- *Flow control:* Ensures that a fast sender does not overwhelm a slow receiver.
- *Physical addressing:* Identifies a machine (sender as well. receiver) in the network using its physical address.
- *Multiple access control* for multipoint links, it provides mechanisms to access the shared communication channel between multiple machines.

- **Network Layer**

The network layer is responsible for the transfer of data packets from source to destination machines across the communication subnet i.e. across multiple networks.

The main functions of the Network layer are:

- *Addressing:* The network layer identities a machine on the basis of its logical address. The logical address identifies the network to which the machine belongs.
- *Routing:* A key design issue is to route packets from source to destination. Therefore routine algorithms are used.
- *Congestion control:* since the data packet has to travel through the communication subnet, congestion control is another function of this layer.

- **Transport Layer**

This is an end-to-end layer, which ensures process-to-process delivery from the source application to the destination application.

Its job is to ensure that the whole message arrives intact and in order, at the destination.

It isolates the upper layers from the lower layers so that the complexities, physical and logical characteristics of the subnet are hidden from the end users.

The main functions of the Transport layer are:

- *Port addressing:* The transport layer has to identify the application in the host machine for whom the data is intended. It does this by an address called port address.
- *Segmentation and reassembly:* A large message may be split up if needed. This layer takes care of sequencing and reassembly at the destination.
- *Services:* It also determines what type of service to provide to the upper layers, connection-less or connection-oriented.
- *Flow and Error control:* It provides end-to-end flow control and also ensures end-to-end error free delivery of data.

- **Session Layer**

  This layer allows users on different machines to establish sessions between them. Sessions offer various services like dialog control (halfduplex, full-duplex), token management (preventing two parties from performing the same critical operations simultaneously), authorization, synchronization, checkpoint mechanism to allow transmission to continue from the last point in case of a crash.

- **Presentation layer**

  This layer is concerned with the syntax and semantics of data transmitted i.e. it defines the format of data to be exchanged between applications and offers services like format transformation, encryption, compression etc.

- **Application Layer**

  This layer provides a means for user applications to access the network. It contains a variety of services commonly needed by users like file transfer, e-mail, remote terminal access, access to the World-Wide-Web etc one widely used application protocol is HTTP.

5.1.1. Layered tasks

- ***Definitions***
  - **Layer:** The network software is designed or organized as a series of layers or levels to reduce their design complexity. The number of layers, their names and functions vary from network to network.
  - **Services:** The purpose of each layer is :
    - Offer certain services to the higher layers.
    - Use the services of the layer below it.

    A layer on one machine carries on a conversation only with its corresponding layer on another machine that is layer 'n' on one machine communicates only with layer 'n' on another machine. The rules and conventions used for this conversation are known as **layer-n-protocol**.

    Corresponding layer entities on different machines are called **peers** i.e. peers communicate using a protocol.

