MODULE 1: INTRODUCTION

Data Communications

• Data Communication is defined as exchange of data between 2 devices over a transmission-medium.
• A communication-system is made up of
→ Hardware (physical equipment) and
→ Software (programs)
• For data-communication, the communicating-devices must be part of a communication-system.
• Four attributes of a communication-system:

1) Delivery
- The system must deliver data to the correct destination.

2) Accuracy
- The system must deliver the data accurately.
- Normally, the corrupted-data are unusable.

3) Timeliness
- The system must deliver audio/video data in a timely manner.
- This kind of delivery is called real-time transmission.
- Data delivered late are useless.

4) Jitter
- Jitter refers to the variation in the packet arrival-time.
- In other words, jitter is the uneven delay in the delivery of audio/video packets.

Components of Communication System

• Five components of a communication-system (Figure 1.1):
1) Message
2) Sender
3) Receiver
4) Transmission-Medium
5) Protocol

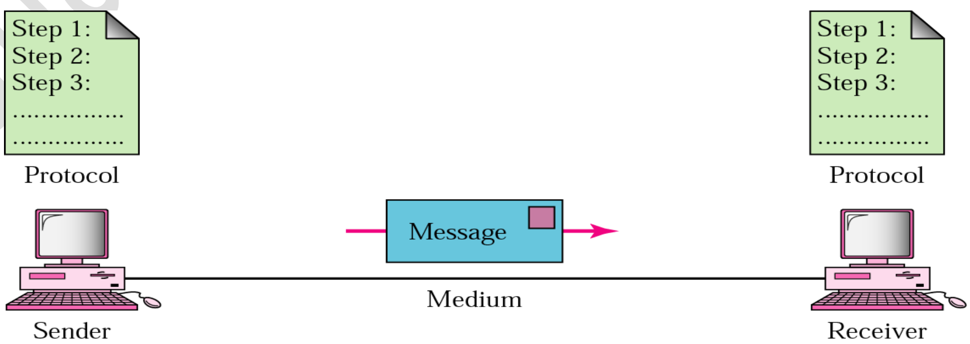


Fig 1.1

Canara Engineering College

1) Message

- Message is the information (or data) to be communicated.
- Message may consist of
  → Number/text
  → Picture or
  → Audio/Video

2) Sender

- Sender is the device that sends the data-message.
- Sender can be
  → Computer and
  → Mobile phone

3) Receiver

- Receiver is the device that receives the message.
- Receiver can be
  → Computer and
  → Mobile phone

4) Transmission Medium

- Transmission-medium is physical-path by which a message travels from sender to receiver.
- Transmission-medium can be wired or wireless.
- Examples of wired medium:
  → Twisted-pair wire (used in landline telephone)
  → Coaxial cable (used in cable TV network)
  → Fiber-optic cable

  Examples of wireless medium:
  → Radio waves
  → Microwaves
  → Infrared waves (ex: operating TV using remote control)

5) Protocol

- A protocol is a set of rules that govern data-communications.
- In other words, a protocol represents an agreement between the communicating-devices.
- Without a protocol, 2 devices may be connected but not communicating.

Data Representation

• Five different forms of information:

1) Text

- Text is represented as a bit-pattern. (Bit-pattern □□sequence of bits: 0s or 1s).
- Different sets of bit-patterns are used to represent symbols (or characters).
- Each set is called a code.
- The process of representing symbols is called encoding.
- Popular encoding system: ASCII, Unicode.

2) Number

- Number is also represented as a bit-pattern.

Canara Engineering College

- ASCII is not used to represent number. Instead, number is directly converted to binary-form.

3) Image

- Image is also represented as a bit-pattern.
- An image is divided into a matrix of pixels (picture-elements).
- A pixel is the smallest element of an image. (Pixel □□Small dot)
- The size of an image depends upon number of pixels (also called resolution).
- For example: An image can be divided into 1000 pixels or 10,000 pixels.
- Two types of images:

  i) Black & White Image

  ¤ If an image is black & white, each pixel can be represented by a value either 0 or 1.

  ¤ For example: Chessboard

  ii) Color Image

  ¤ There are many methods to represent color images.

  ¤ RGB is one of the methods to represent color images.

  ¤ RGB is called so called '.' each color is combination of 3 colors: red, green & blue.

4) Audio

- Audio is a representation of sound.
- By nature, audio is different from text, numbers, or images. Audio is continuous, not discrete.

5) Video

- Video is a representation of movie.
- Video can either

  → be produced as a continuous entity (e.g., by a TV camera), or

  → be a combination of images arranged to convey the idea of motion.

Direction of Data Flow

• Three ways of data-flow between 2 devices (Figure 1.2):

1) Simplex
2) Half-duplex
3) Full-duplex

Canara Engineering College

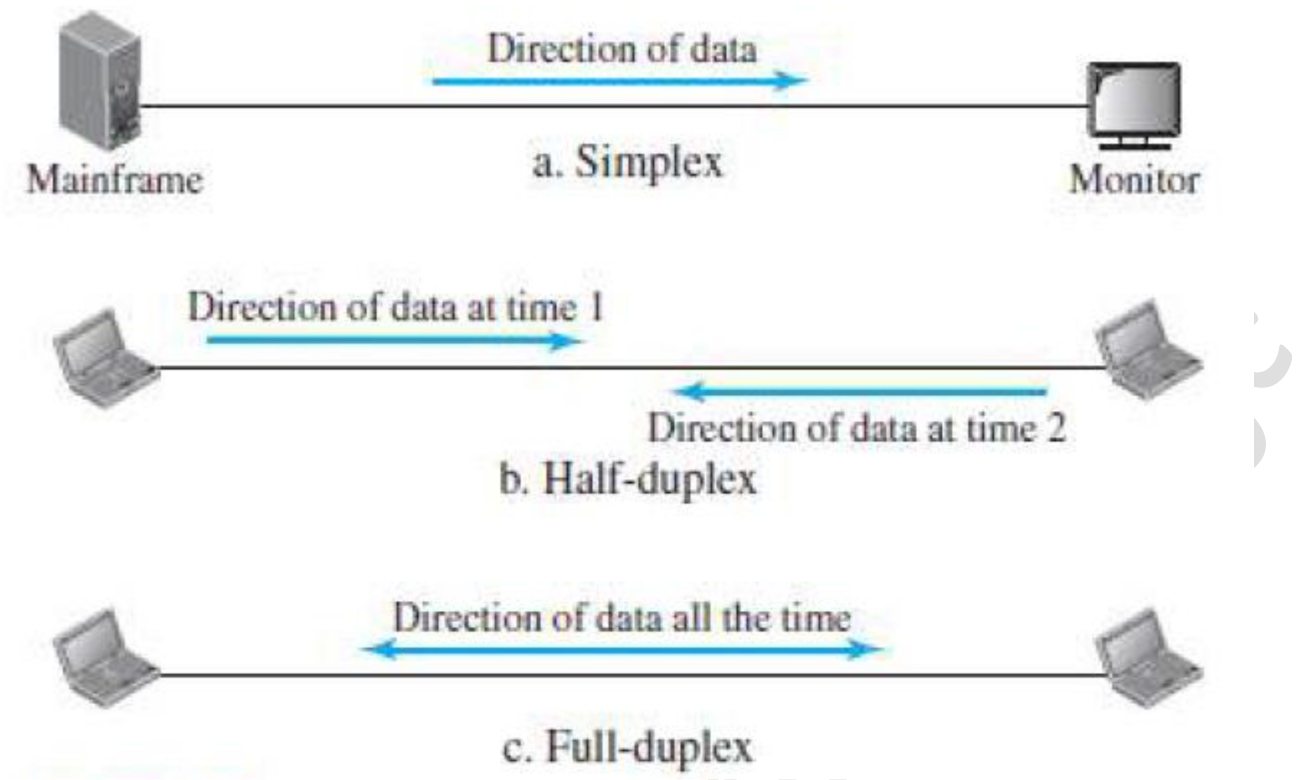


Fig 1.2

1) Simplex

- The communication is unidirectional
  (For ex: The simplex mode is like a one-way street).
- On a link, out of 2 devices:
  i) Only one device can transmit.
  ii) Another device can only receive.
  For example (Figure 1.2a):
- The monitor can only accept output.
- Entire-capacity of channel is used to send the data in one direction.

2) Half-Duplex

- Both the stations can transmit as well as receive but not at the same time.
  (For ex: The half-duplex mode is like a one-lane road with 2 directional traffic).
- When one station is sending, the other can only receive and vice-versa.
  For example (Figure 1.2b): Walkie-talkies
- Entire-capacity of a channel is used by one of the 2 stations that are transmitting the data.

3) Full-Duplex

- Both stations can transmit and receive at the same time.
  (For ex: The full-duplex is like a 2-way street with traffic flowing in both directions at the same time).
  For example (Figure 1.2c):
- Mobile phones (When 2 people are communicating by a telephone line, both can listen and talk at the same time)

Canara Engineering

- Entire-capacity of a channel is shared by both the stations that are transmitting the data.

Networks

• A network is defined as a set of devices interconnected by communication-links.
• This interconnection among computers facilitates information sharing among them.
• Computers may connect to each other by either wired or wireless media.
• Often, devices are referred to as nodes.
• A node can be any device capable of sending/receiving data in the network.
• For example: Computer & Printer
• The best-known computer network is the Internet.

Network Criteria

• A network must meet following 3 criteria's:

1) Performance

- Performance can be measured using i) Transit-time or ii) Response-time.
  i) Transit Time is defined as time taken to travel a message from one device to another.
  ii) Response Time is defined as the time elapsed between enquiry and response.
- The network-performance depends on following factors:
  i) Number of users
  ii) Type of transmission-medium
  iii) Efficiency of software
- Performance is evaluated by 2 networking-metrics: i) throughput and ii) delay.
- Good performance can be obtained by achieving higher throughput and smaller delay times

2) Reliability

- Reliability is measured by
  → Frequency of network-failure
  → Time taken to recover from a network-failure
  → Network's robustness in a disaster
- More the failures are, less is the network's reliability.

3) Security

- Security refers to the protection of data from the unauthorized access or damage.
- It also involves implementing policies for recovery from data-losses.

Physical Structures

Type of Connection

• Two types of connections (Figure 1.3):

1) Point-to-Point

- Only two devices are connected by a dedicated-link (Figure 1.3a).
- Entire-capacity of the link is reserved for transmission between those two devices.

Canara Engineering College

- For example: Point-to-Point connection b/w remote-control & TV for changing the channels.

2) Multipoint (Multi-drop)

- Three or more devices share a single link.
- The capacity of the channel is shared, either spatially or temporally (Figure 1.3b).
  - i) If link is used simultaneously by many devices, then it is spatially shared connection.
  - ii) If user takes turns while using the link, then it is time shared (temporal) connection.
  - (spatially→space or temporally→time)

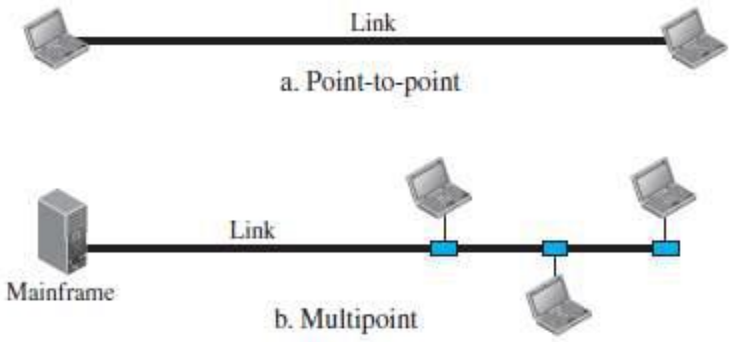


Figure 1.3 *Types of connections: point-to-point and multipoint*

Physical Topology

• The physical-topology defines how devices are connected to make a network.
• Four basic topologies are:

1) Mesh
2) Star
3) Bus and
4) Ring

Mesh Topology

• All the devices are connected to each other (Figure 1.7).
• There exists a dedicated point-to-point link between all devices.
• There are n(n-1) physical channels to link n devices.
• Every device not only sends its own data but also relays data from other nodes.
• For 'n' nodes,

→ There are n(n-1) physical-links
→ There are n(n-1)/2 duplex-mode links

• Every device must have (n–1) I/O ports to be connected to the other (n-1) devices.

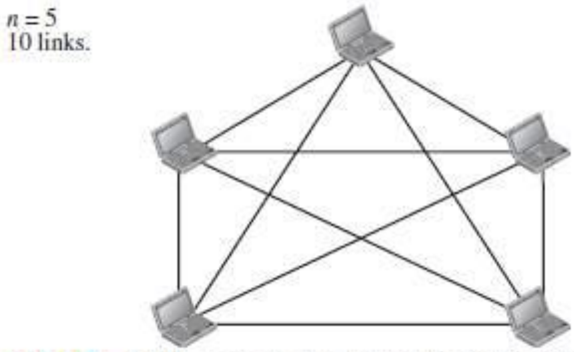


Figure 1.7 A fully connected mesh topology (five devices)

• Advantages:

1) Congestion reduced: Each connection can carry its own data load.
2) Robustness: If one link fails, it does not affect the entire system.
3) Security: When a data travels on a dedicated-line, only intended-receiver can see the data.
4) Easy fault identification & fault isolation: Traffic can be re-routed to avoid problematic links.

• Disadvantages:

1) Difficult installation and reconfiguration.
2) Bulk of wiring occupies more space than available space.
3) Very expensive: as there are many redundant connections.
4) Not mostly used in computer networks. It is commonly used in wireless networks.
5) High redundancy of the network-connections.

Bus Topology

• All the devices are connected to the single cable called bus (Figure 1.4).
• Every device communicates with the other device through this bus.
• A data from the source is broadcasted to all devices connected to the bus.
• Only the intended-receiver, whose physical-address matches, accepts the data.

Figure 1.4 A bus topology connecting three stations

• Devices are connected to the bus by drop-lines and taps.
• A drop-line is a connection running between the device and the bus.
• A tap is a connector that links to the bus or
• Advantages:

1) Easy installation.
2) Cable required is the least compared to mesh/star topologies.
3) Redundancy is eliminated.
4) Costs less (Compared to mesh/star topologies).
5) Mostly used in small networks. Good for LAN.

• Disadvantages:

1) Difficult to detect and troubleshoot fault.

