Unit – I

1.1 DATA COMMUNICATIONS

Data communication 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.

Data communication refers to the exchange of data between a source/ sender and a destination/ receiver. Data communication is said to be local if communicating devices are in the same building or a similarly restricted geographical area.

Data communications are the exchange of data between two devices via some form of transmission medium 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). The effectiveness of a data communications system depends on four fundamental characteristics:

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

2. **Accuracy**. The system must deliver 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 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. For example, let us assume that video packets are sent every 30 ms. If some of the packets arrive with 30-ms delay and others with 40-ms delay, an uneven quality in the video is the result.

1.1.1 Components

A data communications system has five components:

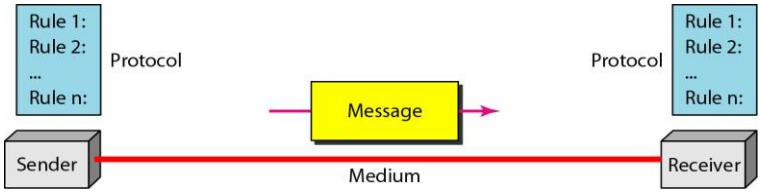


Figure 1.1 Five *components of data communication*

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.

1.1.2 Data Representation

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

1. ***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.

2. ***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.

3. ***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. 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.

4. ***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.

5. ***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.

1.1.3 Data Flow

Communication between two devices can be simplex, half-duplex, or full-duplex as shown in figure.

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 made (also, called duplex), 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. This sharing can occur in two ways: Either the link must contain two physically separate transmission paths, one for sending and the other for receiving; or the capacity of the channel is divided between signals travelling in both directions. One common example of full-duplex communication is the telephone network. The full-duplex mode is used when communication in both directions is required all the time.

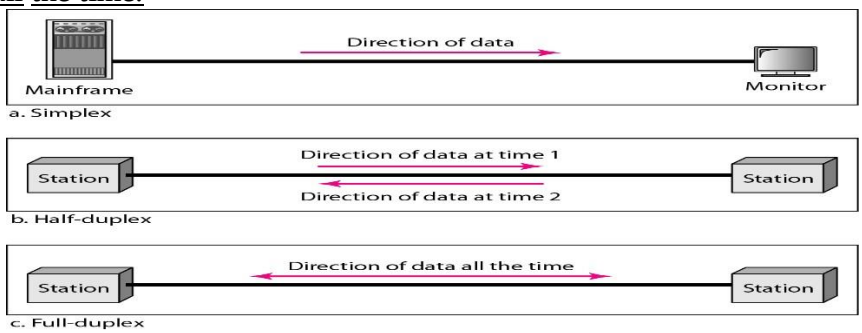


Figure 1.2 Data *flow (simplex, half-duplex, and full-duplex)*

1.2 Networks

A network is a group of two or more computer systems or other devices that are linked together to exchange data. Networks share resources, exchange files and electronic communications. For example, networked computers can share files or multiple computers on the network can share the same printer.

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.

1.2.1 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:

Performance can be measured by considering transit time and response time. **Transit time** is the amount of time required for a message to travel from one device to another. **Response time** is the elapsed time between an inquiry and a response.

The performance of a network depends on a number of factors, including the number of users, the type of transmission medium, the capabilities of the connected hardware, and the efficiency of the software. Performance is often evaluated by two networking metrics: throughput and delay.

B. 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.

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.

1.2.2 Physical Structures

Type of Connection

There are two possible types of connections:

1. **Point-to-Point connection:** it provides a dedicated link between two devices. The entire capacity of the link is reserved for transmission between those two devices. (see Figure 1.3a). Ex: a point-to-point connection between the remote control and the television's control system.
2. **Multipoint / multidrop connection:** it is one in which more than two specific devices share a single link (see Figure 1.3b). 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.

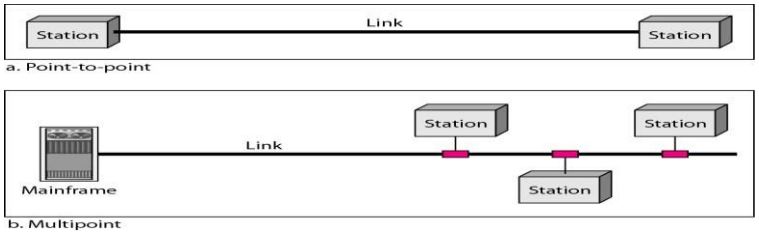


Figure 1.3 Types *of connections: point-to-point and multipoint*

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 /nodes to one another. There are four basic topologies possible: **mesh, star, bus, and ring** (see Fig 1.4).

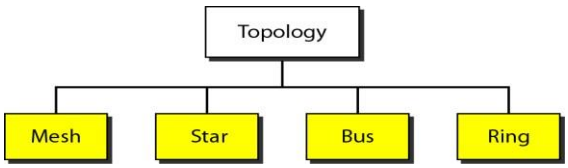


Figure 1.4 Categories *of topology*

A. Mesh topology

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. Number of physical links in a fully connected mesh network with *n* 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 a mesh topology, we need ***n(n -1) /2*** duplex-mode links.

One practical example of a mesh topology is the connection of telephone regional offices in which each regional office needs to be connected to every other regional office. It can also be used as backbone network.

Advantages:

1. Use of dedicated links eliminates traffic problems that can occur with shared links.
2. MT is Robust, If one link becomes unusable, it does not incapacitate the entire system.
3. Enables Privacy or security, message travels along a dedicated line and only intended recipient sees it.
4. Point-to-point links make fault identification easy. Traffic can be routed to avoid links with suspected problems.

Disadvantages:

1. Installation and reconnection are difficult, b'coz every device must be connected to every other device.
2. Sheer bulk of wiring require more accommodation space.

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

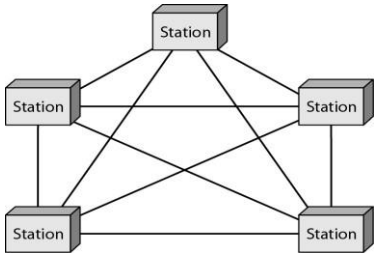


Figure 1.5 ***A fully connected mesh topology (five devices)***

B.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 (see Figure 1.6). Ex: The star topology is used in local-area networks (LANs).

Advantages:

1. Less Expensive, each device needs only one port and one link.
2. Easy to install and reconfigure.
3. Less cabling & additions, moves and deletions involve only one connection
4. Is robust: If one link fails only that is affected, others remain active.
5. As long as hub is working, it can be used to monitor link problems.

Disadvantages:

1. If the hub goes down, the whole system is dead.
2. Although a STAR requires less cabling than a mesh, it requires more cabling than some other topologies.

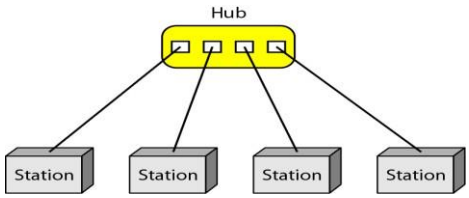


Figure 1.6 ***A star topology connecting four stations***

C. Bus Topology

A **bus topology,** is multipoint. One long cable acts as a **backbone** to link all the devices in a network (see Figure 1.7). 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. Ex: Ethernet LAN's.

Advantages

1. Ease of installation.
2. A bus uses less cabling than mesh or star topologies.

Disadvantages:

1. Difficult reconnection and fault isolation.
2. Difficult to add new devices as it may require modification or replacement of the backbone.
3. A fault in the bus cable stops all transmission. The damaged area reflects signals back in the originating direction, creating noise in both directions.

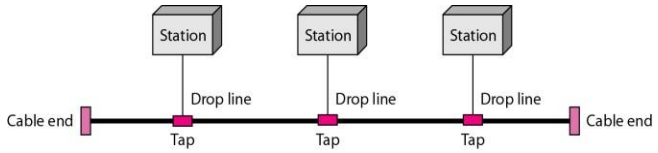


Figure 1.7 A *bus topology connecting three stations*

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 (see Figure 1.8).

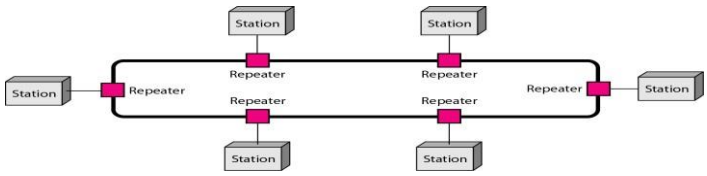


Figure 1.8 A *ring topology connecting six stations*

Advantages

1. Relatively easy to install and reconfigure.
2. Fault isolation is simplified, 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 location.