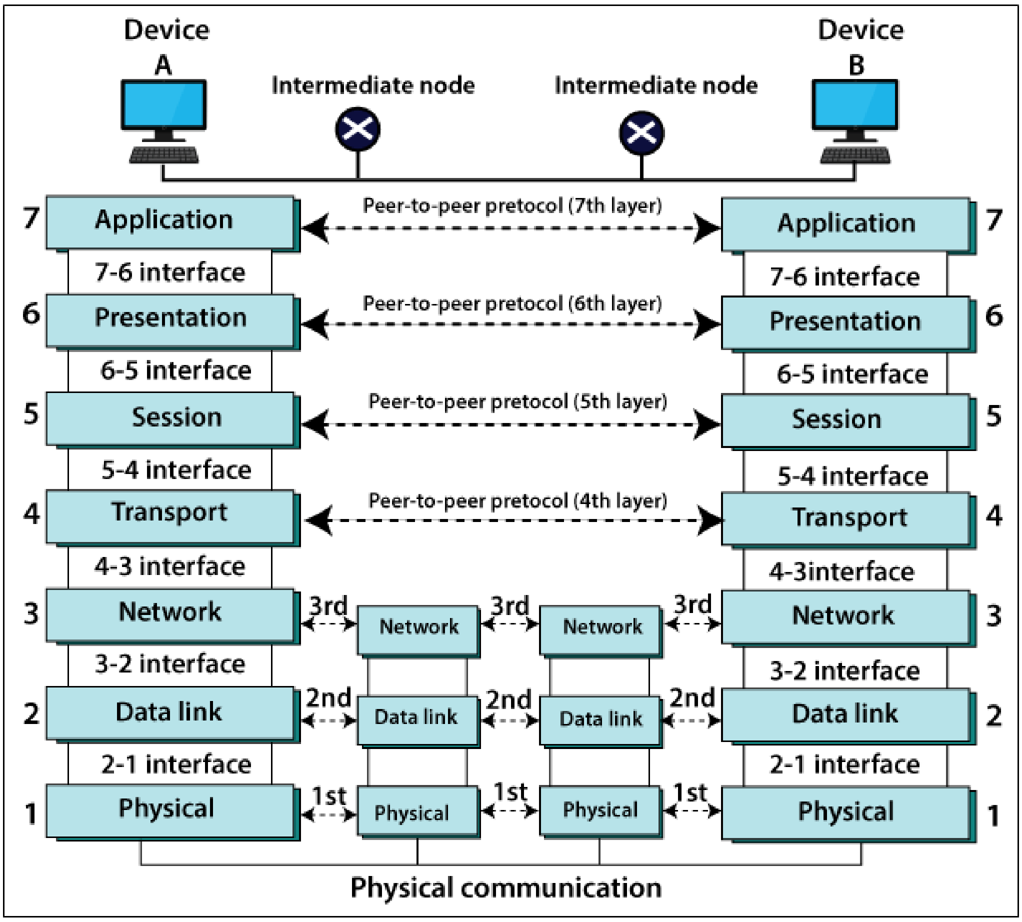
    The peers may negotiate on different parameters in the protocol like maximum packet size, timer values, etc.
  - **Interface:** Between each pair of adjacent layers, there is an interface. This defines which primitive operations and services the lower layer offers to the upper one. Interfaces should be clear-cut and the amount of Information passed between the layers should be minimum.

- **Relationship between services and protocol:** A service is a set of operations that a layer (service provider) provides to the layer above it (service user). These operations are specified in an abstract manner with no details of how to provide that service.

  On the other hand, a protocol specifies a set of rules to carry out communication between peers i.e. communicating entities in the same layer on different machines. Protocols clearly define "how to carry out the communication.

  Entities use protocols to implement services. Different layers use different protocols for communication.

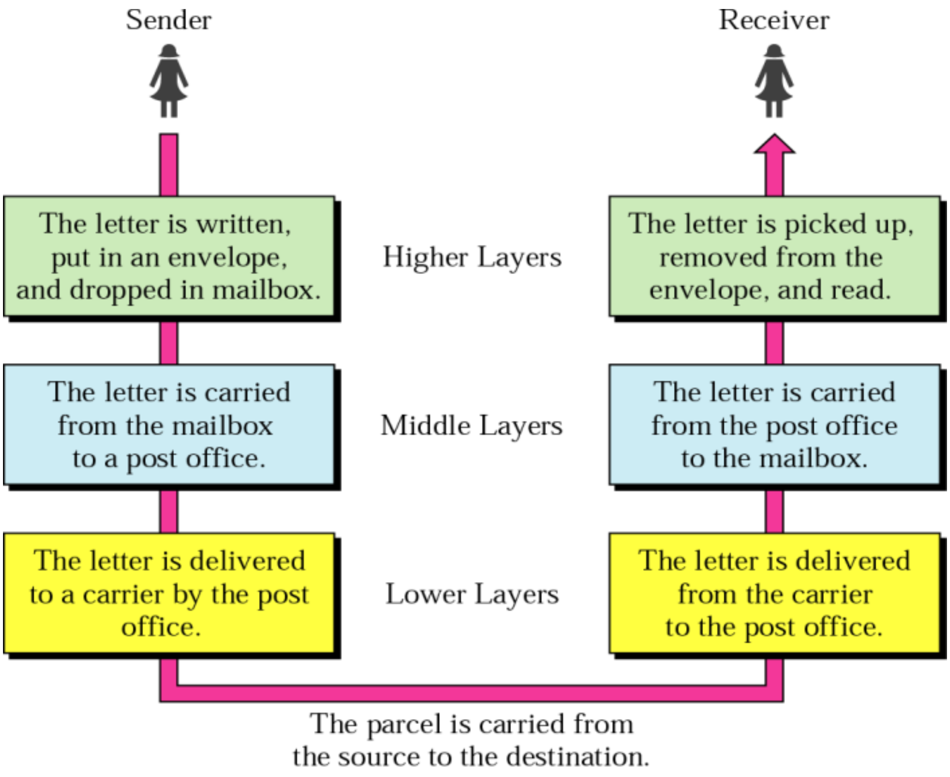
  Thus, services correspond to the interface between adjacent layers while protocols correspond to communication between peer entities in the corresponding or same layer on different machines.