2) Signal reflection at the taps can cause degradation in quality.
3) A fault/break in the cable stops all transmission.
4) There is a limit on
    i) Cable length
    ii) Number of nodes that can be connected.
5) Security is very low because all the devices receive the data sent from the source.

Star Topology

• All the devices are connected to a central controller called a hub (Figure 1.5).
• There exists a dedicated point-to-point link between a device & a hub.
• The devices are not directly linked to one another. Thus, there is no direct traffic between devices.
• The hub acts as a junction:
If device-1 wants to send data to device-2, the device-1 sends the data to the hub, then the hub relays the data to the device-2.

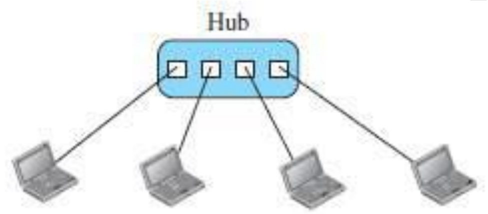


Figure 1.5 *A star topology connecting four stations*

• Advantages:
    1) Less expensive: Each device needs only one link & one I/O port to connect it to any devices.
    2) Easy installation & reconfiguration: Nodes can be added/removed w/o affecting the network.
    3) Robustness: If one link fails, it does not affect the entire system.
    4) Easy to detect and troubleshoot fault.
    5) Centralized management: The hub manages and controls the whole network.
• Disadvantages:
    1) Single point of failure: If the hub goes down, the whole network is dead.
    2) Cable length required is the more compared to bus/ring topologies.
    3) Number of nodes in network depends on capacity of hub.

Ring Topology

• Each device is connected to the next, forming a ring (Figure 1.6).
• There are only two neighbors for each device.
• Data travels around the network in one direction till the destination is reached.
• Sending and receiving of data takes place by the help of token.
• Each device has a repeater.
• A repeater
    → Receives a signal on transmission-medium &
    → Regenerates & passes the signal to next device.

Canara Engineering College

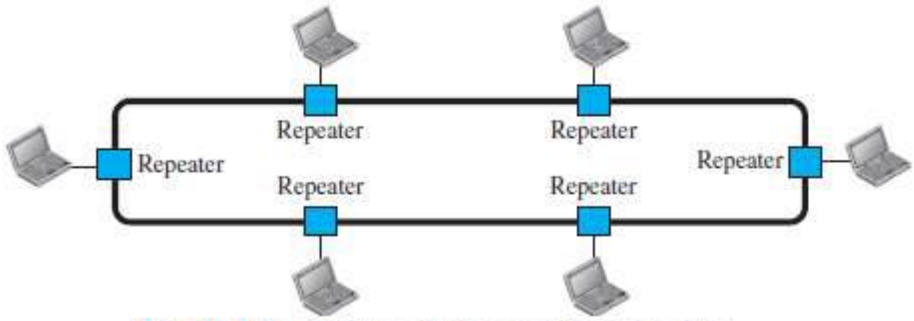


Figure 1.6 *A ring topology connecting six stations*

• Advantages:

1) Easy installation and reconfiguration.
2) To add/delete a device, requires changing only 2 connections.
3) 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 and its location.
4) Congestion reduced: Because all the traffic flows in only one direction.

• Disadvantages:

1) Unidirectional traffic.
2) A fault in the ring/device stops all transmission.

The above 2 drawbacks can be overcome by using dual ring.

3) There is a limit on
    i) Cable length &
    ii) Number of nodes that can be connected.
4) Slower: Each data must pass through all the devices between source and destination.

Network Types

• Two popular types of networks:

1) LAN (Local Area Network) &
2) WAN (Wide Area Network)

LAN

• LAN is used to connect computers in a single office, building or campus (Figure 1.8).
• LAN is usually privately owned network.
• A LAN can be simple or complex.
    1) Simple: LAN may contain 2 PCs and a printer.
    2) Complex: LAN can extend throughout a company.
• Each host in a LAN has an address that uniquely defines the host in the LAN.
• A packet sent by a host to another host carries both source host's and destination host's addresses.
• LANs use a smart connecting switch.
• The switch is able to
    → Recognize the destination address of the packet &
    → Guide the packet to its destination.
• The switch

Canara Engineering College

→ Reduces the traffic in the LAN &

→ Allows more than one pair to communicate with each other at the same time.

• Advantages:

1) Resource Sharing

Computer resources like printers and hard disks can be shared by all devices on the network.

2) Expansion

Nowadays, LANs are connected to WANs to create communication at a wider level.

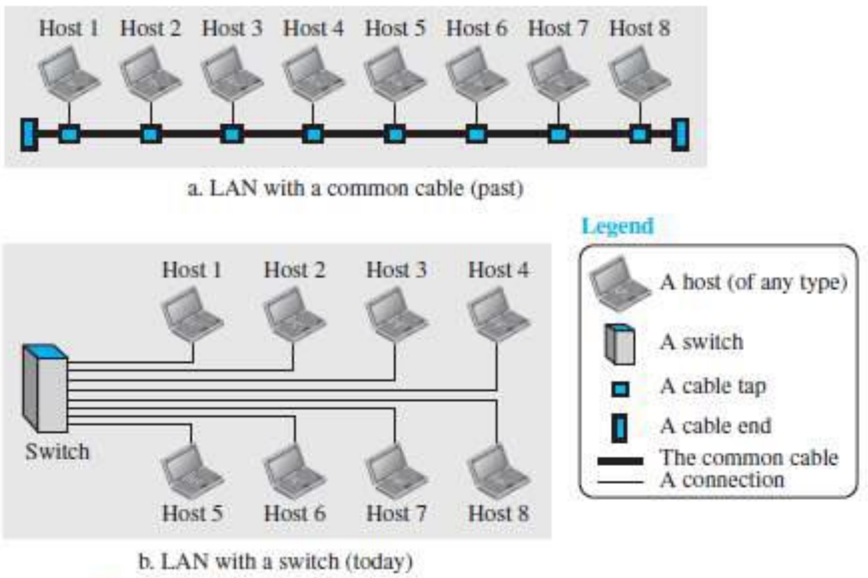


Figure 1.8 *An isolated LAN in the past and today*

WAN

• WAN is used to connect computers anywhere in the world.

• WAN can cover larger geographical area. It can cover cities, countries and even continents.

• WAN interconnects connecting devices such as switches, routers, or modems.

• Normally, WAN is

→ Created & run by communication companies (Ex: BSNL, Airtel)

→ Leased by an organization that uses it.

• A WAN can be of 2 types:

1) Point-to-point WAN

A point-to-point WAN is a network that connects 2 communicating devices through a transmission media (Figure 1.9).

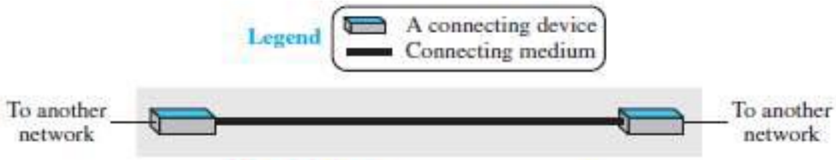


Figure 1.9 *A point-to-point WAN*

Switched WAN

- A switched WAN is a network with more than two ends.
- The switched WAN can be the backbones that connect the Internet.

- A switched WAN is a combination of several point-to-point WANs that are connected by switches (Figure 1.10).

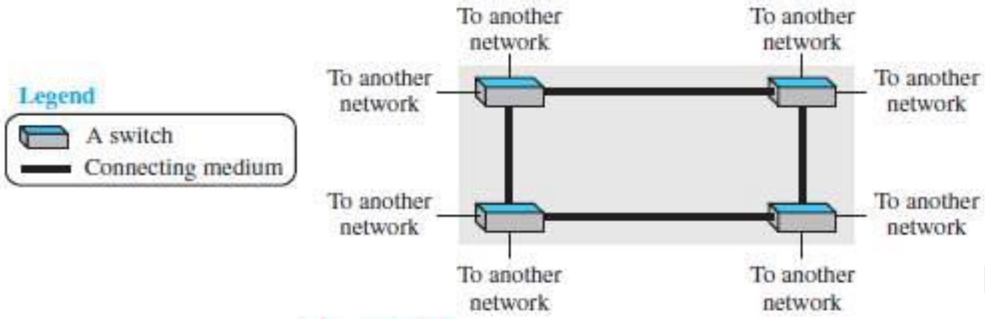


Figure 1.10 *A switched WAN*

Internetwork

• A network of networks is called an internet. (Internet inter-network) (Figure 1.12).
• For example (Figure 1.11):
Assume that an organization has two offices,

i) First office is on the east coast &
ii) Second office is on the west coast.

Each office has a LAN that allows all employees in the office to communicate with each other.
To allow communication between employees at different offices, the management leases a point-to-point dedicated WAN from a ISP and connects the two LANs.(ISP(Internet Service Provider) such as a telephone company ex: BSNL).

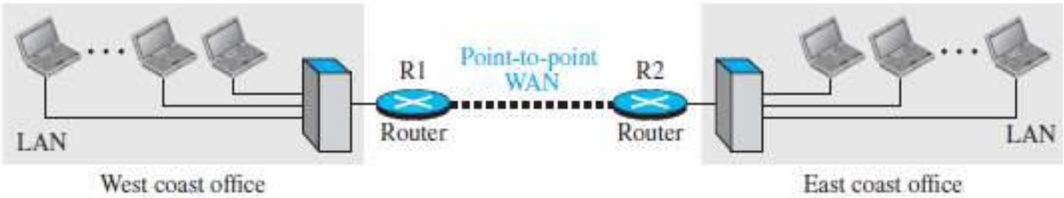


Figure 1.11 *An internetwork made of two LANs and one point-to-point WAN*

When a host in the west coast office sends a message to another host in the same office, the router blocks the message, but the switch directs the message to the destination.
On the other hand, when a host on the west coast sends a message to a host on the east coast, router R1 routes the packet to router R2, and the packet reaches the destination.

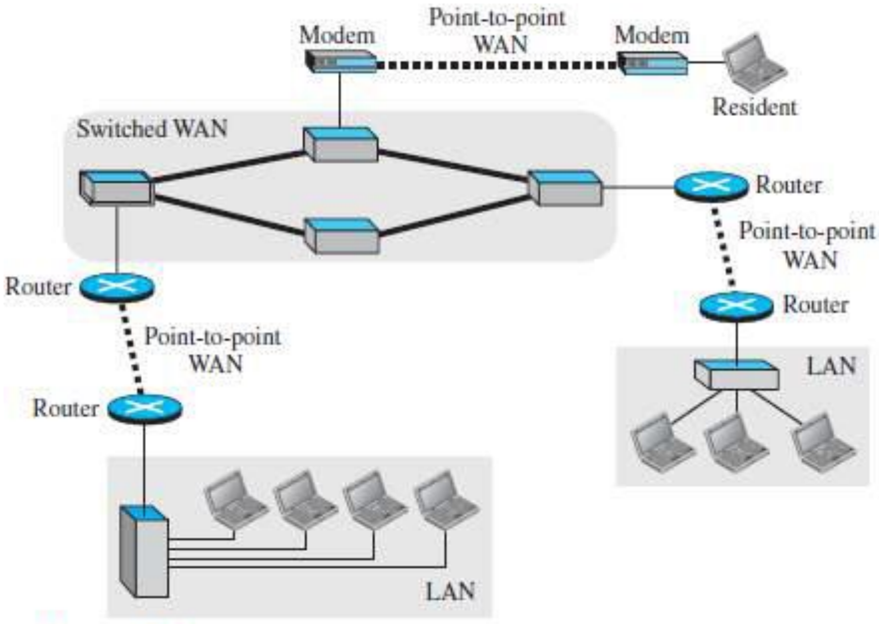


Figure 1.12 *A heterogeneous network made of four WANs and three LANs*

Difference between LAN and WAN

Parameters	**LAN**	**WAN**
Expands to	Local Area Network	Wide Area Network
Meaning	LAN is used to connect computers in a single office, building or campus	WAN is used to connect computers in a large geographical area such as countries
Ownership of network	Private	Private or public
Range	Small: up to 10 km	Large: Beyond 100 km
Speed	High: Typically 10, 100 and 1000 Mbps	Low: Typically 1.5 Mbps
Propagation Delay	Short	Long
Cost	Low	High
Congestion	Less	More
Design & maintenance	Easy	Difficult
Fault Tolerance	More Tolerant	Less Tolerant
Media used	Twisted pair	Optical fiber or radio waves
Used for	College, Hospital	Internet
Interconnects	LAN interconnects hosts	WAN interconnects connecting devices such as switches, routers, or modems

Switching

- An internet is a switched network in which a switch connects at least two links together.
- A switch needs to forward data from a network to another network when required.
- Two types of switched networks are 1)Circuit-switched and 2) Packet-switched networks.

Circuit-Switched Network

A dedicated connection, called a circuit, is always available between the two end systems. The switch can only make it active or inactive.

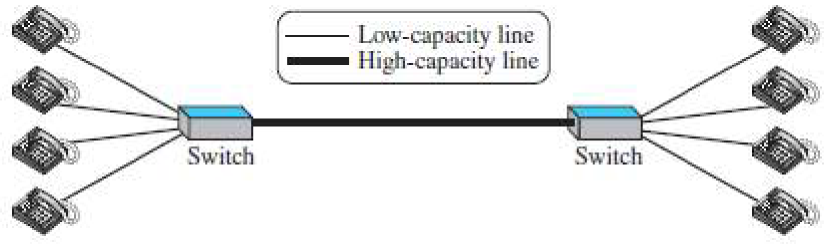


Figure 1.13 *A circuit-switched network*

¤ As shown in Figure 1.13, the 4 telephones at each side are connected to a switch.
¤ The switch connects a telephone at one side to a telephone at the other side.
¤ A high-capacity line can handle 4 voice communications at the same time.
¤ The capacity of high line can be shared between all pairs of telephones.
¤ The switch is used for only forwarding.

Advantage:

A circuit-switched network is efficient only when it is working at its full capacity.

Disadvantage:

Most of the time, the network is inefficient because it is working at partial capacity.

Packet-Switched Network

In a computer network, the communication between the 2 ends is done in blocks of data called packets. The switch is used for both storing and forwarding because a packet is an independent entity that can be stored and sent later.