Disadvantages

1. Unidirectional traffic can be a disadvantage.
2. a break in the ring can disable the entire network. This can be solved by using a dual ring or a switch closing the break.

Hybrid Topology

A network can be hybrid. For example, we can have a main star topology with each branch
connecting several stations in a bus topology as shown in Figure 1.9.

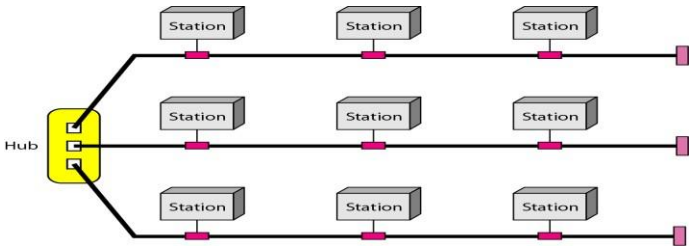


Figure 1.9 *A hybrid topology: a star backbone with three bus networks*

1.2.3 Categories of Networks

There are many types of computer networks. Common types of networks include the following:

Local-area network (LAN): The computers are geographically close together (that is, in the same building).

Wide-area network (WAN): The computers are farther apart and are connected by telephone lines or radio waves.

Metropolitan-area network (MAN): A data network designed for a town or city.

Home-area network (HAN): A network contained within a user's home that connects a person's digital devices.

Virtual private network (VPN): A network that is constructed by using public wires — usually the Internet — to connect to a private network, such as a company's internal network.

Storage area network (SAN): A high-speed network of storage devices that also connects those storage devices with servers.

Local Area Network

A local area network (LAN) is usually privately owned and links the devices in a single office, building, or campus (see Figure 1.10). Depending on the needs of an organization and the type of technology used, a LAN can be as simple as two PCs and a printer in someone's home office; or it can extend throughout a company and include audio and video peripherals. LAN size is limited to a few kilometers.

LANs are designed to allow resources to be shared between personal computers or workstations. The resources to be shared can include hardware (e.g., a printer), software (e.g., an application program), or data. LANs are distinguished from other types of networks by their size, transmission media and topology. LAN will use only one type of transmission medium. The most common LAN topologies are bus, ring, and star. It support normally 100 or 1000 Mbps speed.

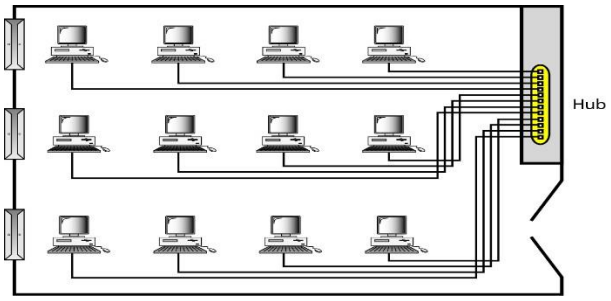


Figure 1.10 ***An isolated LAN connecting 12 computers to a hub in a closet***

WAN

A WAN is a data communication system spanning states, countries, or the whole world. A WAN can be a switched WAN or a Point-to-point WAN. Switched WAN connects the end systems, which comprises of a router that connects to another LAN or WAN (Ex: ATM Networks). Point-to-point WAN is a leased line from a telephone or a cable TV provider that connects a home computer or small LAN to an ISP.

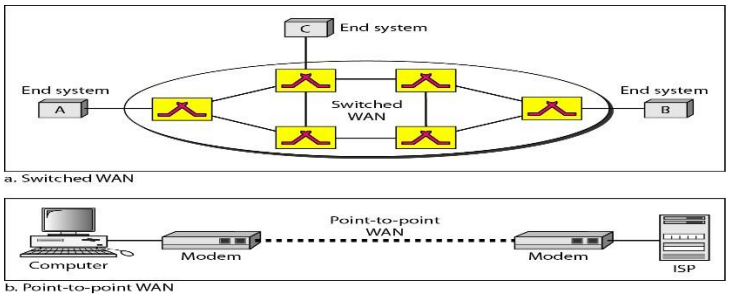


Figure 1.11 ***WANs: a switched WAN and a point-to-point WAN***

Metropolitan Area Networks

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

2.2.4 Interconnection of Networks: Internetwork

An internet is a network of networks or is a collection of many separate networks. i.e When two or more networks are connected, they become an internetwork, or internet.

As an example, consider an organization that has two offices, one on the east with a Star topology LAN and the other on the west with a Bus topology LAN. A switched WAN is used as a backbone WAN for connecting these LAN's to the president's computer. Three point-to-point WAN's are required to connect the LAN's to the switched WAN. These point-to-point WAN's can be a high speed DSL line or a cable modem line.

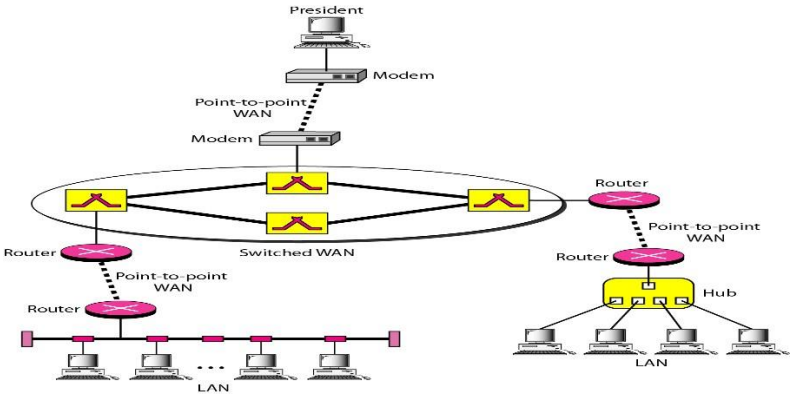


Figure 1.12 A *heterogeneous network made of four WANs and two LANs*

1.3 THE INTERNET

The Internet is a structured, organized system.

1.3.1 A Brief History

An internet (lowercase letter i) is two or more networks that can communicate with each other. The Internet (uppercase letter I), a collaboration of more than hundreds of thousands of interconnected networks.

In the mid-1960s, mainframe computers in research organizations were stand-alone devices. Computers from different manufacturers were unable to communicate with one another. The **Advanced Research Projects Agency** (ARPA) in the Department of Defense (DoD).

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 **interface message processor (IMP).** The IMPs, in turn, 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 Research Institute (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 Project. Cerf and Kahn's land-mark 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.

Authorities made a decision to split TCP into two protocols: Transmission Control Protocol (TCP) and Internetworking Protocol (IP). 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.

1.3.2 The Internet Today

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. The figure shows a conceptual (not geographic) view of the Internet.

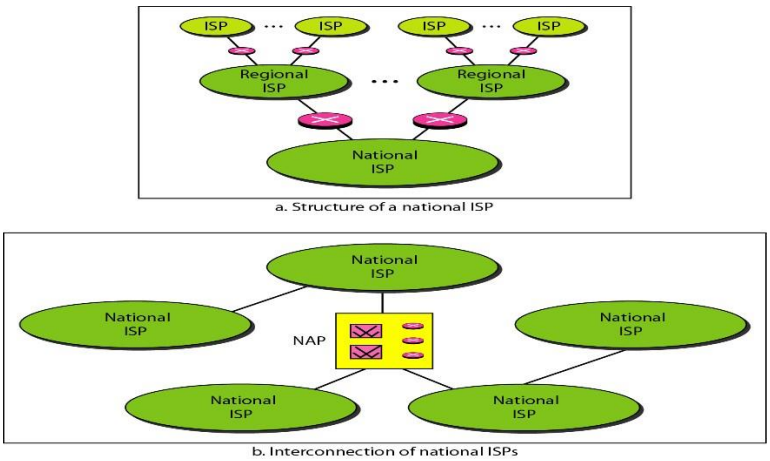


Figure 1.13 ***Hierarchical organization of the Internet***

International Internet Service Providers

At the top of the hierarchy are the international service providers that connect nations together.

National Internet Service Providers

The national Internet service providers are backbone networks created and maintained by specialized companies. To provide connectivity between the end users, these backbone networks are connected by complex switching stations (normally run by a third party) called **network access points** (NAPs). Some national ISP networks are also connected to one another by private switching stations called peering points. These normally operate at a high data rate.

Regional Internet Service Providers

Regional internet service providers or regional ISPs are smaller ISPs that are connected to one or more national ISPs. They are at the third level of the hierarchy with a smaller data rate.