Communication between layers

- **Example**

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


Sender
Receiver
The letter is written, put in an envelope, and dropped in mailbox.
Higher Layers
The letter is picked up, removed from the envelope, and read.
The letter is carried from the mailbox to a post office.
Middle Layers
The letter is carried from the post office to the mailbox.
The letter is delivered to a carrier by the post office.
Lower Layers
The letter is delivered from the carrier to the post office.
The parcel is carried from the source to the destination.

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

- *At the Sender Site:*

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

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

- *0n the Way:*

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

- *At the Receiver Site*

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

5.1.2. *Encapsulation and Layered Communication*

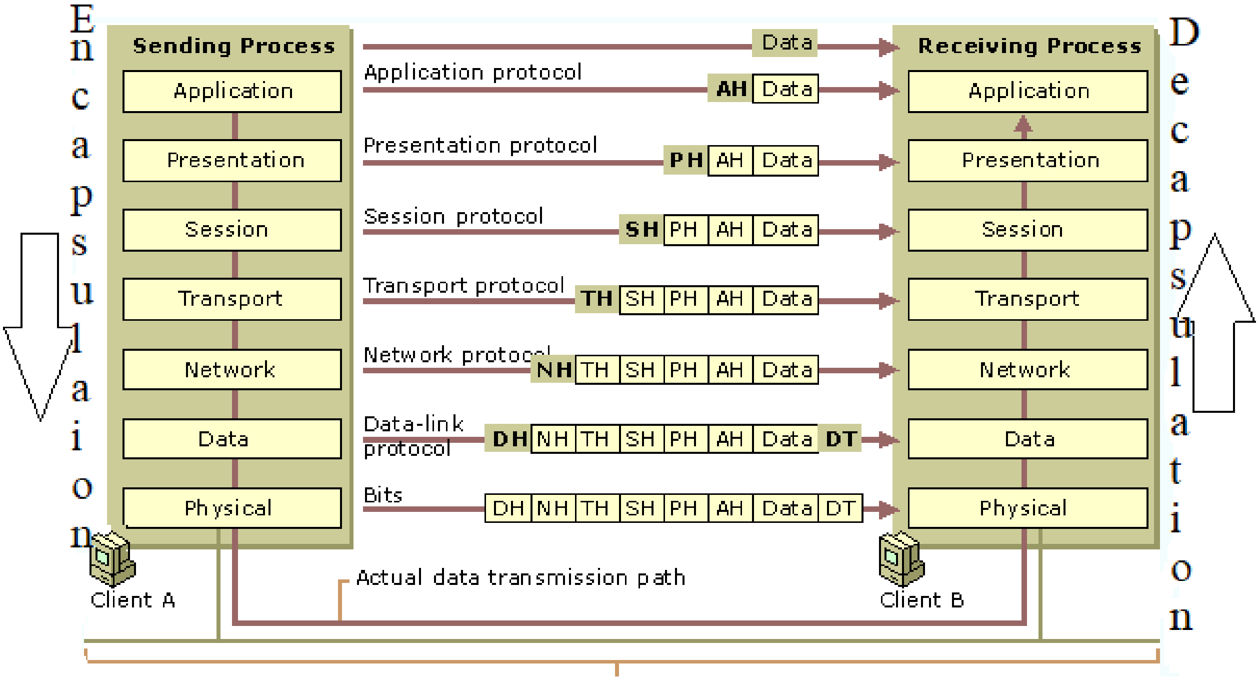
As data is passed from the user application down the virtual layers of the OSI model, each layer adds a **header** (and sometimes a **trailer**) containing protocol information specific to that layer. These headers are called **Protocol Data Units (PDUs),** and the process of adding these headers is called **encapsulation**.