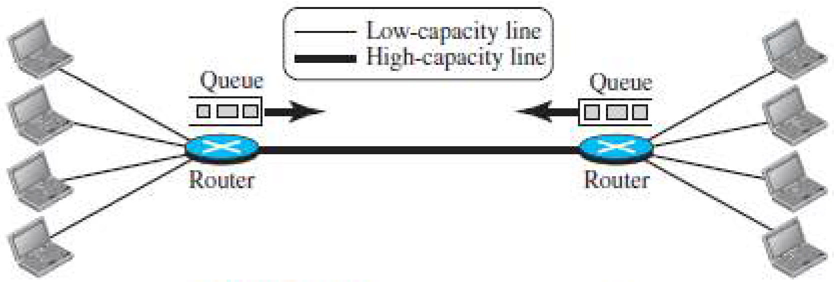


Figure 1.14 *A packet-switched network*

¤ As shown in Figure 1.14, the 4 computers at each side are connected to a router.
¤ A router has a queue that can store and forward the packet.
¤ The high-capacity line has twice the capacity of the low-capacity line.
¤ If only 2 computers (one at each site) need to communicate with each other, there is no waiting for the packets.
¤ However, if packets arrive at one router when high-capacity line is at its full capacity, the packets should be stored and forwarded.

Advantages:

A packet-switched network is more efficient than a circuit switched network.

Disadvantage:

The packets may encounter some delays.

The Internet Today

• A network of networks is called an internet. (Internet inter-network)

• Internet is made up of (Figure 1.15)

1) Backbones
2) Provider networks &
3) Customer networks

1) Backbones

- Backbones are large networks owned by communication companies such as BSNL and Airtel.
- The backbone networks are connected through switching systems, called peering points.

2) Provider Networks

- Provider networks use the services of the backbones for a fee.
- Provider networks are connected to backbones and sometimes to other provider networks.

3) Customer Networks

- Customer networks actually use the services provided by the Internet.
- Customer networks pay fees to provider networks for receiving services.
    - Backbones and provider networks are also called Internet Service Providers (ISPs).
    - The backbones are often referred to as international ISPs.
- The provider networks are often referred to as national or regional ISPs.

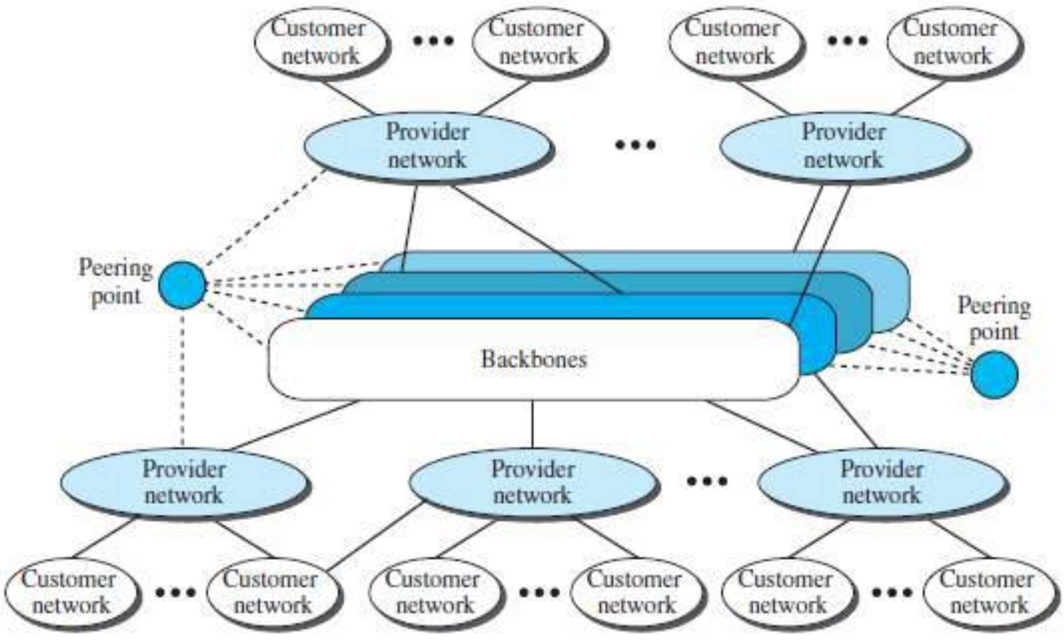


Figure 1.15 *The Internet today*

Accessing the Internet

• The Internet today is an internetwork that allows any user to become part of it.

• However, the user needs to be physically connected to an ISP.

• The physical connection is normally done through a point-to-point WAN.

1) Using Telephone Networks

- Most residences have telephone service, which means they are connected to a telephone network.
- Most telephone networks have already connected themselves to the Internet. Thus, residences can connect to the Internet using a point-to-point WAN.
- This can be done in two ways:

  A) Dial-up service

  ¤ A modem can be added to the telephone line.

  ¤ A modem converts data to voice.

  ¤ The software installed on the computer

  → Dials the ISP &

  → Imitates making a telephone connection.

  ¤ Disadvantages:

  i) The dial-up service is very slow.

  ii) When line is used for Internet connection, it cannot be used for voice connection.

  iii) It is only useful for small residences.

  B) DSL Service

  ¤ DSL service also allows the line to be used simultaneously for voice & data communication.

  ¤ Some telephone companies have upgraded their telephone lines to provide higher speed Internet services to residences.

2) Using Cable Networks

- A residence can be connected to the Internet by using cable service.
- Cable service provides a higher speed connection.
- The speed varies depending on the number of neighbors that use the same cable.

3) Using Wireless Networks

- A residence can use a combination of wireless and wired connections to access the Internet.
- A residence can be connected to the Internet through a wireless WAN.

4) Direct Connection to the Internet

- A large organization can itself become a local ISP and be connected to the Internet.
- The organization

  → Leases a high-speed WAN from a carrier provider and

  → Connects itself to a regional ISP.

Standards And Administration

Internet Standards

• An Internet standard is a thoroughly tested specification useful to those who work with the Internet.
• The Internet standard 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 with no official status and a 6-month lifetime.
• Upon recommendation from the Internet authorities, a draft may be published as a 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. (working document →a work in progress RFC(Request for Comment)

Canara Engineering College

Maturity Levels

• An RFC, during its lifetime, falls into one of 6 maturity levels (Figure 1.16):

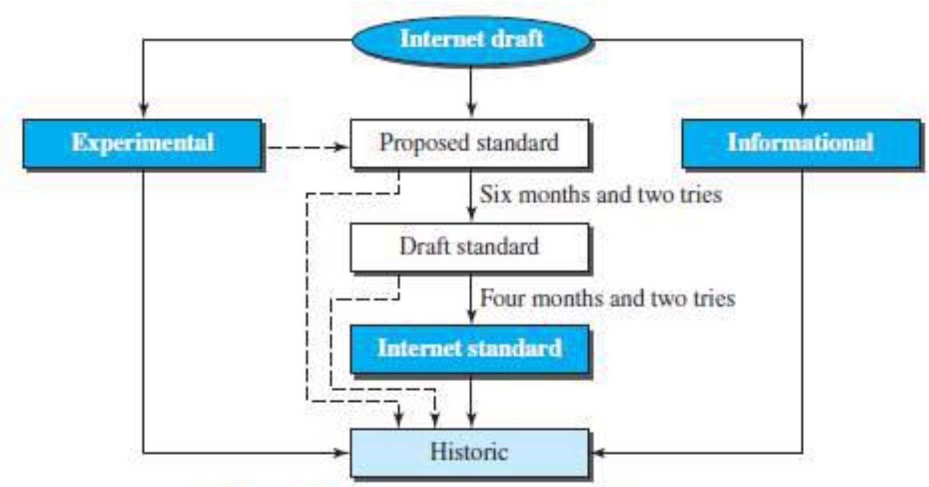


Figure 1.16 *Maturity levels of an RFC*

1) Proposed Standard

- Proposed standard is specification that is stable, well-understood & of interest to Internet community.
- Specification is usually tested and implemented by several different groups.

2) Draft Standard

- A proposed standard is elevated to draft standard status after at least 2 successful independent and interoperable implementations.

3) Internet Standard

- A draft standard reaches Internet standard status after demonstrations of successful implementation.

4) Historic

- The historic RFCs are significant from a historical perspective.
- They either

  → Have been superseded by later specifications or

  → Have never passed the necessary maturity levels to become an Internet standard.

5) Experimental

- An RFC classified as experimental describes work related to an experimental situation.
- Such an RFC should not be implemented in any functional Internet service.

6) Informational

- An RFC classified as informational contains general, historical, or tutorial information related to the Internet.
- Usually, it is written by a vendor.

  ISOC (Internet Society)

  IAB (Internet Architecture Board)

  IETF (Internet Engineering Task Force)

  IRTF (Internet Research Task Force)

  IESG (Internet Engineering Steering Group)

  IRSG (Internet Research Steering Group)

Requirement Levels

• RFCs are classified into 5 requirement levels:

1) Required

- An RFC labeled required must be implemented by all Internet systems to achieve minimum conformance.
- For example, IP and ICMP are required protocols.

2) Recommended

- An RFC labelled recommended is not required for minimum conformance.
- It is recommended because of its usefulness.
- For example, FTP and TELNET are recommended protocols.

3) Elective

- An RFC labelled elective is not required and not recommended.
- However, a system can use it for its own benefit.

4) Limited Use

- An RFC labelled limited use should be used only in limited situations.
- Most of the experimental RFCs fall under this category.

5) Not Recommended

- An RFC labelled not recommended is inappropriate for general use.
- Normally a historic RFC may fall under this category.

Internet Administration

1) ISOC

• ISOC is a nonprofit organization formed to provide support for Internet standards process (Fig 1.17).

• ISOC maintains and supports other Internet administrative bodies such as IAB, IETF, IRTF, and IANA.

2) IAB

• IAB is the technical advisor to the ISOC.

• Two main purposes of IAB:

i) To oversee the continuing development of the TCP/IP Protocol Suite

ii) To serve in a technical advisory capacity to research members of the Internet community.

• Another responsibility of the IAB is the editorial management of the RFCs.

• IAB is also the external liaison between the Internet and other standards organizations and forums.

• IAB has 2 primary components: i) IETF and ii) IRTF.

i) IETF

- IETF is a forum of working groups managed by the IESG.
- IETF is responsible for identifying operational problems & proposing solutions to the problems
- IETF also develops and reviews specifications intended as Internet standards.
- The working groups are collected into areas, and each area concentrates on a specific topic.
- Currently 9 areas have been defined. The areas include applications, protocols, routing, network management next generation (IPng), and security.

ii) IRTF

- IRTF is a forum of working groups managed by the IRSG.
- IRTF focuses on long-term research topics related to Internet protocols, applications, architecture, and technology.

Canara Engineering College

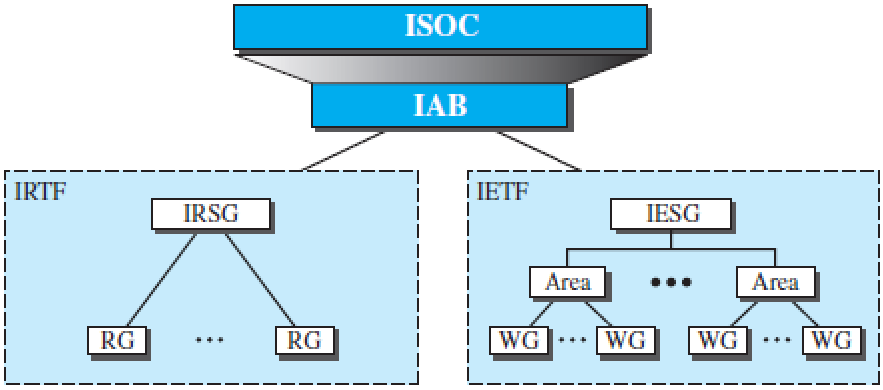


Figure 1.17 *Internet administration*

Canara Engineering College

NETWORK MODELS

Protocol Layering

• A protocol defines the rules that both the sender and receiver and all intermediate devices need to follow to be able to communicate effectively.

• When communication is simple, we may need only one simple protocol.

When communication is complex, we need to divide the task b/w different layers. We need a protocol at each layer, or protocol layering.

Scenarios

First Scenario

• In the first scenario, communication is so simple that it can occur in only one layer (Figure 2.1).

• Assume Maria and Ann are neighbours with a lot of common ideas.

• Communication between Maria and Ann takes place in one layer, face to face, in the same language

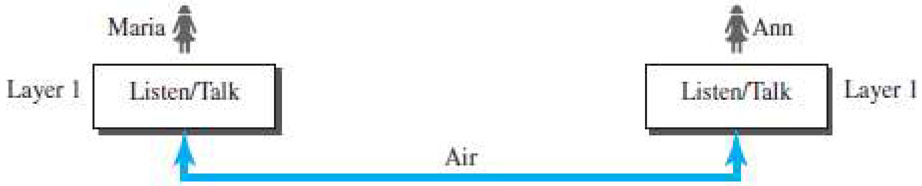


Figure 2.1 *A single-layer protocol*

Second Scenario

• Maria and Ann communicate using regular mail through the post office (Figure 2.2).

• However, they do not want their ideas to be revealed by other people if the letters are intercepted.

• They agree on an encryption/decryption technique.

• The sender of the letter encrypts it to make it unreadable by an intruder; the receiver of the letter decrypts it to get the original letter.