Local Internet Service Providers

Local Internet service providers provide direct service to the end users. The local ISPs can be connected to regional ISPs or directly to national ISPs. Most end users are connected to the local ISPs.

1.4 PROTOCOLS AND STANDARDS

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 is it communicated, and when is it 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.

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?.

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"): 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 (meaning "by law" or "by regulation"): Those standards that have been legislated by an officially recognized body are de jure standards.

Network standards are important to ensure that hardware and software can work together. Without standards you could not easily develop a network to share information. Networking standards can be categorized in one of two ways**: formal** and **de facto (informal).**

Formal standards are developed by industry organizations or governments. Formal standards exist for network layer software, data link layer, hardware and so on. Formal standardization is a lengthy process of developing the specification, identifying choices and industry acceptance.

The second category of networking standards is de facto standards. These standards typically emerge in the marketplace and are supported by technology vendors but have no official backing. For example, Microsoft Windows is a de facto standard, but is not formally recognized by any standards organization. It is simply widely recognized and accepted.

Standards Organizations

Standards are developed through the cooperation of standards creation committees, forums, and government regulatory agencies.

Standards Creation Committees

While many organizations are dedicated to the establishment of standards, data telecommunications in North America rely primarily on those published by the following:

1. **International Organization for Standardization (ISO)**: The ISO is a multinational body whose membership is drawn mainly from the standards creation committees of various governments throughout the world. The ISO is active in developing cooperation in the realms of scientific, technological, and economic activity.

2. **International Telecommunication Union -Telecommunication Standards Sector (ITU-T)**: By the early 1970s, a number of countries were defining national standards for telecommunications, but there was still little international compatibility. The United Nations responded by forming, as

part of its International Telecommunication Union (ITU), a committee, the Consultative Committee for International Telegraphy and Telephony (CCITT). This committee was devoted to the research and establishment of standards for telecommunications in general and for phone and data systems in particular. On March 1, 1993, the name of this committee was changed to the International Telecommunication Union - Telecommunication Standards Sector (ITU-T).

3. **American National Standards Institute (ANSI):** the American National Standards Institute is a completely private, nonprofit corporation not affiliated with the U.S. federal government. However, all ANSI activities are undertaken with the welfare of the United States and its citizens occupying primary importance.

4. **Institute of Electrical and Electronics Engineers (IEEE):** The Institute of Electrical and Electronics Engineers is the largest professional engineering society in the world. International in scope, it aims to advance theory, creativity, and product quality in the fields of electrical engineering, electronics, and radio as well as in all related branches of engineering. As one of its goals, the IEEE oversees the development and adoption of international standards for computing and communications.

5. **Electronic Industries Association (EIA):** Aligned with ANSI, the Electronic Industries Association is a nonprofit organization devoted to the promotion of electronics manufacturing concerns. Its activities include public awareness education and lobbying efforts in addition to standards development. In the field of information technology, the EIA has made significant contributions by defining physical connection interfaces and electronic signaling specifications for data communication.

Forums

Telecommunications technology development is moving faster than the ability of standards committees to ratify standards. Standards committees are procedural bodies and by nature slow-moving. To accommodate the need for working models and agreements and to facilitate the standardization process, many special-interest groups have developed forums made up of representatives from interested corporations. The forums work with universities and users to test, evaluate, and standardize new technologies. By concentrating their efforts on a particular technology, the forums are able to speed acceptance and use of those technologies in the telecommunications community. The forums present their conclusions to the standards bodies.

Regulatory Agencies

All communications technology is subject to regulation by government agencies such as the

Federal Communications Commission (FCC) in the United States. The purpose of these agencies is to protect the public interest by regulating radio, television, and wire/cable communications. The FCC has authority over interstate and international commerce as it relates to communications.

Internet Standards

An Internet standard is a thoroughly tested specification that is useful to and adhered to by those who work with the Internet. It is a formalized regulation that must be followed. There is a strict procedure by which a specification attains Internet standard status. A specification begins as an Internet draft. An Internet draft is a working document (a work in progress) with no official status and a 6-month lifetime. Upon recommendation from the Internet authorities, a draft may be published as a Request for Comment (**RFC**). Each RFC is edited, assigned a number, and made available to all interested parties. RFCs go through maturity levels and are categorized according to their requirement level.

Network Models

The two best-known standards are the OSI model and the Internet model. The OSI (Open Systems Interconnection) model defines a seven-layer network; the Internet model defines a five-layer network.

1.5 LAYERED TASKS

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. The figure shows the steps in this task.

Sender, Receiver, and Carrier

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

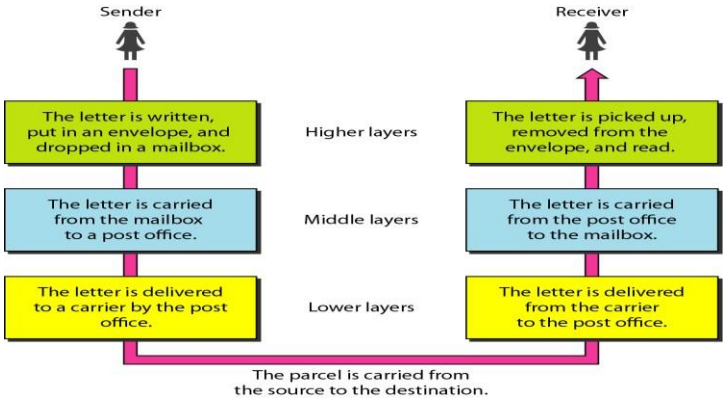


Figure 2.1 Tasks involved in sending a letter

At the Sender Site

The activities that take place at the sender site, in order, are:

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 carder transports the letter.

On the Way

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

At the Receiver Site

Lower layer. The carrier transports the letter to the post office.

Middle layer. The letter is sorted and delivered to the recipient's mailbox.

Higher layer. The receiver picks up the letter, opens the envelope, and reads it.

Hierarchy

According to our analysis, there are three different activities at the sender site and another three activities at the receiver site. The task of transporting the letter between the sender and the receiver is done by the carrier. Something that is not obvious immediately is that the tasks must be done in the order given in the hierarchy. At the sender site, the letter must be written and dropped in the mailbox before being picked up by the letter carrier and delivered to the post office. At the receiver site, the letter must be dropped in the recipient mailbox before being picked up and read by the recipient.

Services

Each layer at the sending site uses the services of the layer immediately below it. The sender at the higher layer uses the services of the middle layer. The middle layer uses the services of the lower layer. The lower layer uses the services of the carrier.

1.6 THE OSI MODEL

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.

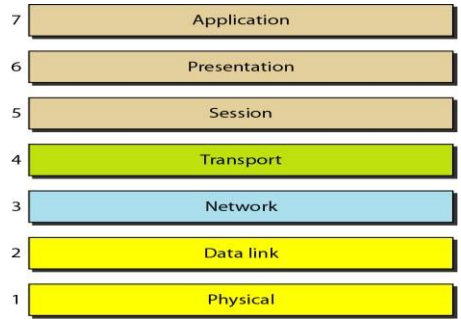


Figure 2.2 ***Seven layers of the OSI model***

Layered Architecture

The OSI model is composed of seven ordered layers: physical (layer 1), data link (layer 2), network
(layer 3), transport (layer 4), session (layer 5), and presentation (layer 6), and application (layer 7).

The following figure shows the layers involved when a message is sent from device A to device B. As the message travels from A to B, it may pass through many intermediate nodes. These intermediate nodes usually involve only the first three layers of the OSI model. Within a single machine, each layer calls upon the services of the layer just below it. Layer 3, for example, uses the services provided by layer
2 and provides services for layer 4.

Between machines, layer x on one machine communicates with layer x on another machine. This communication is governed by an agreed-upon series of rules and conventions called **protocols**. The
processes on each machine that communicate at a given layer are called **peer-to-peer processes**.

Peer-to-Peer Processes

At the physical layer, communication is direct: In the figure below, device A sends a stream of bits
to device B (through intermediate nodes). At the higher layers, however, communication must move down through the layers on device A, over to device B, and then back up through the layers. Each layer in the sending device adds its own information to the message it receives from the layer just above it and passes the whole package to the layer just below it. At layer 1 the entire package is converted to a form that can be transmitted to the receiving device. At the receiving machine, the message is unwrapped layer by layer, with each process receiving and removing the data meant for it. For example, layer 2 removes the data meant for it, and then passes the rest to layer 3. Layer 3 then removes the data meant for it and passes the rest to layer 4, and so on.