The PDU of each layer is identified with a different term:

Layer	*PDU Name*
Application	-
Presentation	-
Session	-
Transport	**Segments**
Network	**Packets**
Data-Link	**Frames**
Physical	**Bits**

Each layer **communicates with the corresponding layer** on the receiving device. For example, on the sending device, source and destination hardware addressing is placed in a Data-link header. On the receiving device, that Data-link header is processed and stripped away (**decapsulated)** before being sent up to the Network and other upper layers.

The **Decapsulation** is the process of removing the header and trailer information from a packet, as it moves toward its destination. The destination device receives the data in its original form.



5.2. The TCP/IP Reference Model

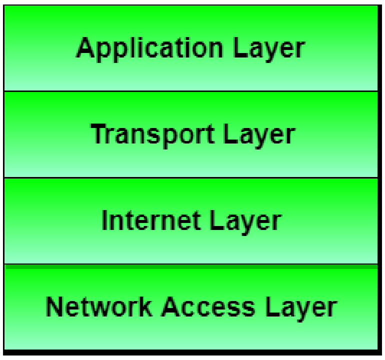
TCP/IP means Transmission Control Protocol and Internet Protocol. It is the network model used in the current Internet architecture as well.

- **History:** The TCP/IP model was built on the **ARPANET** modes which was developed by **Advanced Research Projects Agency (ARPA)** of the U.S. Dept. of Defense (DOD) in the late 1960's. ARPANET was very successful but was **unable to connect diverse networks**. This led to the development of a new architecture for internetworking.

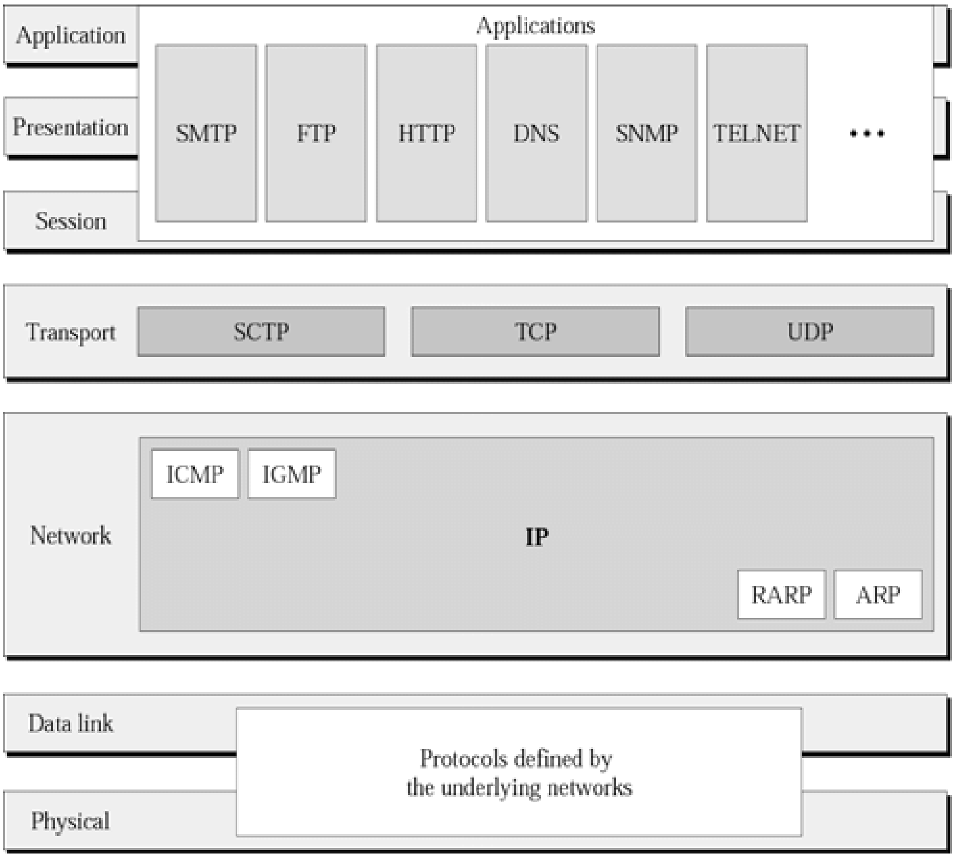
- **Goals:** The TCP/IP reference model was developed prior to OSI model. The major design goals of this model were,

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

- **Different Layers of TCP/IP Reference Model:** There are 4 layers that form the TCP/IP reference model:



- **TCP/IP protocol suite:** Protocols used in the various layers of TCP/IP model are as shown in following diagram:



- **Network Access Layer:**

- A network layer is the lowest layer of the TCP/IP.
- A network layer is the combination of the Physical layer and Data Link layer defined in the OSI reference model.
- It defines how the data should be sent physically through the network.
- This layer is mainly responsible for the transmission of the data between two devices on the same network.
- The functions carried out by this layer are encapsulating the IP datagram into frames transmitted by the network and mapping of IP addresses into physical addresses.
- The protocols used by this layer are ethernet, token ring, FDDI, X.25, frame relay.