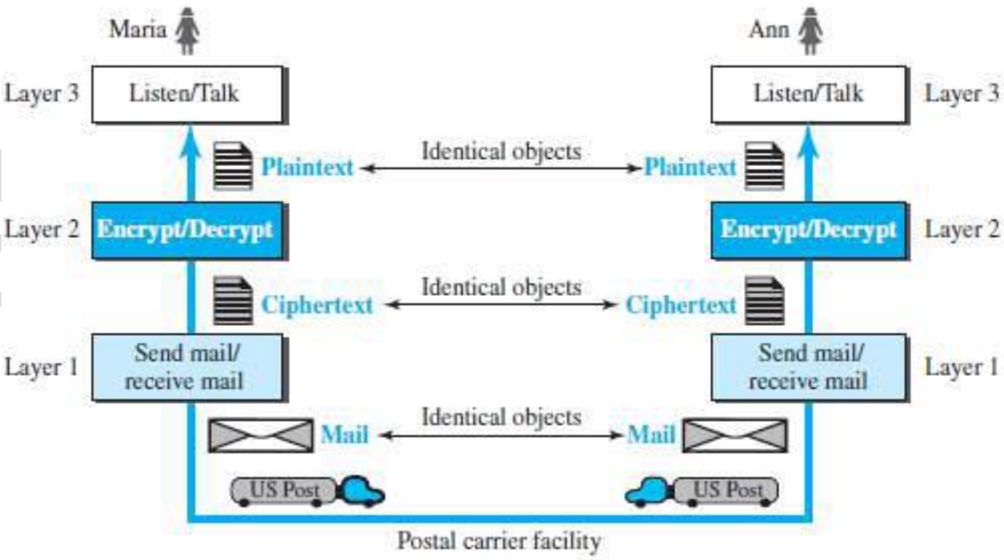


Figure 2.2 *A three-layer protocol*

Protocol Layering

• Protocol layering enables us to divide a complex task into several smaller and simpler tasks.

• Modularity means independent layers.

• A layer (module) can be defined as a black box with inputs and outputs, without concern about how inputs are changed to outputs.

• If two machines provide the same outputs when given the same inputs, they can replace each other.

• Advantages:

1) It allows us to separate the services from the implementation.

2) There are intermediate systems that need only some layers, but not all layers.

• Disadvantage:

1) Having a single layer makes the job easier. There is no need for each layer to provide a service to the upper layer and give service to the lower layer.

Principles of Protocol Layering

1) First Principle

• If we want bidirectional communication, we need to make each layer able to perform 2 opposite tasks, one in each direction.

• For example, the third layer task is to listen (in one direction) and talk (in the other direction).

2) Second Principle

• The two objects under each layer at both sites should be identical.

• For example, the object under layer 3 at both sites should be a plaintext letter.

Logical Connections

• We have layer-to-layer communication (Figure 2.3).

• There is a logical connection at each layer through which 2 end systems can send the object created from that layer.

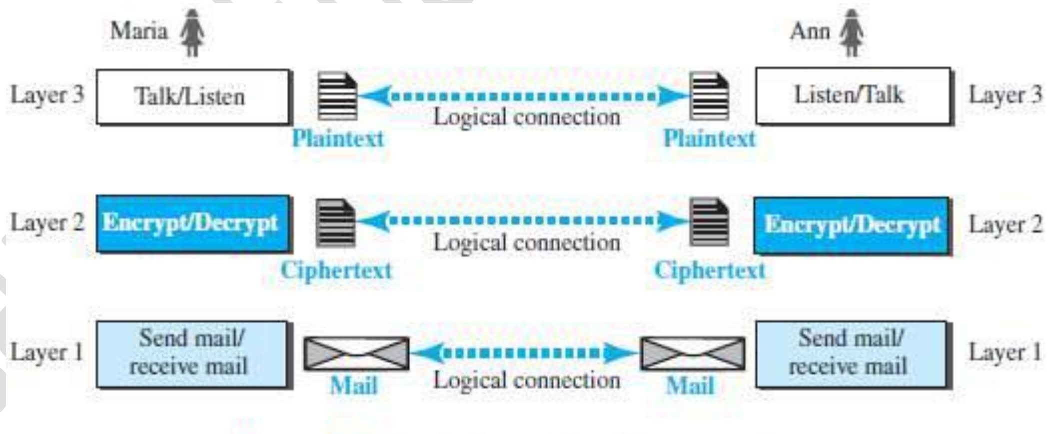


Figure 2.3 *Logical connection between peer layers*

TCP/IP Protocol Suite

• TCP/IP is a protocol-suite used in the Internet today.

• Protocol-suite refers a set of protocols organized in different layers.

• It 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 the services provided by one or more lower level protocols.

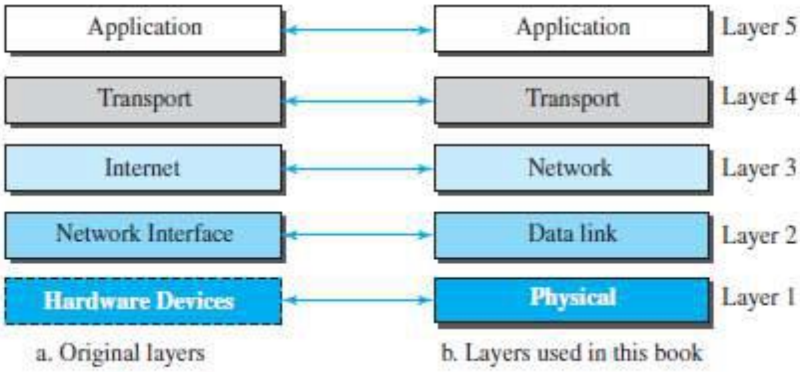


Figure 2.4 *Layers in the TCP/IP protocol suite*

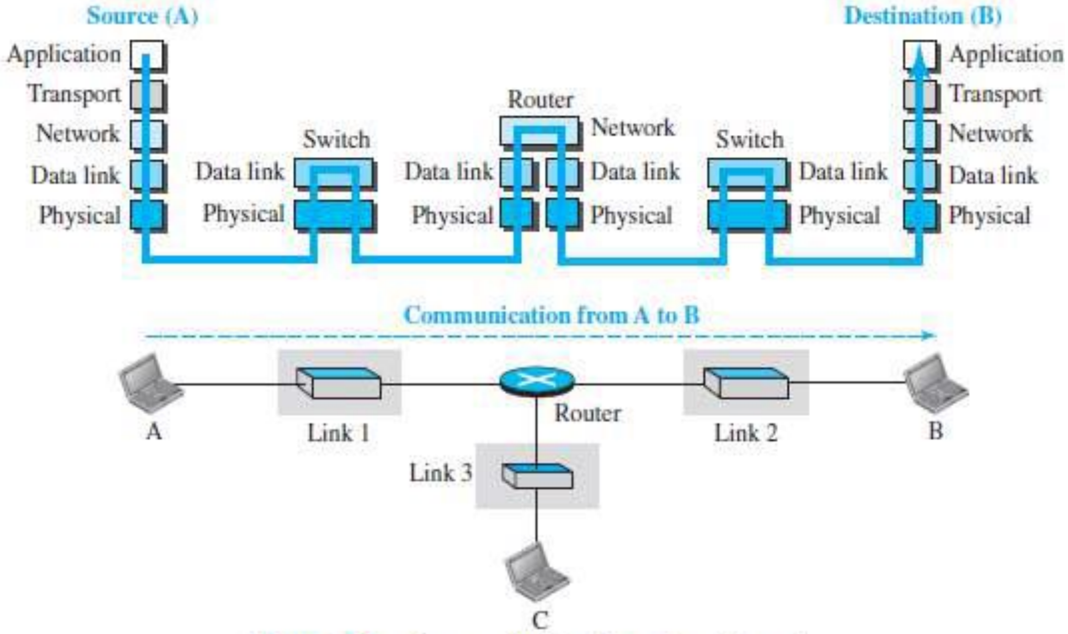


Figure 2.5 *Communication through an internet*

• Let us assume that computer A communicates with computer B (Figure 2.4).
• As the Figure 2.5 shows, we have five communicating devices:
 1) Source host(computer A) 2) Link-layer switch in link 1
 3) Router 4) Link-layer switch in link 2
 5) Destination host (computer B).
• Each device is involved with a set of layers depending on the role of the device in the internet.
• The two hosts are involved in all five layers.
• The source host
 → Creates a message in the application layer and
 → Sends the message down the layers so that it is physically sent to the destination host.
• The destination host
 → Receives the message at the physical layer and
 → Then deliver the message through the other layers to the application layer.

• The router is involved in only three layers; there is no transport or application layer.
• A router is involved in n combinations of link and physical layers. where n = number of links the router is connected to.
• The reason is that each link may use its own data-link or physical protocol.
• A link-layer switch is involved only in two layers: i) data-link and ii) physical.

Layers in the TCP/IP Protocol Suite

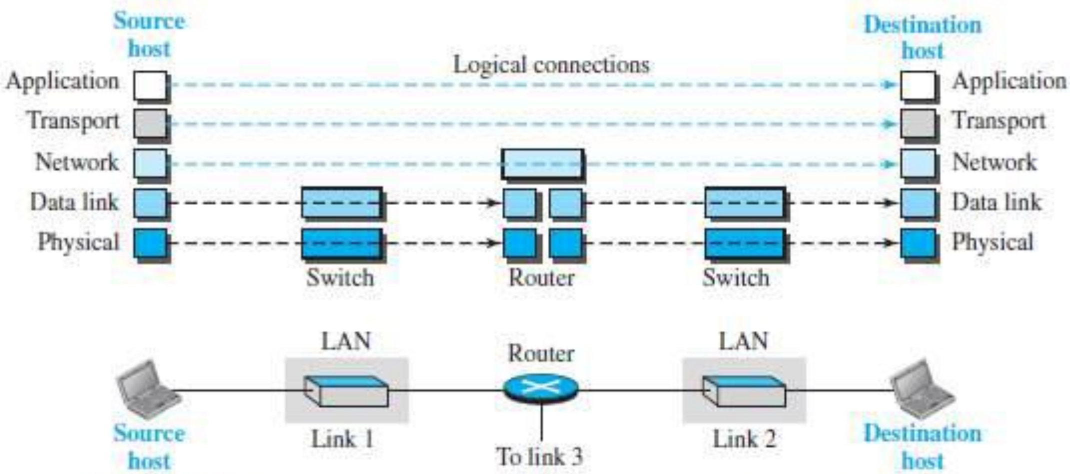


Figure 2.6 *Logical connections between layers of the TCP/IP protocol suite*

• As shown in the figure 2.6, the duty of the application, transport, and network layers is end-to-end.
• However, the duty of the data-link and physical layers is hop-to-hop. A hop is a host or router.
• The domain of duty of the top three layers is the internet.
The domain of duty of the two lower layers is the link.
• In top 3 layers, the data unit should not be changed by any router or link-layer switch.
In bottom 2 layers, the data unit is changed only by the routers, not by the link-layer switches.

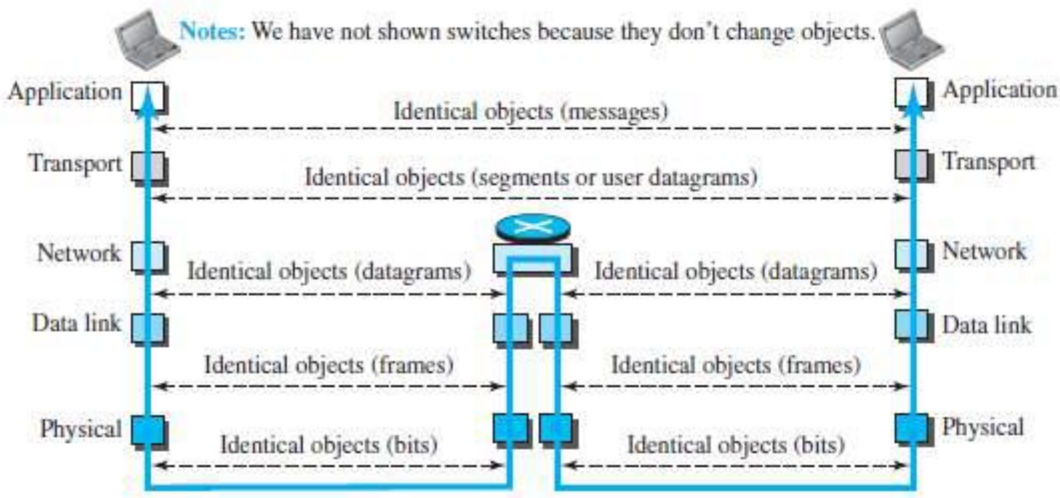


Figure 2.7 *Identical objects in the TCP/IP protocol suite*

• Identical objects exist between two hops. Because router may fragment the packet at the network layer and send more packets than received (Figure 2.7).
• The link between two hops does not change the object.

Description of Each Layer
Physical Layer
• The physical layer is responsible for movements of individual bits from one node to another node.
• Transmission media is another hidden layer under the physical layer.
• Two devices are connected by a transmission medium (cable or air).
• The transmission medium does not carry bits; it carries electrical or optical signals.
• The physical layer
→ Receives bits from the data-link layer &
→ Rends through the transmission media.

Data Link Layer
• Data-link-layer (DLL) is responsible for moving frames from one node to another node over a link.
• The link can be wired LAN/WAN or wireless LAN/WAN.
• The data-link layer
→ Gets the datagram from network layer
→ Encapsulates the datagram in a packet called a frame.
→ Sends the frame to physical layer.
• TCP/IP model does not define any specific protocol.
• DLL supports all the standard and proprietary protocols.
• Each protocol may provide a different service.
• Some protocols provide complete error detection and correction; some protocols provide only error correction.

Network Layer
• The network layer is responsible for source-to-destination transmission of data.
• The network layer is also responsible for routing the packet.
• The routers choose the best route for each packet.
• Why we need the separate network layer?
1) The separation of different tasks between different layers.
2) The routers do not need the application and transport layers.
• TCP/IP model defines 5 protocols:
1) IP (Internetworking Protocol) 2) ARP (Address Resolution Protocol)
3) ICMP (Internet Control Message Protocol) 4) IGMP (Internet Group Message Protocol)

1) IP
- IP is the main protocol of the network layer.
- IP defines the format and the structure of addresses.
- IP is also responsible for routing a packet from its source to its destination.
- It is a connection-less & unreliable protocol.
  i) Connection-less means there is no connection setup b/w the sender and the receiver.
  ii) Unreliable protocol means

Canara Engineering College

→ IP does not make any guarantee about delivery of the data.
→ Packets may get dropped during transmission.

- It provides a best-effort delivery service.
- Best effort means IP does its best to get the packet to its destination, but with no guarantees.
  IP does not provide following services
  → flow control
  → error control
  → congestion control services.
- If an application requires above services, the application should rely only on the transport layer protocol.

2) ARP

- ARP is used to find the physical-address of the node when its Internet-address is known.
- Physical address is the 48-bit address that is imprinted on the NIC or LAN card.
- Internet address (IP address) is used to uniquely & universally identify a device in the internet.

3) ICMP

- ICMP is used to inform the sender about datagram-problems that occur during transit.