The following figure gives an overall view of the OSI layers. The process starts at layer 7 (the application layer), then moves from layer to layer in descending, sequential order. At each layer, a

header, or possibly a **trailer**, can be added to the data unit. Commonly, the trailer is added only at layer 2. When the formatted data unit passes through the physical layer (layer 1), it is changed into an electromagnetic signal and transported along a physical link. Upon reaching its destination, the signal passes into layer 1

and is transformed back into digital form. The data units then move back up through the OSI layers. As each block of data reaches the next higher layer, the headers and trailers attached to it at the corresponding sending layer are removed, and actions appropriate to that layer are taken. By the time it reaches layer 7, the message is again in a form appropriate to the application and is made available to the recipient.

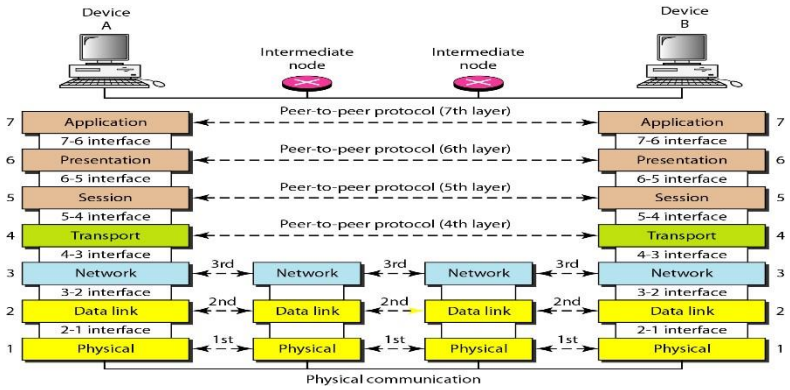


Figure 2.3 The *interaction between layers in the OSI model*

Interfaces between Layers

The passing of the data and network information down through the layers of the sending device and back up through the layers of the receiving device is made possible by an interface between each pair of adjacent layers. Each interface defines the information and services a layer must provide for the layer above it.

Organization of the Layers

The seven layers can be thought of as belonging to three subgroups. Layers 1, 2, and 3 - physical, data link, and network - are the network support layers; they deal with the physical aspects of moving data from one device to another (such as electrical specifications, physical connections, physical addressing, and transport timing and reliability). Layers 5, 6, and 7 - session, presentation, and application – are the user support layers; they allow interoperability among unrelated software systems. Layer 4, the transport layer, links the two subgroups. The upper OSI layers are almost always implemented in software; lower layers are a combination of hardware and software, except for the physical layer, which is mostly hardware.

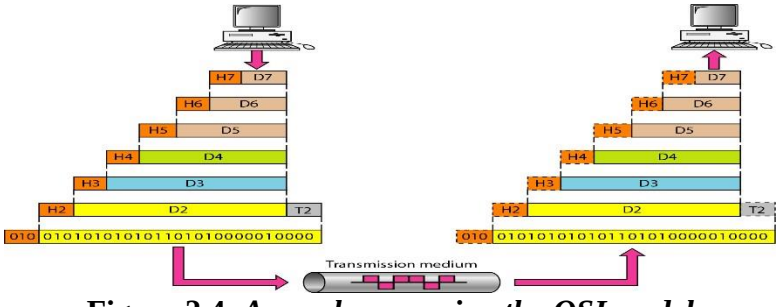


Figure 2.4 *An exchange using the OSI model*

Encapsulation

In the above figure, A packet (header and data) at level 7 is encapsulated in a packet at level 6. The whole packet at level 6 is encapsulated in a packet at level 5, and so on. In other words, the data portion of a packet at level N - 1 carries the whole packet (data and header and maybe trailer) from level N. The concept is called **encapsulation**; level N - 1 is not aware of which part of the encapsulated packet is data and which part is the header or trailer. For level N- 1, the whole packet coming from level N is treated as one integral unit.

1.7 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 following figure shows the position of the physical layer with respect to the transmission medium and the data link layer. **The physical layer is responsible for movements of individual bits from one hop (node) to the next.**

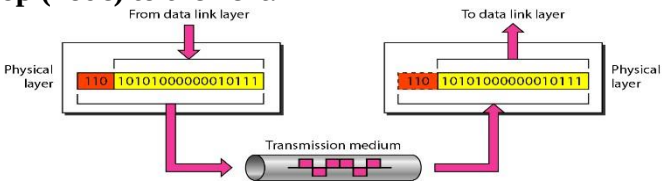


Figure 2.5 Physical *layer*

The physical layer is also concerned with the following:

1. **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.
2. **Representation of bits**. The physical layer data consists of a stream of bits (sequence of 0s or ls). To be transmitted, bits must be encoded into signals - electrical or optical. The physical layer defines the type of encoding (how 0s and 1s are changed to signals).
3. **Data rate**. The transmission rate--the number of bits sent each second.
4. **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.
5. **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.
6. **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).
7. **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 (duplex) mode, two devices can send and receive at the same time.

2. Data Link Layer

The data link layer transforms a raw transmission facility, to a reliable link. It makes the physical layer appear error-free to the upper layer (network layer). The figure shows the relationship of the data link layer to the network and physical layers. **The data link layer is responsible for moving frames from one hop (node) to the next.**

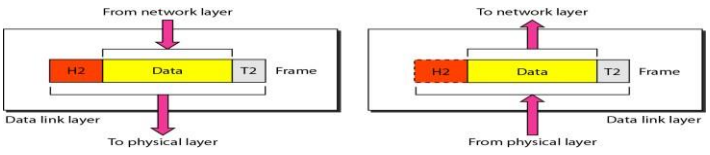


Figure 2.6 *Data link layer*

Other responsibilities of the data link layer include the following:

1. **Framing**. The data link layer divides the stream of bits received from the network layer into manageable data units called **frames.**
2. **Physical addressing**. the data link layer adds a header to the frame, that contains the sender and/or receiver address of the frame.
3. **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.
4. **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.
5. **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.

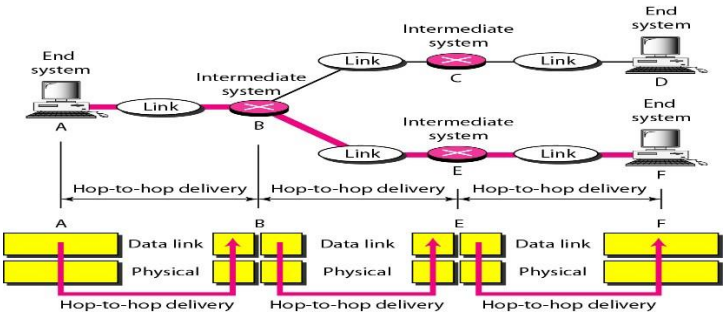


Figure 2.7 ***Hop-to-hop delivery***

The figure illustrates hop-to-hop (node-to-node) delivery by the data link layer. Communication at the data link layer occurs between two adjacent nodes. To send data from A to F, three partial deliveries are made. First, the data link layer at A sends a frame to the data link layer at B (a router). Second, the data link layer at B sends a new frame to the data link layer at E. Finally, the data link layer at E sends a new frame to the data link layer at F.

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. The figure shows the relationship of the network layer to the data link and transport layers. **The network layer is responsible for the delivery of individual packets from the source host to the destination host.**

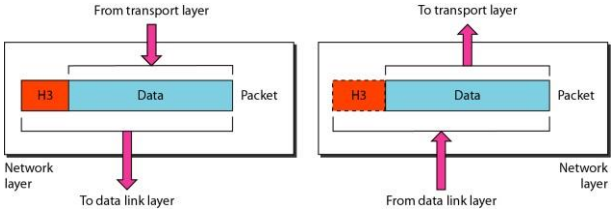


Figure 2.8 Network ***layer***

Other responsibilities of the network layer include the following:

1. **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.

2. **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.

The figure illustrates end-to-end delivery by the network layer.

The network layer at A sends the packet to the network layer at B. When the packet arrives at router B, the router makes a decision based on the final destination (F) of the packet. As we will see in later chapters, router B uses its routing table to find that the next hop is router E. The network layer at B, therefore, sends the packet to the network layer at E. The network layer at E, in turn, sends the packet to the network layer at F.