- **Internet Layer:**

- An internet layer is the second layer of the TCP/IP model.
- An internet layer is also known as the network layer.
- The main responsibility of the internet layer is to send the packets from any network and they arrive at the destination irrespective of the route they take.
- Following are the protocols used in this layer:
  - **IP Protocol:** IP protocol is used in this layer and it is the most significant part of the entire TCP/IP suite. This protocol implements logical host addresses known as **IP addresses.** The IP addresses are used by the internet and higher layers to identify the device and to provide internetwork routing.
  - **ARP Protocol:**
  - ARP stands for Address Resolution Protocol.
  - ARP is a network layer protocol which is used to find the physical address from the IP address.
  - The two terms are mainly associated with the ARP Protocol:
    - **ARP request:** When a sender wants to know the physical address of the device, it broadcasts the ARP request to the network.
    - **ARP reply:** Every device attached to the network will accept the ARP request and process the request, but only recipient recognize the IP address and send back its physical address in the form of ARP reply. The recipient adds the physical address both to its cache memory and to the datagram header.
  - **ICMP Protocol:**
  - ICMP stands for Internet Control Message Protocol.
  - It is a mechanism used by the hosts or routers to send notifications regarding datagram problems back to the sender.

- An ICMP protocol mainly uses two terms:
  - **ICMP Test:** ICMP Test is used to test whether the destination is reachable or not.
  - **ICMP Reply:** ICMP Reply is used to check whether the destination device is responding or not.
- The core responsibility of the ICMP protocol is to report the problems, not correct them. The responsibility of the correction lies with the sender.

- **Transport Layer:** The transport layer is responsible for the reliability, flow control and correction of data which is being sent over the network. The two protocols used in the transport layer are User Datagram Protocol and Transmission Control Protocol.

➢ **User Datagram Protocol (UDP):**

- It provides connectionless service and end-to-end delivery of transmission.
- It is an unreliable protocol as it discovers the errors but not specify the error.
- User Datagram Protocol discovers the error and ICMP protocol reports the error to the sender that user datagram has been damaged.
- UDP consists of the following fields:
  - **Source port address:** The source port address is the address of the application program that has created the message.
  - **Destination port address:** The destination port address is the address of the application program that receives the message.
  - **Total length:** It defines the total number of bytes of the user datagram in bytes.
  - **Checksum:** The checksum is a 16-bit field used in error detection.
- UDP does not specify which packet is lost. UDP contains only checksum; it does not contain any ID of a data segment.

➢ **Transmission Control Protocol (TCP)**

- It provides a full transport layer services to applications.
- It creates a virtual circuit between the sender and receiver, and it is active for the duration of the transmission.
- TCP is a reliable protocol as it detects the error and retransmits the damaged frames. Therefore, it ensures all the segments must be received and acknowledged before the transmission is considered to be completed and a virtual circuit is discarded.
- At the sending end, TCP divides the whole message into smaller units known as segment, and each segment contains a sequence number which is required for reordering the frames to form an original message.
- At the receiving end, TCP collects all the segments and reorders them based on sequence numbers.