4) IGMP

- IGMP is used to send the same message to a group of recipients.

Transport Layer

• TL protocols are responsible for delivery of a message from a process to another process.
• The transport layer
  → gets the message from the application layer
  → encapsulates the message in a packet called a segment and
  → sends the segment to network layer.
• TCP/IP model defines 3 protocols:
  1) TCP (Transmission Control Protocol)
  2) UDP (User Datagram Protocol) &
  3) SCTP (Stream Control Transmission Protocol)

1) TCP

- TCP is a reliable connection-oriented protocol.
- A connection is established b/w the sender and receiver before the data can be transmitted.
- TCP provides
  → flow control
  → error control and
  → congestion control

2) UDP

- UDP is the simplest of the 3 transport protocols.
- It is an unreliable, connectionless protocol.
- It does not provide flow, error, or congestion control.
- Each datagram is transported separately & independently.

Canara Engineering College

- It is suitable for application program that
  → needs to send short messages &
  → cannot afford the retransmission.

3) SCTP

- SCTP provides support for newer applications such as voice over the Internet.
- It combines the best features of UDP and TCP.

Application Layer

• The two application layers exchange messages between each other.
• Communication at the application layer is between two processes (two programs running at this layer).
• To communicate, a process sends a request to the other process and receives a response.
• Process-to-process communication is the duty of the application layer.
• TCP/IP model defines following protocols:

1) SMTP (Simple Mail Transfer Protocol) is used to transport email between a source and destination.
2) TELNET (Terminal Network) is used for accessing a site remotely.
3) FTP(File Transfer Protocol) is used for transferring files from one host to another.
4) DNS (Domain Name System) is used to find the IP address of a computer.
5) SNMP(Simple Network Management Protocol) is used to manage the Internet at global and local levels.
6) HTTP(Hyper Text Transfer Protocol) is used for accessing the World Wide Web (WWW).

Encapsulation and Decapsulation

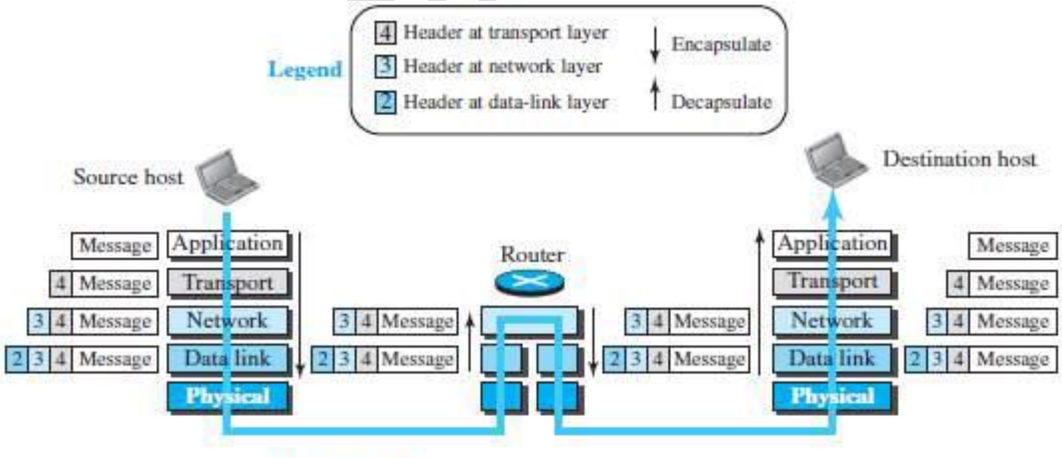


Figure 2.8 *Encapsulation/Decapsulation*

Encapsulation at the Source Host

• At the source, we have only encapsulation (Figure 2.8).

1) At the application layer, the data to be exchanged is referred to as a message.

- A message normally does not contain any header or trailer.
- The message is passed to the transport layer.

2) The transport layer takes the message as the payload.

- Transport Layer adds its own header to the payload.

- The header contains
  → identifiers of the source and destination application programs
  → information needed for flow, error control, or congestion control.
- The transport-layer packet is called the segment (in TCP) and the user datagram (in UDP).
- The segment is passed to the network layer.

3) The network layer takes the transport-layer packet as payload.

- Network Layer adds its own header to the payload.
- The header contains
  → addresses of the source and destination hosts
  → some information used for error checking of the header &
  → fragmentation information.
- The network-layer packet is called a datagram.
- The datagram is passed to the data-link layer.

4) The data-link layer takes the network-layer packet as payload.

- Data-link layer adds its own header to the payload.
- The header contains the physical addresses of the host or the next hop (the router).
- The link-layer packet is called a frame.
- The frame is passed to the physical layer for transmission

B) Decapsulation and Encapsulation at the Router

• At the router, we have both encapsulation & encapsulation and because the router is connected to two or more links.

1) Data-link layer
→ receives frame from physical layer
→ decapsulates the datagram from the frame and
→ passes the datagram to the network layer.

2) The network layer
→ inspects the source and destination addresses in the datagram header and
→ consults forwarding table to find next hop to which the datagram is to be delivered.

The datagram is then passed to the data-link layer of the next link.

3) The data-link layer of the next link
→ encapsulates the datagram in a frame and
→ passes the frame to the physical layer for transmission.

C) Decapsulation at the Destination Host

• At the destination host, each layer
→ decapsulates the packet received from lower layer
→ removes the payload and
→ delivers the payload to the next-higher layer

Addressing

• We have logical communication between pairs of layers.
• Any communication that involves 2 parties needs 2 addresses: source address and destination address.

Canara Engineering College

• We need 4 pairs of addresses (Figure 2.9):

1) At the application layer, we normally use names to define

→ site that provides services, such as vtunotesbysri.com, or

→ e-mail address, such as abc@gmail.com.

2) At the transport layer, addresses are called port numbers.

- Port numbers define the application-layer programs at the source and destination.
- Port numbers are local addresses that distinguish between several programs running at the same time.

3) At the network-layer, addresses are called logical addresses (IP Address).

- IP address uniquely defines the connection of a device to the Internet.
- The IP addresses are global, with the whole Internet as the scope.

4) At the data link-layer, addresses are called Physical addresses (MAC Address)

- The MAC addresses defines a specific host or router in a network (LAN or WAN).
- The MAC addresses are locally defined addresses.

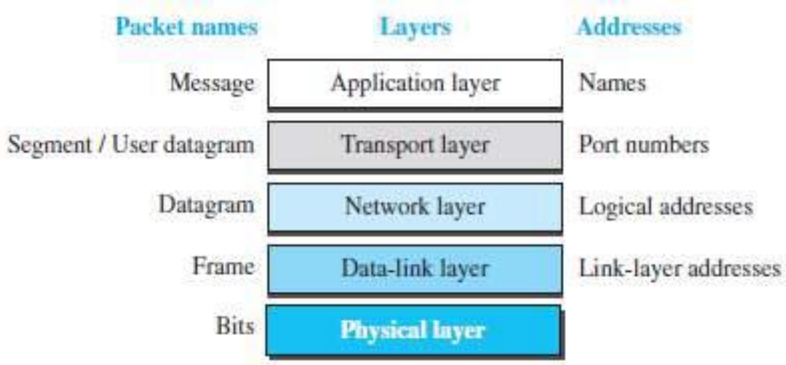


Figure 2.9 *Addressing in the TCP/IP protocol suite*

Multiplexing and Demultiplexing

• Multiplexing means a protocol at a layer can encapsulate a packet from several next-higher layer protocols (one at a time) (Figure 2.10).

• Demultiplexing means a protocol can decapsulate and deliver a packet to several next-higher layer protocols (one at a time).

1) At transport layer, either UDP or TCP can accept a message from several application-layer protocols.

2) At network layer, IP can accept

→ a segment from TCP or a user datagram from UDP.

→ a packet from ICMP or IGMP.

3) At data-link layer, a frame may carry the payload coming from IP or ARP.

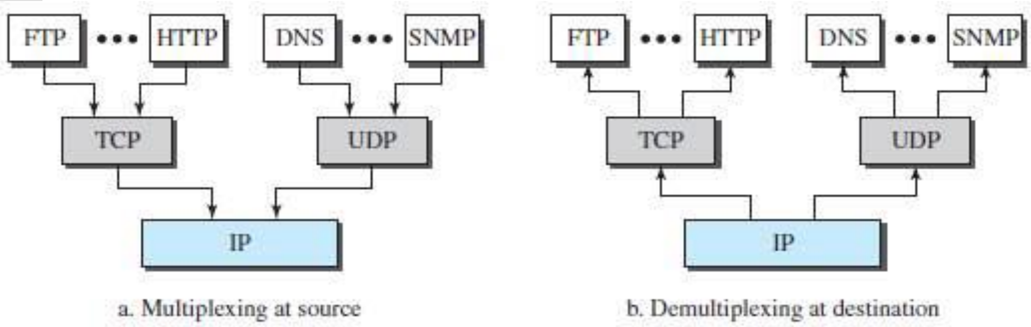


Figure 2.10 *Multiplexing and demultiplexing*

Canara Engineering College

OSI Model

- OSI model was developed by ISO.
- ISO is the organization, OSI is the model.
- Purpose: OSI was developed to allow systems with diff. platforms to communicate with each other.
- Platform means hardware, software or operating system.
- OSI is a network-model that defines the protocols for network communications.
- OSI has 7 layers as follows (Figure 2.11):
  1) Application Layer
  2) Presentation Layer
  3) Session Layer
  4) Transport Layer
  5) Network Layer
  6) Data Link Layer
  7) Physical Layer
- Each layer has specific duties to perform and has to co-operate with the layers above & below it.

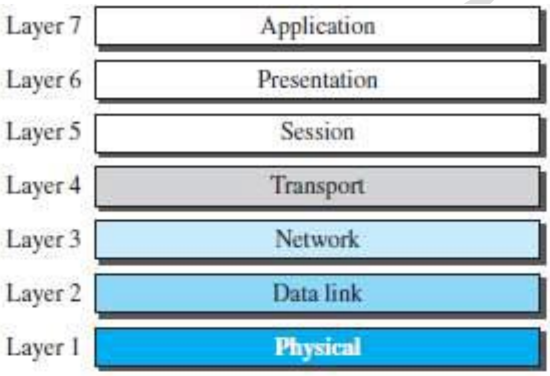


Figure 2.11 *The OSI model*

OSI vs. TCP/IP

1) The four bottommost layers in the OSI model & the TCP/IP model are same (Figure 2.12).
However, the Application-layer of TCP/IP model corresponds to the Session, Presentation & Application Layer of OSI model.
Two reasons for this are:
1) TCP/IP has more than one transport-layer protocol.
2) Many applications can be developed at Application layer
The OSI model specifies which functions belong to each of its layers.
In TCP/IP model, the layers contain relatively independent protocols that can be mixed and matched depending on the needs of the system.

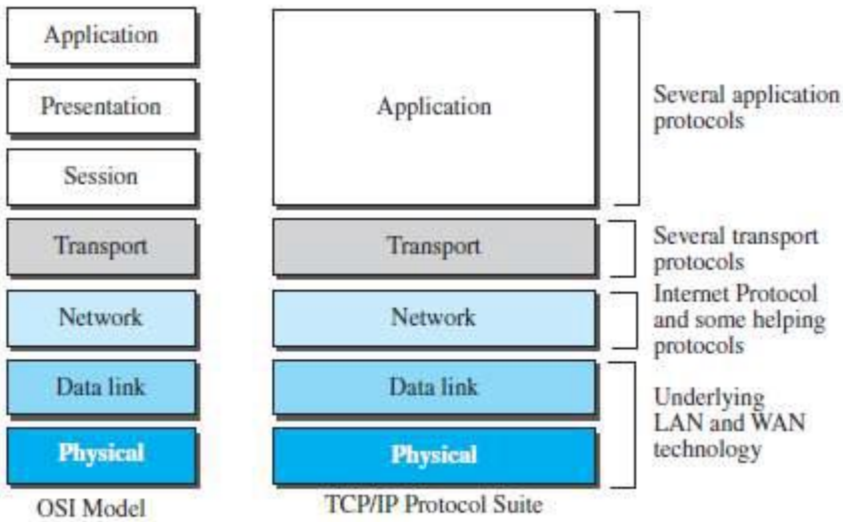


Figure 2.12 *TCP/IP and OSI model*

Lack of OSI Model's Success

• OSI was completed when TCP/IP was fully in place and a lot of time and money had been spent on the suite; changing it would cost a lot.

• Some layers in the OSI model were never fully defined.

• When OSI was implemented by an organization in a different application, it did not show a high enough level of performance

Canara Engineering College

DATA AND SIGNALS

Analog & Digital Data

• To be transmitted, data must be transformed to electromagnetic-signals.

• Data can be either analog or digital.

1) Analog Data refers to information that is continuous.

For example:The sounds made by a human voice.

2) Digital Data refers to information that has discrete states.

For example:Data are stored in computer-memory in the form of 0s and 1s.

Analog & Digital Signals

• Signals can be either analog or digital (Figure 3.2).

1) Analog Signal has infinitely many levels of intensity over a period of time.

2) Digital Signal can have only a limited number of defined values.

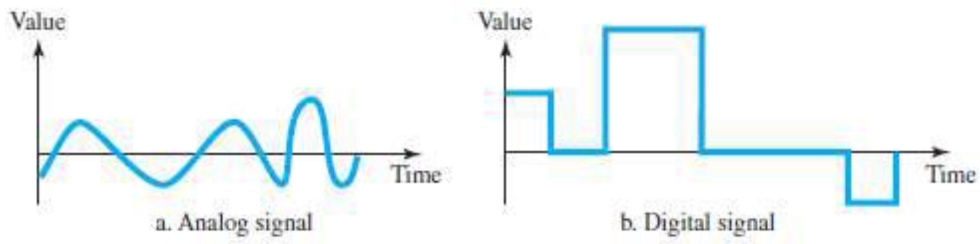


Figure 3.2 *Comparison of analog and digital signals*

Periodic & Non-Periodic Signals

• The signals can take one of 2 forms: periodic or non-periodic.

1) Periodic Signal

- Signals which repeat itself after a fixed time period are called Periodic Signals.
- The completion of one full pattern is called a cycle.

2) Non-Periodic Signal

- Signals which do not repeat itself after a fixed time period are called Non-Periodic Signals.