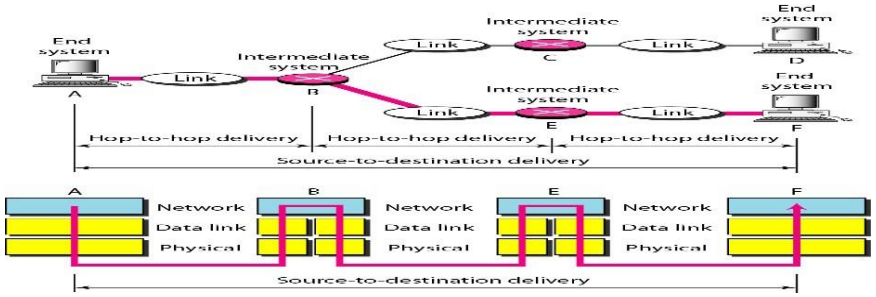


Figure 2.9 ***Source-to-destination delivery***

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.

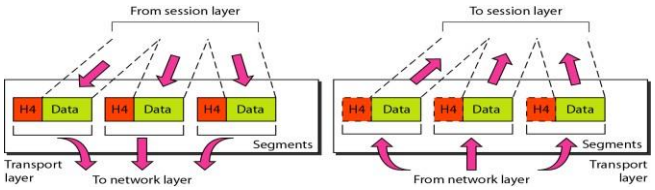


Figure 2.10 ***Transport layer***

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. The figure shows the relationship of the transport layer to the network and session layers. **The transport layer is responsible for the delivery of a message from one process to another.**

Other responsibilities of the transport layer include the following:

1. **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 **service-point address (or port address)**. The network layer gets each packet to the correct computer whereas the transport layer gets the entire message to the correct process on that computer.

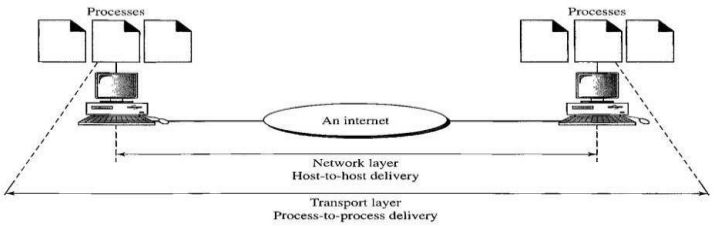
2. **Segmentation and reassembly**. A message is divided into transmittable segments, with each segment containing a sequence number. These numbers enable the transport layer to reassemble the message correctly upon arriving at the destination and to identify and replace packets that were lost in transmission.

3. **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.

4. **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.

5. **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. The figure illustrates process-to-process delivery by the transport layer.



Session Layer

The session layer is the network dialog controller. It establishes, maintains, and synchronizes the interaction among communicating systems. **The session layer is responsible for dialog control and synchronization.** Specific responsibilities of the session layer include the following:

1. **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.

2. **Synchronization**. The session layer allows a process to add checkpoints, or synchronization points, to a stream of data. For example, if a system is sending a file of 2000 pages, it is advisable to insert checkpoints after every 100 pages to ensure that each 100-page unit is received and acknowledged independently. In this case, if a crash happens during the transmission of page 523, the only pages that need to be resent after system recovery are pages 501 to 523. Pages previous to 501 need not be resent.

The figure illustrates the relationship of the session layer to the transport and presentation layers.

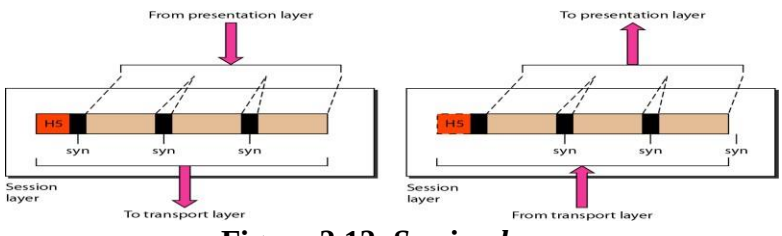


Figure 2.12 ***Session layer***

Presentation layer

The presentation layer is concerned with the **syntax and semantics of the information** exchanged between two systems. The figure shows the relationship between the presentation layer and the application and session layers. **The presentation layer is responsible for translation, compression, and encryption.**

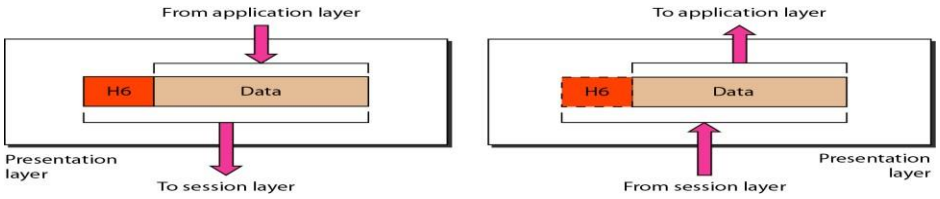


Figure 2.13 ***Presentation layer***

Specific responsibilities of the presentation layer include the following:

1. **Translation.** The processes 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.

2. **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.

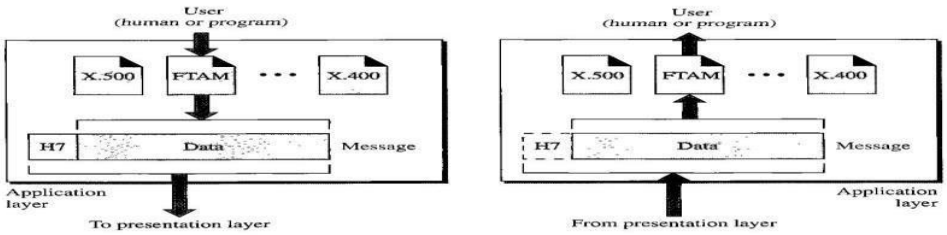
3. **Compression.** Data compression reduces the number of bits contained in the information. Data compression becomes particularly important in the transmission of multimedia such as text, audio, and video.

Application Layer

The application layer is responsible for providing services to the user. 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. The figure shows the relationship of the application layer to the user and the presentation layer.

Specific services provided by the application layer include the following:

1. **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. To do so, the application creates a software emulation of a terminal at the remote host. The user's computer talks to the software terminal which, in turn, talks to the host, and vice versa. The remote host believes it is communicating with one of its own terminals and allows the user to log on.
2. **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.
3. **Mail services**. This application provides the basis for e-mail forwarding and storage.
4. **Directory services**. This application provides distributed database sources and access for global information about various objects and services.



1.8 TCP/IP PROTOCOL SUITE

The TCP/IP protocol suite is made of five layers: physical, data link, network, transport, and application. The first four layers provide physical standards, network interfaces, internetworking, and transport functions that correspond to the first four layers of the OSI model. The three topmost layers in the OSI model, are represented in TCP/IP by a single layer called the application layer.

TCP/IP is a hierarchical protocol made up of interactive modules, each of which provides a specific functionality. 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.

Physical and Data Link Layers

At the physical and data link layers, TCP/IP does not define any specific protocol. It supports all the standard and proprietary protocols. A network in a TCP/IP internetwork can be a local-area network or a wide-area network.

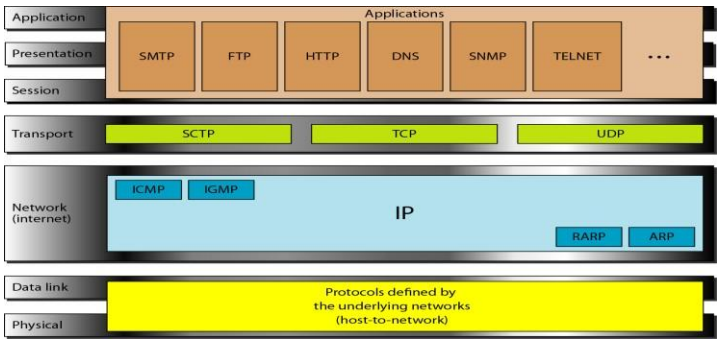


Figure 2.16 ***TCP/IP and OSI model***

Network Layer

At the network layer (internetwork layer), TCP/IP supports the Internetworking Protocol. IP, in turn, uses four supporting protocols: ARP, RARP, ICMP, and IGMP.

Internetworking Protocol (IP)

The Internetworking Protocol (IP) is the transmission mechanism used by the TCP/IP protocols. It is an unreliable and connectionless protocol - a best-effort delivery service. The term best effort means that IP provides no error checking or tracking. IP assumes the unreliability of the underlying layers and does its best to get a transmission through to its destination, but with no guarantees.

IP transports data in packets called datagram, each of which is transported separately. Datagram can travel along different routes and can arrive out of sequence or be duplicated. IP does not keep track of the routes and has no facility for reordering datagram once they arrive at their destination.