- **Application Layer:**

- An application layer is the topmost layer in the TCP/IP model.
- It is responsible for handling high-level protocols, issues of representation.
- This layer allows the user to interact with the application.
- When one application layer protocol wants to communicate with another application layer, it forwards its data to the transport layer.
- There is an ambiguity occurs in the application layer. Every application cannot be placed inside the application layer except those who interact with the communication system.
- For example: text editor cannot be considered in application layer while web browser using HTTP protocol to interact with the network where HTTP protocol is an application layer protocol.
- Following are the main protocols used in the application layer:
  - **HTTP:** HTTP stands for Hypertext transfer protocol. This protocol allows us to access the data over the World Wide Web. It transfers the data in the form of plain text, audio, video. It is known as a Hypertext transfer protocol as it has the efficiency to use in a hypertext environment where there are rapid jumps from one document to another.
  - **SMTP:** SMTP stands for Simple mail transfer protocol. The TCP/IP protocol that supports the e-mail is known as a Simple mail transfer protocol. This protocol is used to send the data to another e-mail address.
  - **DNS:** DNS stands for Domain Name System. An IP address is used to identify the connection of a host to the internet uniquely. But, people prefer to use the names instead of addresses. Therefore, the system that maps the name to the address is known as Domain Name System.
  - **TELNET:** It is an abbreviation for Terminal Network. It establishes the connection between the local computer and remote computer in such a way that the local terminal appears to be a terminal at the remote system.
  - **FTP:** FTP stands for File Transfer Protocol. FTP is a standard internet protocol used for transmitting the files from one computer to another computer.

5.3. Comparison of the OSI and TCP/IP Reference Models

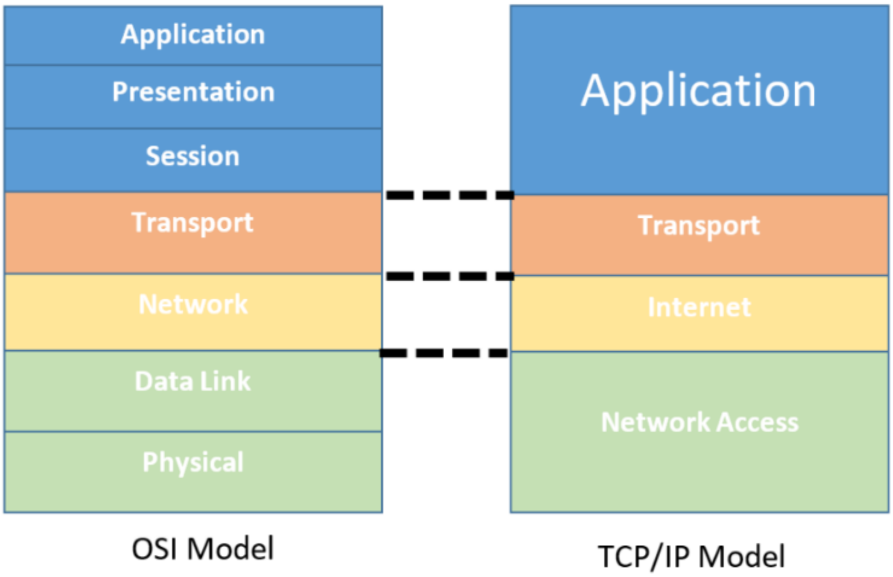
- **Similarities:**

- Both are layered network models based on the concept of an independent protocol stack.
- The functionalities of the Network, Transport and Application layers are similar.
- The higher layers are end-to-end layers, whereas the lower ones deal with host to intermediate processors.

- **Differences:**

- OSI has 7 while TCP/IP has 4 layers. The common layers are Network, Transport and Application layer.
- OSI model supports connection oriented and connection less service in the Network layer and only connection – oriented communication in the Transport layer.
- TCP/IP has only connectionless mode in the Internet layer but supports connection oriented and connectionless in the Transport layer.
- The OSI model makes a clear distinction between services, interfaces and protocols, whereas TCP/IP model does not clearly distinguish between them.
- The OSI reference model is a general purpose model. Hence, this model was not biased toward one particular set of protocols.
- In TCP/IP, the protocols were developed first and the model fits the protocol description. Hence, it does not fit any other protocol stack.
- OSI model does not take care of internetworking. The TCP/IP protocols were designed first and the main goal was to handle internetworking.
- The session and presentation layers in the OSI model are nearly empty whereas the data link and network layers are overfull.
- The TCP / IP model on the other hand, does not distinguish - nor mention physical and data link layers.
- The OSI model is very useful to discuss and understand - computer networks but the OSI model has not been implemented. In contrast, the protocols in the TCP/IP model are universally used.

- **Diagrammatic Comparison between OSI Reference Model and TCP/IP Reference Model**


Application
Presentation
Session
Transport
Network
Data Link
Physical
OSI Model
Application
Transport
Internet
Network Access
TCP/IP Model