DIGITAL SIGNALS

• Information can be represented by a digital signal.

• For example:

1) 1 can be encoded as a positive voltage.0 can be encoded as a zero voltage (Figure 3.17a).

2) A digital signal can have more than 2 levels (Figure 3.17b).

Canara Engineering College

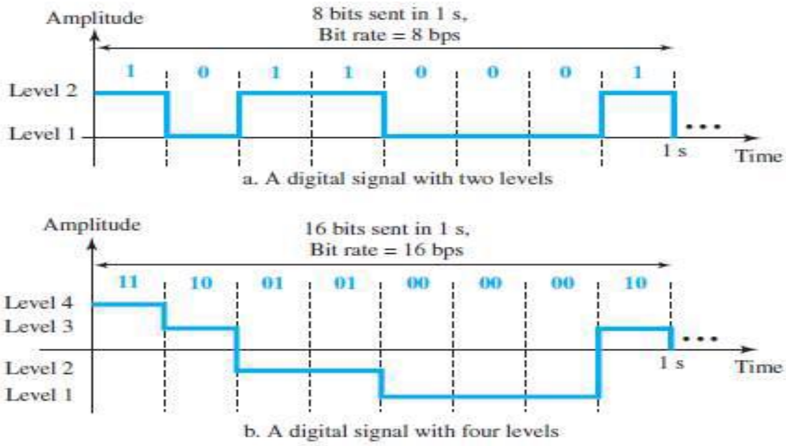


Figure 3.17 *Two digital signals: one with two signal levels and the other with four signal levels*

Qn 1. A digital signal has eight levels. How many bits are needed per level? We calculate the number of bits from the following formula. Each signal level is represented by 3 bits.

$$\textbf{Number of bits per level} = \log_2 8 = 3$$

Bit Rate

• The bit rate is the number of bits sent in 1s.

• The bit rate is expressed in bits per second (bps).

Qn 2. Assume we need to download text documents at the rate of 100 pages per second. What is the required bit rate of the channel?

Solution:

A page is an average of 24 lines with 80 characters in each line. If we assume that one character requires 8 bits, the bit rate is

$$100 \times 24 \times 80 \times 8 = 1{,}536{,}000 \text{ bps} = 1.536 \text{ Mbps}$$

Qn 3. A digitized voice channel, as we will see in Chapter 4, is made by digitizing a 4-kHz bandwidth analog voice signal. We need to sample the signal at twice the highest frequency (two samples per hertz). We assume that each sample requires 8 bits. What is the required bit rate?

Solution

The bit rate can be calculated as

$$2 \times 4000 \times 8 = 64{,}000 \text{ bps} = 64 \text{ kbps}$$

Bit Length

• The bit length is the distance one bit occupies on the transmission medium.

Bit length = propagation speed × bit duration

Digital Signal as a Composite Analog Signal

• A digital signal is a composite analog signal.
• A digital signal, in the time domain, comprises connected vertical and horizontal line segments.
1) A vertical line in the time domain means a frequency of infinity (sudden change in time);
2) A horizontal line in the time domain means a frequency of zero (no change in time).
• Fourier analysis can be used to decompose a digital signal.
1) If the digital signal is periodic, the decomposed signal has a frequency domain representation with an infinite bandwidth and discrete frequencies (Figure 3.18a).
2) If the digital signal is non-periodic, the decomposed signal has a frequency domain representation with an infinite bandwidth and continuous frequencies (Figure 3.18b).

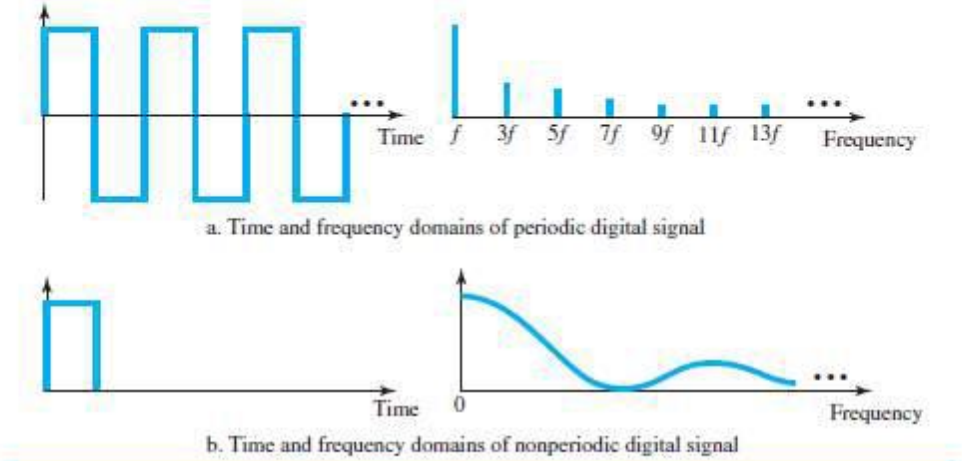


Figure 3.18 *The time and frequency domains of periodic and nonperiodic digital signals*

Transmission of Digital Signals

• Two methods for transmitting a digital signal:
1) Baseband transmission
2) Broadband transmission (using modulation).

Baseband Transmission

• Baseband transmission means sending a digital signal over a channel without changing the digital signal to an analog signal (Figure 3.19).

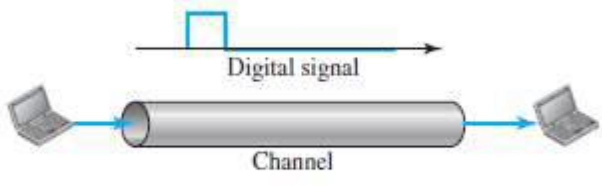


Figure 3.19 *Baseband transmission*

• Baseband transmission requires that we have a low-pass channel.
• Low-pass channel means a channel with a bandwidth that starts from zero.
• For example, we can have a dedicated medium with a bandwidth constituting only one channel.

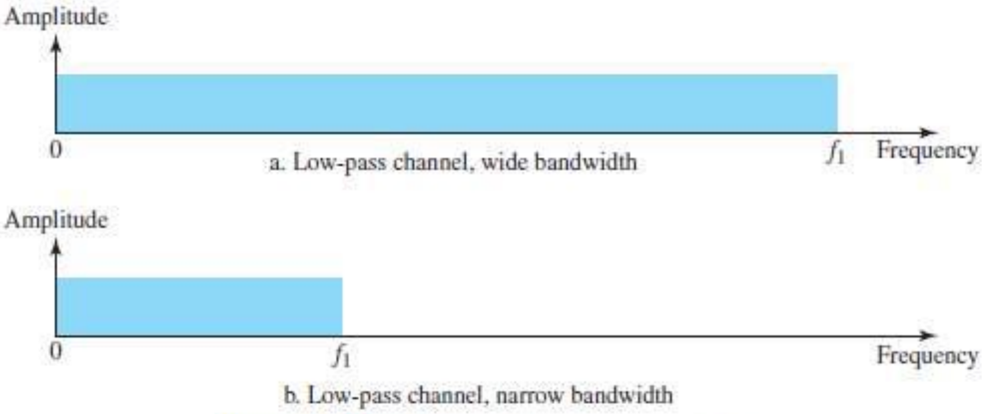


Figure 3.20 *Bandwidths of two low-pass channels*

• Two cases of a baseband communication:

Case 1: Low-pass channel with a wide bandwidth (Figure 3.20a)

Case 2: Low-pass channel with a limited bandwidth (Figure 3.20b)

Case 1: Low-Pass Channel with Wide Bandwidth

- To preserve the shape of a digital signal, we need to send the entire spectrum i.e. the continuous range of frequencies between zero and infinity.
- This is possible if we have a dedicated medium with an infinite bandwidth between the sender and receiver.
- If we have a medium with a very wide bandwidth, 2 stations can communicate by using digital signals with very good accuracy (Figure 3.21).
- Although the output signal is not an exact replica of the original signal, the data can still be deduced from the received signal.

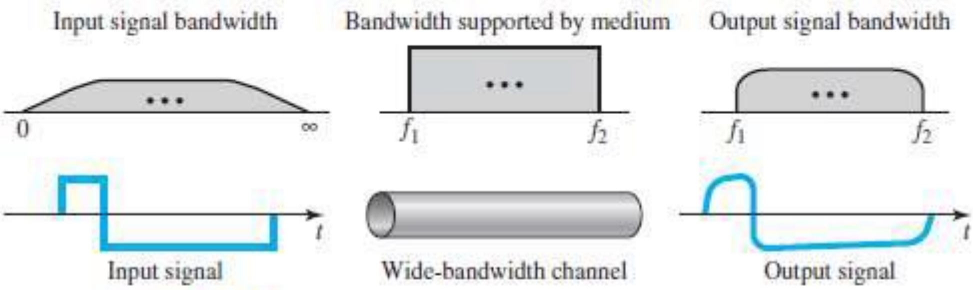


Figure 3.21 *Baseband transmission using a dedicated medium*

Case 2: Low-Pass Channel with Limited Bandwidth

- In a low-pass channel with limited bandwidth, we approximate the digital signal with an analog signal.
- The level of approximation depends on the bandwidth available.

A) Rough Approximation

¤ Assume that we have a digital signal of bit rate N (Figure 3.22).

¤ If we want to send analog signals to roughly simulate this signal, we need to consider the worst case, a maximum number of changes in the digital signal.

¤ This happens when the signal carries the sequence 01010101 . . or 10101010. . . .

¤ To simulate these two cases, we need an analog signal of frequency f = N/2.

¤ Let 1 be the positive peak value and 0 be the negative peak value.
¤ We send 2 bits in each cycle; the frequency of the analog signal is one-half of the bit rate, or N/2.
¤ This rough approximation is referred to as using the first harmonic (N/2) frequency.
The required bandwidth is

$$\text{Bandwidth} = \frac{N}{2} - 0 = \frac{N}{2}$$

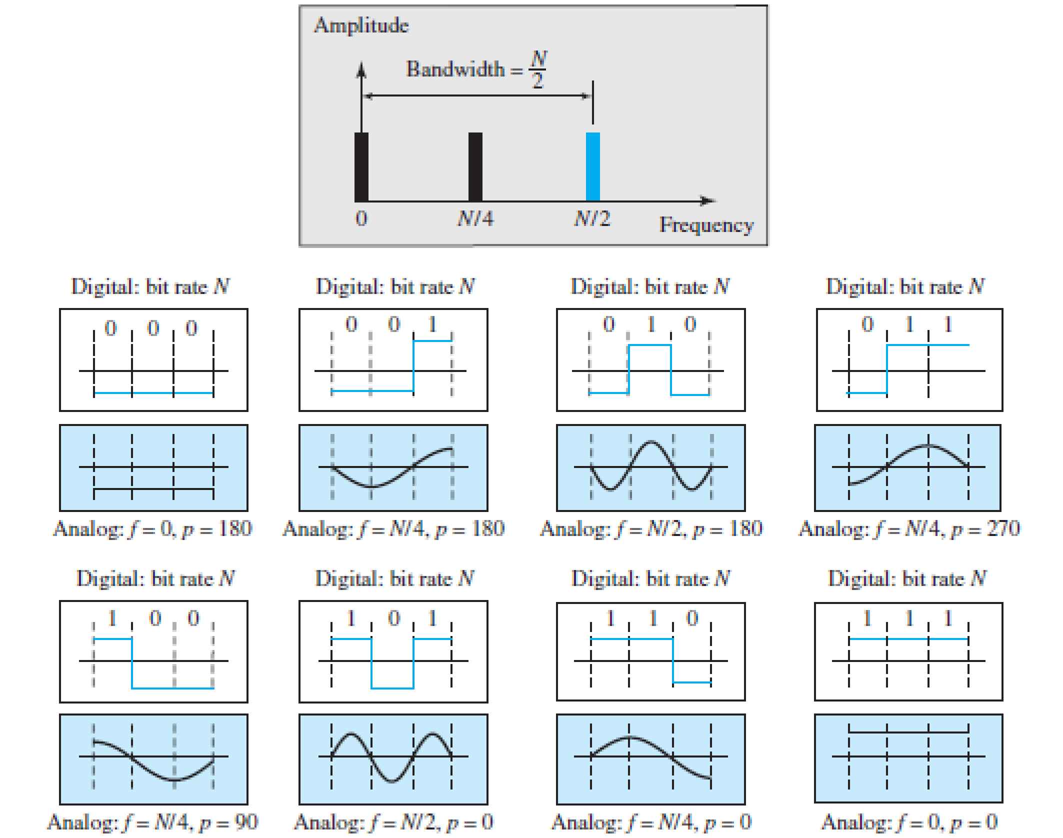


Figure 3.22 *Rough approximation of a digital signal using the first harmonic for worst case*

B) Better Approximation

¤ To make the shape of the analog signal look more like that of a digital signal, we need to add more harmonics of the frequencies (Figure 3.23).
¤ We can increase the bandwidth to 3N/2, 5N/2, 7N/2, and so on.
¤ In baseband transmission, the required bandwidth is proportional to the bit rate;
If we need to send bits faster, we need more bandwidth.