1. **Address Resolution Protocol:** The Address Resolution Protocol (ARP) is used to associate a logical address with a physical address. In physical network, each device on a link is identified by a physical address, usually imprinted on the network interface card (NIC). ARP is used to find the physical address of the node when its Internet address is known.

2. **Reverse Address Resolution Protocol:** The Reverse Address Resolution Protocol (RARP) allows a host to discover its Internet address when it knows only its physical address. It is used when a computer is connected to a network for the first time or when a diskless computer is booted.
3. **Internet Control Message Protocol:** The Internet Control Message Protocol (ICMP) is a mechanism used by hosts and gateways to send notification of datagram problems back to the sender. ICMP sends query and error reporting messages.
4. **Internet Group Message Protocol:** The Internet Group Message Protocol (IGMP) is used to facilitate the simultaneous transmission of a message to a group of recipients.

Transport Layer

The transport layer was represented in TCP/IP by two protocols: TCP and UDP. IP is a host-to-host protocol, meaning that it can deliver a packet from one physical device to another. **UDP and TCP** are transport level protocols responsible for delivery of a message from a process (running program) to another process. **SCTP**, has been devised to meet the needs of some newer applications.

1. ***User Datagram Protocol:*** It is a process-to-process protocol that adds only port addresses, checksum error control, and length information to the data from the upper layer.
2. ***Transmission Control Protocol:*** TCP provides full transport-layer services to applications. TCP is a reliable stream transport protocol. The term stream, means connection-oriented: A connection must be established between both ends of a transmission before either can transmit data.

At the sending end of each transmission, TCP divides a stream of data into smaller units called segments. Each segment includes a sequence number for reordering after receipt, together with an acknowledgment number for the segments received. Segments are carried across the internet inside of IP datagrams. At the receiving end, TCP collects each datagram as it comes in and reorders the transmission based on sequence numbers.

3. ***Stream Control Transmission Protocol:*** SCTP provides support for newer applications such as voice over the Internet. It is a transport layer protocol that combines the best features of UDP and TCP.

Application Layer

The application layer in TCP/IP is equivalent to the combined session, presentation, and application layers in the OSI model. Many protocols are defined at this layer.

1. ***Telnet:*** it provides remote login facilities. Telnet is an application layer protocol used on the Internet or local area networks to provide a bidirectional interactive text-oriented communication facility using a virtual terminal connection. Telnet provided access to a command-line interface on a remote host, including most network equipment and operating systems with a configuration utility.
2. ***File Transfer Protocol (FTP), Trivial File Transfer Protocol(TFTP):***
   A. **File Transfer Protocol (FTP):** is a standard network protocol used to transfer computer files between a client and server on a computer network. FTP is built on client-server model architecture and uses separate control and data connections between the client and the server.

   B. **Trivial File Transfer Protocol (TFTP):** is an Internet software utility for transferring files that is simpler to use than the File Transfer Protocol (FTP) but less capable. It is used where user authentication and directory visibility are not required. TFTP uses the User Datagram Protocol (UDP) rather than the Transmission Control Protocol (TCP).

3. ***Simple Mail Transfer Protocol (SMTP):*** is an Internet standard for electronic mail (email) transmission. SMTP by default uses TCP port 25. SMTP connections secured by SSL, known as SMTPS
4. ***Domain Name System (DNS):*** is a hierarchical decentralized naming system for computers, services, or any resource connected to the Internet. By providing a worldwide, distributed directory service, the Domain Name System is an essential component of the functionality of the Internet.
5. ***BOOTP:*** The Bootstrap Protocol (BOOTP) is a computer networking protocol used in Internet networks to automatically assign an IP address to network devices from a configuration server.
6. ***Simple Network Management Protocol (SNMP),*** is a popular protocol for network management. It is used for collecting information from, and configuring, network devices, such as servers, printers, hubs, switches, and routers on an Internet Protocol (IP) network.
7. **Common Management Information Protocol (CMIP)** It provides an implementation for the services defined by the Common Management Information Service (CMIS), allowing communication between network management applications and management agents.

1.9 ADDRESSING

Four levels of addresses are used in an internet employing the *TCP/IP* protocols: physical (link) addresses, logical (IP) addresses, port addresses, and specific addresses. Each address is related to a specific layer in the TCPIIP architecture, as shown in figure.

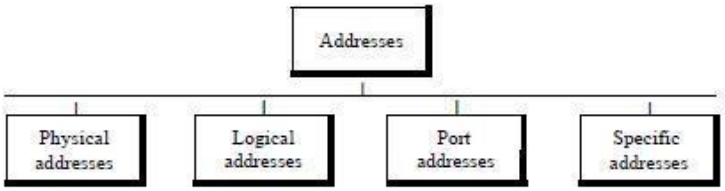


Figure 2.17 *Addresses in TCPIIP*

Physical Address/ link address/ MAC address

It is the address of a node as defined by its LAN or WAN. It is included in the frame used by the data link layer. It is the lowest-level address. The size and format of these addresses vary depending on the network. For example, Ethernet uses a 6-byte (48-bit) physical address that is imprinted on the network interface card (NIC). Most local-area networks use a 48-bit (6-byte) physical address written as 12 hexadecimal digits; every byte (2 hexadecimal digits) is separated by a colon, as shown below:

07:01:02:01:2C:4B
A 6-byte (12 hexadecimal digits) physical address

The relationship of layers and addresses in TCP/IP.

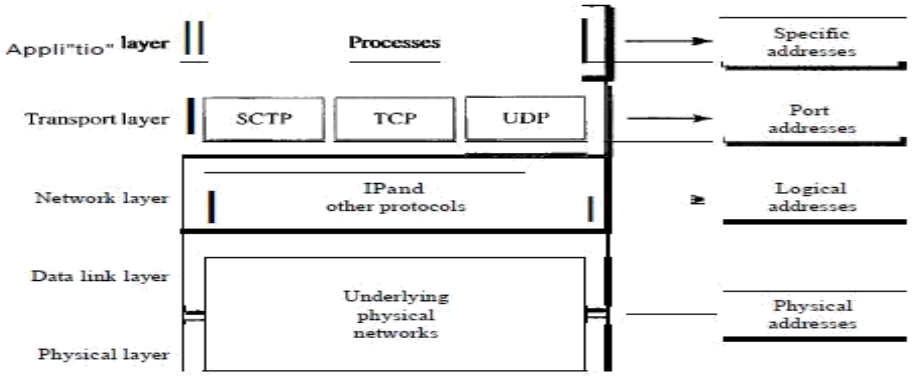


Figure 2.18 *Relationship of layers and addresses in TCPIIP*

Logical Addresses

Logical addresses are necessary for universal communications that are independent of underlying physical networks. Physical addresses are not adequate in an internetwork environment where different networks can have different address formats. A universal addressing system is needed in which each host can be identified uniquely, regardless of the underlying physical network. The logical addresses are designed for this purpose. A logical address in the Internet is currently a 32-bit address that can uniquely define a host connected to the Internet. No two publicly addressed and visible hosts on the Internet can have the same IP address.

Note that the physical addresses will change from hop to hop, but the logical addresses usually remain the same.

Port Addresses

The IP address and the physical address are necessary for a quantity of data to travel from a source to the destination host. However, arrival at the destination host is not the final objective of data communications on the Internet. A system that sends data from one computer to another is not complete. Today, computers are devices that can run multiple processes at the same time. The end objective of Internet communication is a process communicating with another process.

For example, computer A can communicate with computer C by using TELNET. At the same time, computer A communicates with computer B by using the File Transfer Protocol (FTP). For these processes to receive data simultaneously, we need a method to label the different processes. In other words, they need addresses. In the TCP/IP architecture, the label assigned to a process is called a port address. A port address in TCPIIP is 16 bits in length.

Note that although physical addresses change from hop to hop, logical and port addresses remain the same from the source to destination.

Specific Addresses

Some applications have user-friendly addresses that are designed for that specific address. Examples include the e-mail address (for example, forouzan@fhda.edu) and the Universal Resource Locator (URL) (for example, www.mhhe.com). The first defines the recipient of an e-mail; the second is used to find a document on the World Wide Web. These addresses, however, get changed to the corresponding port and logical addresses by the sending computer.