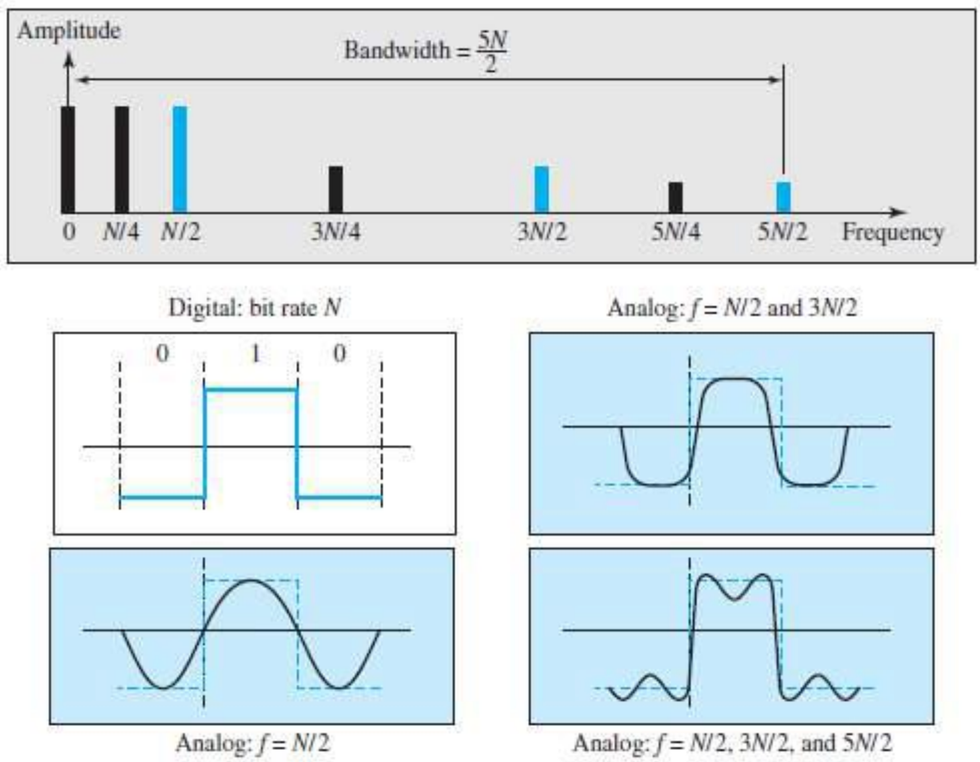


Figure 3.23 *Simulating a digital signal with first three harmonics*

Table 3.2 *Bandwidth requirements*

Bit Rate	*Harmonic 1*	*Harmonics 1, 3*	*Harmonics 1, 3, 5*
$n = 1$ kbps	$B = 500$ Hz	$B = 1.5$ kHz	$B = 2.5$ kHz
$n = 10$ kbps	$B = 5$ kHz	$B = 15$ kHz	$B = 25$ kHz
$n = 100$ kbps	$B = 50$ kHz	$B = 150$ kHz	$B = 250$ kHz

Qn.4. What is the required bandwidth of a low-pass channel if we need to send 1 Mbps by using baseband transmission?

Solution

The answer depends on the accuracy desired.

a. The minimum bandwidth, a rough approximation, is $B =$ bit rate /2, or 500 kHz. We need a low-pass channel with frequencies between 0 and 500 kHz.

b. A better result can be achieved by using the first and the third harmonics with the required bandwidth $B = 3 \times 500$ kHz $= 1.5$ MHz.

c. A still better result can be achieved by using the first, third, and fifth harmonics with $B = 5 \times 500$ kHz $= 2.5$ MHz.

Qn.5. We have a low-pass channel with bandwidth 100 kHz. What is the maximum bit rate of this channel?

Solution

The maximum bit rate can be achieved if we use the first harmonic. The bit rate is 2 times the available bandwidth, or 200 kbps.

Broadband Transmission (Using Modulation)

• Broadband transmission or modulation means changing the digital signal to an analog signal for transmission.
• Modulation allows us to use a bandpass channel (Figure 3.24).
• Bandpass channel means a channel with a bandwidth that does not start from zero.
• This type of channel is more available than a low-pass channel.

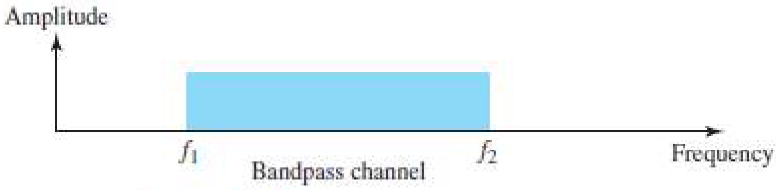


Figure 3.24 *Bandwidth of a bandpass channel*

• If the available channel is a bandpass channel,
We cannot send the digital signal directly to the channel;
We need to convert the digital signal to an analog signal before transmission (Figure 3.25).

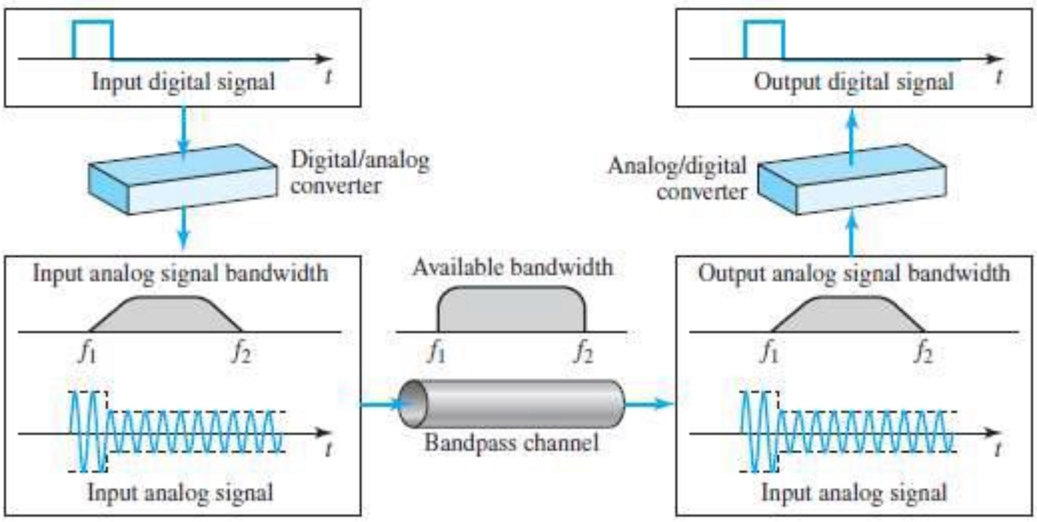
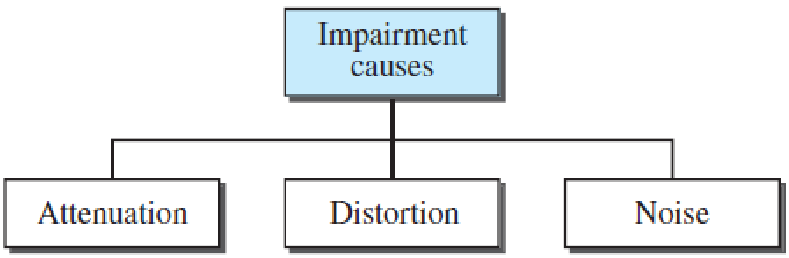


Figure 3.25 *Modulation of a digital signal for transmission on a bandpass channel*

TRANSMISSION IMPAIRMENT

• Signals travel through transmission media, which are not perfect.
• The imperfection causes signal-impairment.
• This means that signal at beginning of the medium is not the same as the signal at end of medium.
• What is sent is not what is received.
• Three causes of impairment are (Figure 3.26):
1) Attenuation
2) Distortion &
3) Noise.

Figure 3.26 *Causes of impairment*



Attenuation

• As signal travels through the medium, its strength decreases as distance increases. This is called attenuation (Figure 3.27).

• As the distance increases, attenuation also increases.

• For example:

Voice-data becomes weak over the distance & loses its contents beyond a certain distance.

• To compensate for this loss, amplifiers are used to amplify the signal.

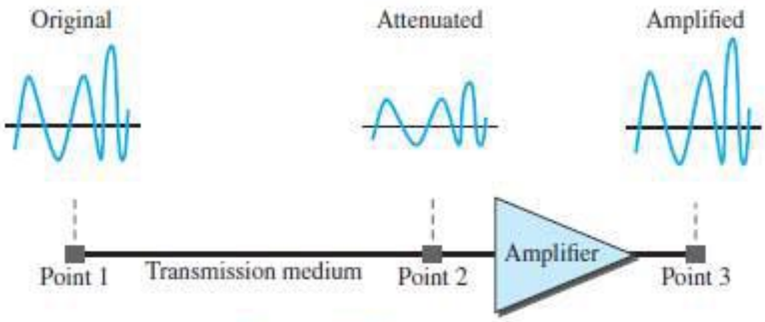


Figure 3.27 *Attenuation*

Decibel

• The decibel (dB) measures the relative strengths of

→ 2 signals or

→ one signal at 2 different points.

• The decibel is negative if a signal is attenuated.

The decibel is positive if a signal is amplified.

$$\mathrm{dB} = 10 \log_{10} \frac{P_2}{P_1}$$

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

• To show that a signal has lost or gained strength, engineers use the unit of decibel.

Qn. 6.

Suppose a signal travels through a transmission medium and its power is reduced to one-half. This means that $P_2 = \frac{1}{2} P_1$. In this case, the attenuation (loss of power) can be calculated as

$$10 \log_{10} \frac{P_2}{P_1} = 10 \log_{10} \frac{0.5P_1}{P_1} = 10 \log_{10} 0.5 = 10(-0.3) = -3 \text{ dB}$$

A loss of 3 dB (–3 dB) is equivalent to losing one-half the power.

Qn. 7. A signal travels through an amplifier, and its power is increased 10 times. Calculate power gained in dB

$$10 \log_{10} \frac{P_2}{P_1} = 10 \log_{10} \frac{10P_1}{P_1} = 10 \log_{10} 10 = 10(1) = 10 \text{ dB}$$

Qn.8.

Sometimes the decibel is used to measure signal power in milliwatts. In this case, it is referred to as dB_m and is calculated as $dB_m = 10 \log_{10} P_m$, where P_m is the power in milliwatts. Calculate the power of a signal if its $dB_m = -30$.

Solution

We can calculate the power in the signal as

$$dB_m = 10 \log_{10} \longrightarrow dB_m = -30 \longrightarrow \log_{10} P_m = -3 \longrightarrow P_m = 10^{-3} \text{ mW}$$

Qn. 9.

The loss in a cable is usually defined in decibels per kilometer (dB/km). If the signal at the beginning of a cable with –0.3 dB/km has a power of 2 mW, what is the power of the signal at 5 km?

Solution

The loss in the cable in decibels is $5 \times (-0.3) = -1.5$ dB. We can calculate the power as

$$dB = 10 \log_{10} (P_2 / P_1) = -1.5 \longrightarrow (P_2 / P_1) = 10^{-0.15} = 0.71$$

$$P_2 = 0.71P_1 = 0.7 \times 2 \text{ mW} = 1.4 \text{ mW}$$

Distortion

- Distortion means that the signal changes its form or shape (Figure 3.29).
- Distortion can occur in a composite signal made of different frequencies.
- Different signal-components
  - → have different propagation speed through a medium.
  - → have different delays in arriving at the final destination.
- Differences in delay create a difference in phase if delay is not same as the period-duration.
- Signal-components at the receiver have phases different from what they had at the sender.

• The shape of the composite signal is therefore not the same.

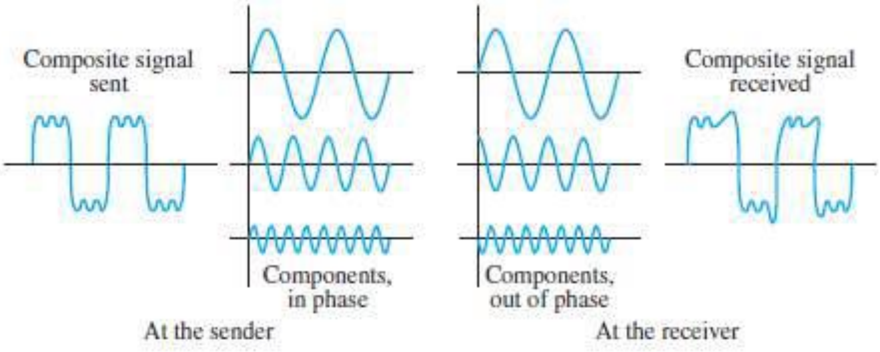


Figure 3.29 *Distortion*

Noise

• Noise is defined as an unwanted data (Figure 3.30).
• In other words, noise is the external energy that corrupts a signal.
• Due to noise, it is difficult to retrieve the original data/information.
• Four types of noise:

i) Thermal Noise

- It is random motion of electrons in wire which creates extra signal not originally sent by transmitter.

ii) Induced Noise

- Induced noise comes from sources such as motors & appliances.
- These devices act as a sending-antenna. The transmission-medium acts as the receiving-antenna.

iii) Crosstalk

- Crosstalk is the effect of one wire on the other.
- One wire acts as a sending-antenna and the other as the receiving-antenna.

iv) Impulse Noise

Impulse Noise is a spike that comes from power-lines, lightning, and so on. (spike - a signal with high energy in a very short time)

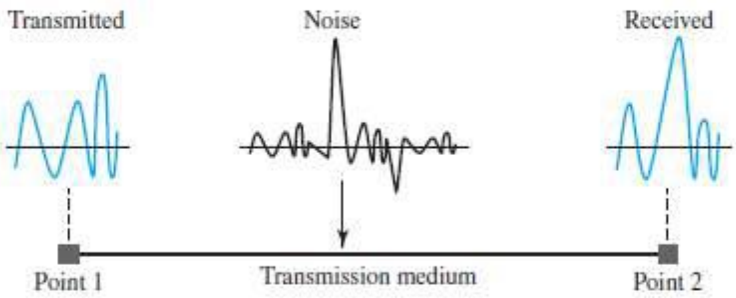


Figure 3.30 *Noise*

Signal-to-Noise Ratio (SNR)

• SNR is used to find the theoretical bit-rate limit.
• SNR is defined as

$$SNR = \frac{\text{average signal power}}{\text{average noise power}}$$

• SNR is actually the ratio of what is wanted (signal) to what is not wanted (noise).

• A high-SNR means the signal is less corrupted by noise.
A low-SNR means the signal is more corrupted by noise.
• Because SNR is the ratio of 2 powers, it is often described in decibel units, SNRdB, defined as

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

Example 3.31

The power of a signal is 10 mW and the power of the noise is 1 μW; what are the values of SNR and SNR_{dB}?

Solution

The values of SNR and SNR_{dB} can be calculated as follows:

$$SNR = (10{,}000\ \mu w) / (1\ \mu w) = 10{,}000 \qquad SNR_{dB} = 10 \log_{10} 10{,}000 = 10 \log_{10} 10^4 = 40$$

Example 3.32

The values of SNR and SNR_{dB} for a noiseless channel are

$$SNR = (\text{signal power}) / 0 = \infty \longrightarrow SNR_{dB} = 10 \log_{10} \infty = \infty$$

We can never achieve this ratio in real life; it is an ideal.

DATA RATE LIMITS

• Data-rate depends on 3 factors:
1) Bandwidth available
2) Level of the signals
3) Quality of channel (the level of noise)
• Two theoretical formulas can be used to calculate the data-rate:
1) Nyquist for a noiseless channel and
2) Shannon for a noisy channel.