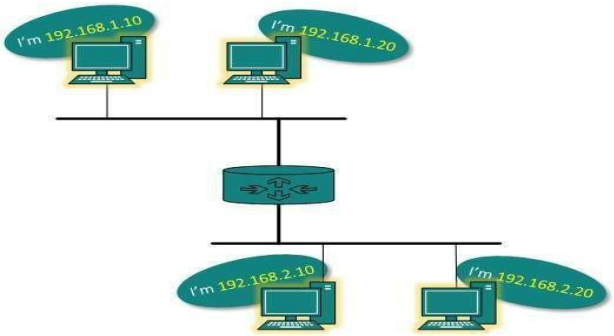
The physical addresses will change from hop to hop, but the logical addresses usually remain the same. The physical addresses change from hop to hop, but the logical and port addresses usually remain the same.

Network addressing

A network address is an identifier for a node or network interface of a telecommunications network. Network addresses are often designed to be unique across the network. A network address always points to host / node / server or it can represent a whole network. Network address is always configured on network interface card and is generally mapped by system with the MAC address of the machine.

More than one type of network address may be used in any one network. In some cases, node may have more than one network address, for example, each link interface may be uniquely identified.

Examples of network addresses are: a telephone number in the public switched telephone network, an Internet Protocol address in the Internet, a MAC address in an Ethernet network segment.

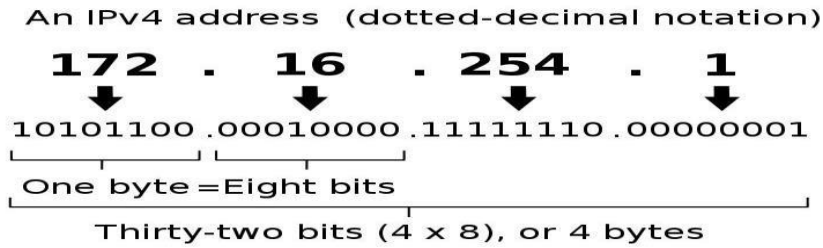


IP addressing enables every host on the TCP/IP network to be uniquely identifiable. IP addresses are divided into many categories:

1. Class A - it uses first octet for network addresses and last three octets for host addressing
2. Class B - it uses first two octets for network addresses and last two for host addressing
3. Class C - it uses first three octets for network addresses and last one for host addressing
4. Class D - it provides flat IP addressing scheme in contrast to hierarchical structure for above three.
5. Class E - It is used as experimental.

The version of the Internet Protocol (IP) that is commonly used is version 4 (IPv4). IPv4 provides hierarchical addressing scheme which enables it to divide the network into sub-networks, each with well-Defined number of hosts. In IPv4 an address consists of 32 bits which limits the address space to
4294967296 (2^32) possible unique addresses. IPv4 reserves some addresses for special purposes such as private networks or multicast addresses.

IPv4 addresses are canonically represented in dot-decimal notation, which consists of four decimal numbers, each ranging from 0 to 255, separated by dots, e.g., 172.16.254.1. Each part represents a group of 8 bits (octet) of the address. In some cases of technical writing, IPv4 addresses may be presented in various hexadecimal, octal, or binary representations.



Public vs. private addressing

An IP address is a unique numerical value that is used to identify a computer on a network. There are two kinds of IP addresses, public (also called globally unique IP addresses) and private.

Public IP addresses are assigned by the Internet Assigned Numbers Authority (IANA). The addresses are guaranteed to be globally unique and reachable on the Internet. This assures that multiple computers do not have the same IP address.

An Internet service provider (ISP) obtains a range of public IP addresses from IANA, and then the ISP assigns the addresses to customers to use when they connect to the Internet through the ISP.

Computer with a public IP address is visible to other computers on the Internet.

Private IP addresses cannot be used on the Internet. IANA has set aside three blocks of IP addresses that cannot be used on the global Internet. These three blocks of addresses are private IP addresses, and they are used for networks that do not directly connect to the Internet.

A private IP address is within one of the following blocks or range of addresses:

192.168.0.0/16: This block allows valid IP addresses within the range 192.168.0.1 to 192.168.255.254.

172.16.0.0/12: This block allows valid IP addresses within the range 172.16.0.1 to 172.31.255.254.

10.0.0.0/8: This block allows valid IP addresses within the range 10.0.0.1 to 10.255.255.254.

Most small businesses prefer to use private IP addresses for the local network, because ISPs generally charge a fee for each public IP address that the small business uses. As a result, using public IP addresses on a local network is costly. Rather than purchasing a globally unique IP address for each client computer that is on your local network, you can purchase one globally unique IP address and use it for the router interface that connects to your ISP.

IPv4 vs. IPv6 addresses

IPv4

The version of the Internet Protocol (IP) that is commonly used is version 4 (IPv4), which has not changed
substantially since RFC 791 was published in 1981. IPv4 is robust, easily implemented, interoperable, and capable of scaling to a global utility that can function with the Internet.

But, the Internet continues to grow exponentially, and the adoption of broadband technologies, such as cable modems, mobile information appliances, such as personal data assistants or PDAs, and cellular phones, means that many more addresses are needed.

IPv6

IPv6 significantly increases the number of addresses that are available. The most obvious difference between IPv6 and IPv4 is the size of the addresses. An IPv4 address is 32 bits long, and an IPv6 address is 128 bits long, which is four times longer than an IPv4 address.

Switching

A network is a set of connected devices. Whenever we have multiple devices, we have the problem of how to connect them to make one- to-one communication possible. A better solution is switching. A switched network consists of a series of interlinked nodes, called switches. Switches are devices capable of creating temporary connections between two or more devices linked to the switch. In a switched network, some of these nodes are connected to the end systems (computers or telephones, for example). Others are used only for routing. The below figure shows a switched network.

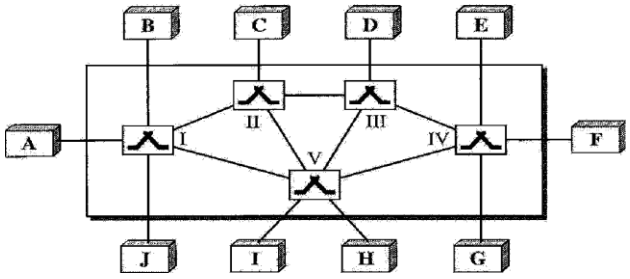


Figure 8.1 *Switched network*

The end systems (communicating devices) are labeled A, B, C, D, and so on, and the switches are labeled I, II, III, IV, and V. Each switch is connected to multiple links.

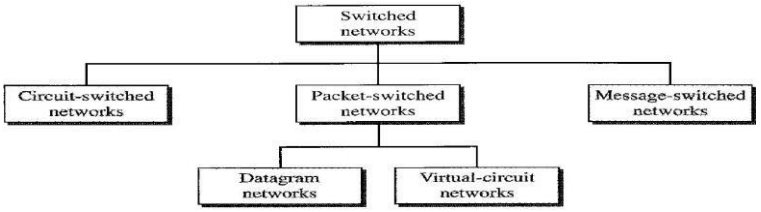


Figure 8.2 *Taxonomy of switched networks*

Three methods of switching are: circuit switching, packet switching, and message switching. Packet-switched networks can further be divided into two subcategories: Virtual-circuit networks and Datagram networks as shown in above figure.

1.10 Circuit-Switched Networks

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

In circuit switching, the resources need to be reserved during the setup phase; the resources remain dedicated for the entire duration of data transfer until the teardown phase. **A circuit-switched network is made of a set of switches connected by physical links, in which each link is divided into *n* channels.**

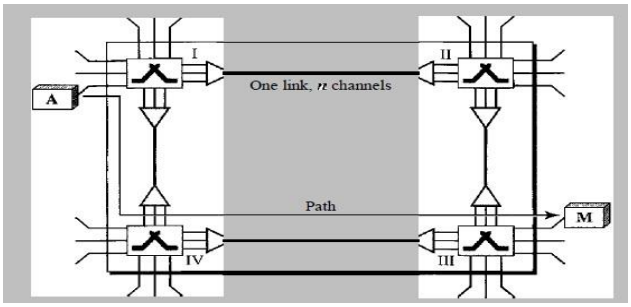
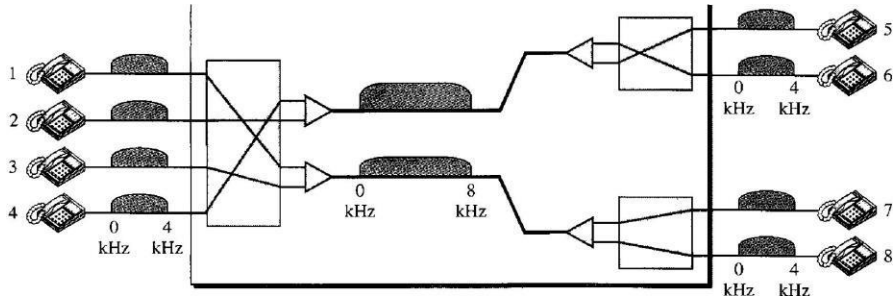


Figure 8.3 *A trivial circuit-switched network*



As a trivial example, let us use a circuit-switched network to connect eight telephones in a small area. Communication is through 4-kHz voice channels. We assume that each link uses FDM to connect a maximum of two voice channels. The bandwidth of each link is then 8 kHz. The above figure shows the situation. Telephone 1 is connected to telephone 7; 2 to *5;* 3 to 8; and 4 to 6. Of course the situation may change when new connections are made. The switch controls the connections.

Three Phases

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

Setup Phase:

Before the two parties (or multiple parties in a conference call) can communicate, a dedicated

circuit (combination of channels in links) needs to be established. The end systems are normally connected through dedicated lines to the switches, so connection setup means creating dedicated channels between

the switches. For example, in above figure, when system A needs to connect to system J, it sends a setup request that includes the address of system J, to switch I. Switch I finds a channel between system A and J that can be dedicated for this purpose. Switch I informs system J of system A's intention at this time.

In the next step to making a connection, an acknowledgment from system J needs to be sent in the opposite direction to system A. Only after system A receives this acknowledgment is the connection established. Note that end-to-end addressing is required for creating a connection between the two end systems.

Data Transfer Phase:

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

Teardown Phase:

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

We need to emphasize several points here:

Circuit switching takes place at the physical layer.

Before starting communication, the stations must make a reservation for the resources to be used during the communication. These resources, such as channels, switch buffers, switch processing time, and switch input/output ports, must remain dedicated during the entire duration of data transfer until the teardown phase.

Data transferred between the two stations are not packetized. The data are a continuous flow sent by the source station and received by the destination station.

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

Circuit-switched networks are not as efficient as the other two types of networks because resources are allocated during the entire duration of the connection. These resources are unavailable to other connections. In a telephone network, people normally terminate the communication when they have finished their conversation.

The delay in this type of network is minimal. The total delay is due to the time needed to create the connection, transfer data, and disconnect the circuit.

In circuit switching, the resources need to be reserved during the setup phase; the resources remain dedicated for the entire duration of data transfer until the teardown phase.

1.11 Packet-switched network

In data communications, we need to send messages from one end system to another. If the message is going to pass through a packet-switched network, it needs to be divided into packets of fixed or variable size. The size of the packet is determined by the network and the governing protocol.

In packet switching, there is no resource allocation for a packet. This means that there is no reserved bandwidth on the links, and there is no scheduled processing time for each packet. Resources are allocated on demand. The allocation is done on a first-come, first-served basis. When a switch receives a packet, no matter what is the source or destination, the packet must wait if there are other packets being processed. As with other systems in our daily life, this lack of reservation may create delay. For example, if we do not have a reservation at a restaurant, we might have to wait.

In a packet-switched network, there is no resource reservation; resources are allocated on demand.

A. Datagram Networks

In a datagram network, each packet is treated independently of all others. Packets in this approach are referred to as **datagram**. Datagram switching is normally done at the network layer. Below figure

shows how the datagram approach is used to deliver four packets from station A to station X. The switches in a datagram network are traditionally referred to as routers.

In this example, all four packets (or datagrams) belong to the same message, but may travel different paths to reach their destination. This approach can cause the datagrams of a transmission to arrive at their destination out of order with different delays between the packets. Packets may also be lost or dropped because of a lack of resources. In most protocols, it is the responsibility of an upper-layer protocol to reorder the datagrams or ask for lost datagrams before passing them on to the application.

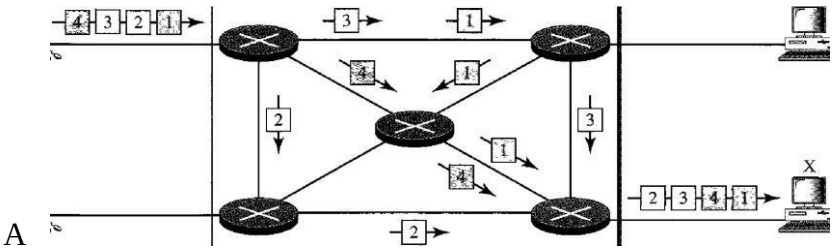


Figure 8.7 *A datagram network with four switches (routers)*

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

Routing Table

In this type of network, each switch (or packet switch) has a routing table which is based on the destination address. The routing tables are dynamic and are updated periodically. The destination addresses and the corresponding forwarding output ports are recorded in the tables. This is different from the table of a circuit switched network in which each entry is created when the setup phase is completed and deleted when the teardown phase is over.

Destination Address

Every packet in a datagram network carries a header that contains, among other information, the destination address of the packet. When the switch receives the packet, this destination address is examined; the routing table is consulted to find the corresponding port through which the packet should be forwarded. The destination address in the header of a packet in a datagram network remains the same during the entire journey of the packet.

The efficiency of a datagram network is better than that of a circuit-switched network; resources are allocated only when there are packets to be transferred. There may be greater delay in a datagram network than in a virtual-circuit network.

B. Virtual-Circuit Networks

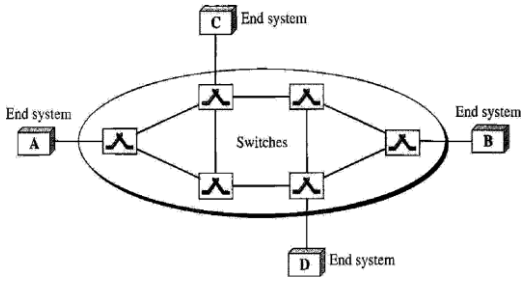


Figure 8.10 *Virtual-circuit network*

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

1. As in a circuit-switched network, there are setup and teardown phases in addition to the data transfer phase.
2. Resources can be allocated during the setup phase, as in a circuit-switched network, or on demand, as in a datagram network.
3. As in a datagram network, data are packetized and each packet carries an address in the header.
   However, the address in the header defines what should be the next switch and the channel on which the packet is being carried.
4. As in a circuit-switched network, all packets follow the same path established during the connection.
5. A virtual-circuit network is normally implemented in the data link layer, while a circuit-switched network is implemented in the physical layer and a datagram network in the network layer.

The above figure is an example of a virtual-circuit network. The network has switches that allow traffic from sources to destinations. A source or destination can be a computer, packet switch, bridge, or any other device that connects other networks.

Addressing

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

Global Addressing

A source or a destination address -an address that can be unique in the scope of the network. However, a global address in virtual-circuit networks is used only to create a virtual-circuit identifier.

Virtual-Circuit Identifier

The identifier that is actually used for data transfer is called the virtual-circuit identifier (VCI). A VCI, is a small number that has scope within the switch; it is used by a frame between two switches. When a frame arrives at a switch, it has a VCI; when it leaves, it has a different VCI. The above figure shows how the VCI in a data frame changes from one switch to another. Note that a VCI does not need to be a large number since each switch can use its own unique set of VCIs.

Three Phases

like a circuit-switched network, a **virtual-circuit network** is also need to go through three phases for connection: **setup, data transfer,** and **teardown**. In the setup phase, the source and destination use their global addresses to help switches make table entries for the connection. In the teardown phase, the source and destination inform the switches to delete the corresponding entry. Data transfer occurs between these two phases.

Data Transfer Phase

To transfer a frame from a source to its destination, all switches need to have a table entry for this virtual circuit. The table, in its simplest form, has four columns. This means that the switch holds four pieces of information for each virtual circuit that is already set up.

Setup Phase

In the setup phase, a switch creates an entry for a virtual circuit. For example, suppose source A needs to create a virtual circuit to B. Two steps are required: the **setup request** and the **acknowledgment.**

a. **Setup Request** A setup request frame is sent from the source to the destination.
b. **Acknowledgment** A special frame, called the acknowledgment frame, completes the entries in the switching tables.

Teardown Phase

In this phase, source A, after sending all frames to B, sends a special frame called a ***teardown request.*** Destination B responds with a teardown confirmation frame. All switches delete the corresponding entry from their tables.

Resource reservation in a virtual-circuit network can be made during the setup or can be on demand during the data transfer phase. In a virtual-circuit network, there is a one-time delay for setup and a one- time delay for teardown. If resources are allocated during the setup phase, there is no wait time for individual packets.

Virtual-circuit networks are used in switched WANs such as Frame Relay and ATM networks. The data link layer of these technologies is well suited to the virtual-circuit technology.