Noiseless Channel: Nyquist Bit-rate

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

$$\text{Bitrate} = 2 \times \text{Bandwidth} \times \log_2 L$$

where bandwidth = bandwidth of the channel
L = number of signal-levels used to represent data
BitRate = bitrate of channel in bps

- According to the formula,
  ¤ By increasing number of signal-levels, we can increase the bit-rate.
  ¤ Although the idea is theoretically correct, practically there is a limit.
  ¤ When we increase the number of signal-levels, we impose a burden on the receiver.

¤ If no. of levels in a signal is 2, the receiver can easily distinguish b/w 0 and 1.
¤ If no. of levels is 64, the receiver must be very sophisticated to distinguish b/w 64 different levels.
¤ In other words, increasing the levels of a signal reduces the reliability of the system.

Example 3.34

Consider a noiseless channel with a bandwidth of 3000 Hz transmitting a signal with two signal levels. The maximum bit rate can be calculated as

$$\mathbf{BitRate = 2 \times 3000 \times \log_2 2 = 6000\ bps}$$

Example 3.35

Consider the same noiseless channel transmitting a signal with four signal levels (for each level, we send 2 bits). The maximum bit rate can be calculated as

$$\mathbf{BitRate = 2 \times 3000 \times \log_2 4 = 12{,}000\ bps}$$

Example 3.36

We need to send 265 kbps over a noiseless channel with a bandwidth of 20 kHz. How many signal levels do we need?

Solution

We can use the Nyquist formula as shown:

$$\mathbf{265{,}000 = 2 \times 20{,}000 \times \log_2 L \longrightarrow \log_2 L = 6.625 \longrightarrow L = 2^{6.625} = 98.7\ levels}$$

Since this result is not a power of 2, we need to either increase the number of levels or reduce the bit rate. If we have 128 levels, the bit rate is 280 kbps. If we have 64 levels, the bit rate is 240 kbps.

Noisy Channel: Shannon Capacity

- In reality, we cannot have a noiseless channel; the channel is always noisy.
- For a noisy channel, the Shannon capacity formula defines the theoretical maximum bit-rate.

$$\text{Capacity} = \text{bandwidth} \times \log_2 (1 + \text{SNR})$$

where bandwidth = bandwidth of channel in bps.
SNR = signal-to-noise ratio and
Capacity = capacity of channel in bps.

- This formula does not consider the no. of levels of signals being transmitted (as done in the Nyquist bit rate).
- This means that no matter how many levels we have, we cannot achieve a data-rate higher than the capacity of the channel.
- In other words, the formula defines a characteristic of the channel, not the method of transmission.

Example 3.37

Consider an extremely noisy channel in which the value of the signal-to-noise ratio is almost zero. In other words, the noise is so strong that the signal is faint. For this channel the capacity C is calculated as

$$C = B \log_2 (1 + \text{SNR}) = B \log_2(1 + 0) = B \log_2 1 = B \times 0 = 0$$

Example 3.38

We can calculate the theoretical highest bit rate of a regular telephone line. A telephone line normally has a bandwidth of 3000 Hz (300 to 3300 Hz) assigned for data communications. The signal-to-noise ratio is usually 3162. For this channel the capacity is calculated as

$$C = B \log_2 (1 + \text{SNR}) = 3000 \log_2(1 + 3162) = 3000 \times 11.62 = 34{,}860 \text{ bps}$$

This means that the highest bit rate for a telephone line is 34.860 kbps. If we want to send data faster than this, we can either increase the bandwidth of the line or improve the signal-to-noise ratio.

Example 3.39

The signal-to-noise ratio is often given in decibels. Assume that $\text{SNR}_{\text{dB}} = 36$ and the channel bandwidth is 2 MHz. The theoretical channel capacity can be calculated as

$$\text{SNR}_{\text{dB}} = 10 \log_{10}\text{SNR} \longrightarrow \text{SNR} = 10^{\text{SNR}_{\text{dB}}/10} \longrightarrow \text{SNR} = 10^{3.6} = 3981$$

$$C = B \log_2(1 + \text{SNR}) = 2 \times 10^6 \times \log_2 3982 = 24 \text{ Mbps}$$

Using Both Limits

In practice, we need to use both methods to find the limits and signal levels. Let us show this with an example.

Example 3.41

We have a channel with a 1-MHz bandwidth. The SNR for this channel is 63. What are the appropriate bit rate and signal level?

Solution

First, we use the Shannon formula to find the upper limit.

$$C = B \log_2(1 + \text{SNR}) = 10^6 \log_2(1 + 63) = 10^6 \log_2 64 = 6 \text{ Mbps}$$

The Shannon formula gives us 6 Mbps, the upper limit. For better performance we choose something lower, 4 Mbps, for example. Then we use the Nyquist formula to find the number of signal levels.

$$4 \text{ Mbps} = 2 \times 1 \text{ MHz} \times \log_2 L \longrightarrow L = 4$$

- **The Shannon capacity gives us the upper limit; the Nyquist formula tells us how many signal levels we need.**

PERFORMANCE

Bandwidth

• One characteristic that measures network-performance is bandwidth.
• Bandwidth of analog and digital signals is calculated in separate ways:

(1) Bandwidth of an Analog Signal (in hz)

- Bandwidth of an analog signal is expressed in terms of its frequencies.
- Bandwidth is defined as the range of frequencies that the channel can carry.
- It is calculated by the difference b/w the maximum frequency and the minimum frequency.

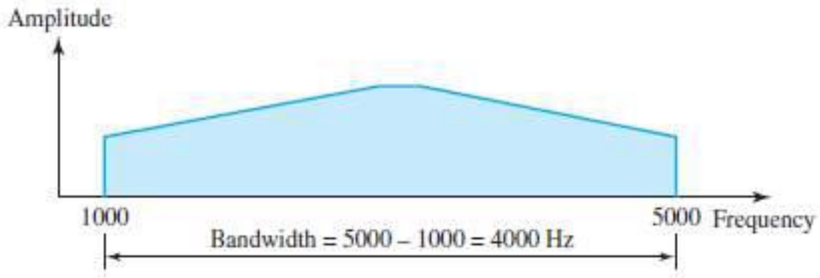


Figure 3.13 *The bandwidth of signals*

In figure 3.13, the signal has a minimum frequency of F1 = 1000Hz and maximum frequency of F2 = 5000Hz.
Hence, the bandwidth is given by F2 – F1= 5000 - 1000 = 4000 Hz

(2) Bandwidth of a Digital Signal (in bps)

Bandwidth refers to the number of bits transmitted in one second in a channel (or link).
For example:
The bandwidth of a Fast Ethernet is a maximum of 100 Mbps. (This means that this network can send 100 Mbps).

Relationship between (1) and (2)

• There is an explicit relationship between the bandwidth in hertz and bandwidth in bits per seconds.
• Basically, an increase in bandwidth in hertz means an increase in bandwidth in bits per second.
• The relationship depends on
→ baseband transmission or
→ transmission with modulation.

Throughput

• The throughput is a measure of how fast we can actually send data through a network.
• Although, bandwidth in bits per second and throughput seem the same, they are actually different.
• A link may have a bandwidth of B bps, but we can only send T bps through this link with T always less than B.
• In other words,
1) The bandwidth is a potential measurement of a link.
2) The throughput is an actual measurement of how fast we can send data.

For example:

¤ We may have a link with a bandwidth of 1 Mbps, but the devices connected to the end of the link may handle only 200 kbps.

¤ This means that we cannot send more than 200 kbps through this link.

Example 3.44

A network with bandwidth of 10 Mbps can pass only an average of 12,000 frames per minute with each frame carrying an average of 10,000 bits. What is the throughput of this network?

Solution

We can calculate the throughput as

$$\textbf{Throughput} = (12{,}000 \times 10{,}000) / 60 = 2\ \textbf{Mbps}$$

The throughput is almost one-fifth of the bandwidth in this case.

Latency (Delay)

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

$$\textbf{Latency} = \textbf{propagation time} + \textbf{transmission time} + \textbf{queuing time} + \textbf{processing delay}$$

Propagation Time

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

$$\textbf{Propagation time} = \textbf{Distance} / (\textbf{Propagation Speed})$$

The propagation speed of electromagnetic signals depends on the medium and on the frequency of the signal. For example, in a vacuum, light is propagated with a speed of 3×10^8 m/s. It is lower in air; it is much lower in cable.

Example 3.45

What is the propagation time if the distance between the two points is 12,000 km? Assume the propagation speed to be 2.4×10^8 m/s in cable.

Solution

We can calculate the propagation time as

$$\textbf{Propagation time} = (12{,}000 \times 10{,}000) / (2.4 \times 2^8) = 50\ \textbf{ms}$$

The example shows that a bit can go over the Atlantic Ocean in only 50 ms if there is a direct cable between the source and the destination.

Transmission Time

- The time required for transmission of a message depends on
  → size of the message and
  → bandwidth of the channel.
- The transmission time is given by

$$\textbf{Transmission time} = (\textbf{Message size}) / \textbf{Bandwidth}$$

Example 3.46

What are the propagation time and the transmission time for a 2.5-KB (kilobyte) message (an e-mail) if the bandwidth of the network is 1 Gbps? Assume that the distance between the sender and the receiver is 12,000 km and that light travels at 2.4×10^8 m/s.

Solution

We can calculate the propagation and transmission time as

$$\text{Propagation time} = (12{,}000 \times 1000) / (2.4 \times 10^8) = 50 \text{ ms}$$

$$\text{Transmission time} = (2500 \times 8) / 10^9 = 0.020 \text{ ms}$$

Note that in this case, because the message is short and the bandwidth is high, the dominant factor is the propagation time, not the transmission time. The transmission time can be ignored.

Example 3.47

What are the propagation time and the transmission time for a 5-MB (megabyte) message (an image) if the bandwidth of the network is 1 Mbps? Assume that the distance between the sender and the receiver is 12,000 km and that light travels at 2.4×10^8 m/s.

Solution

We can calculate the propagation and transmission times as

$$\text{Propagation time} = (12{,}000 \times 1000) / (2.4 \times 10^8) = 50 \text{ ms}$$

$$\text{Transmission time} = (5{,}000{,}000 \times 8) / 10^6 = 40 \text{ s}$$

Note that in this case, because the message is very long and the bandwidth is not very high, the dominant factor is the transmission time, not the propagation time. The propagation time can be ignored.

Queuing Time

- Queuing-time is the time needed for each intermediate-device to hold the message before it can be processed.

(Intermediate device may be a router or a switch)

- The queuing-time is not a fixed factor. This is because
  i) Queuing-time changes with the load imposed on the network.

ii) When there is heavy traffic on the network, the queuing-time increases.

- An intermediate-device
  → queues the arrived messages and
  → processes the messages one by one.
- If there are many messages, each message will have to wait.

Processing Delay

- Processing delay is the time taken by the routers to process the packet header.

Bandwidth Delay Product

• Two performance-metrics of a link are 1) Bandwidth and 2) Delay
• The bandwidth-delay product is very important in data-communications.
• Let us elaborate on this issue, using 2 hypothetical cases as examples.
Case 1: The following figure shows case 1 (Figure 3.32).

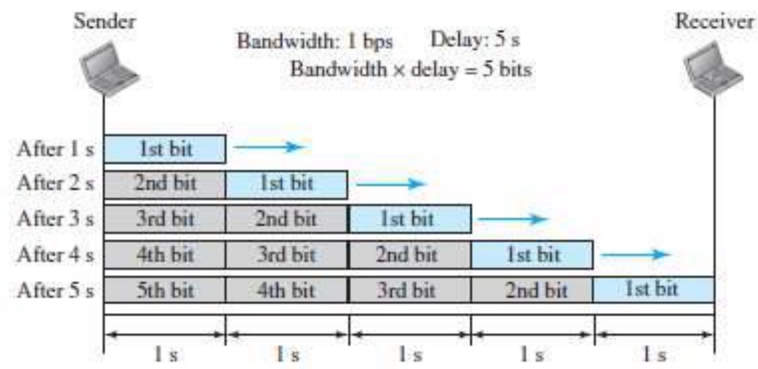


Figure 3.32 *Filling the link with bits for case 1*

Let us assume,
Bandwidth of the link = 1 bps Delay of the link = 5s.
From the figure 3.32, bandwidth-delay product is 1 x 5 = 5. Thus, there can be maximum 5 bits on the line. There can be no more than 5 bits at any time on the link.

Case 2: The following figure shows case 2 (Figure 3.33).

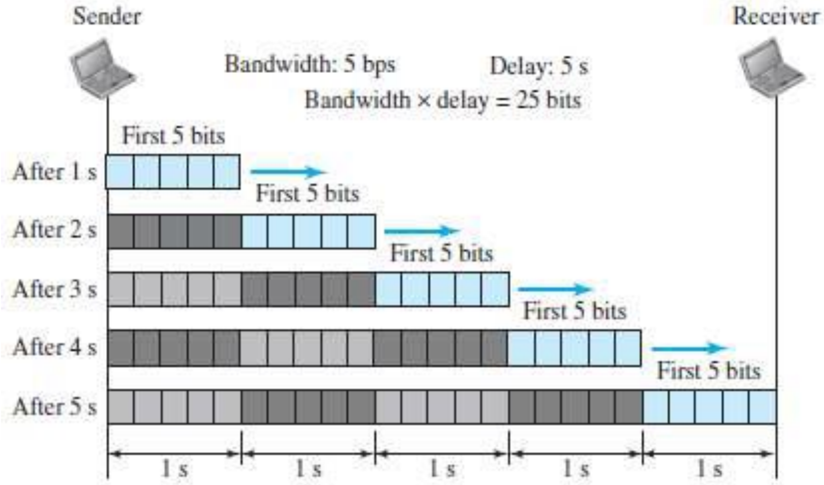


Figure 3.33 *Filling the link with bits in case 2*

Let us assume,
Bandwidth of the link = 4 bps Delay of the link = 5s.
From the figure 3.33, bandwidth-delay product is 5 x 5 = 25. Thus, there can be maximum 25 bits on the line.
At each second, there are 5 bits on the line, thus the duration of each bit is 0.20s.

- The above 2 cases show that the (bandwidth X delay) is the number of bits that can fill the link.
- This measurement is important if we need to
  - → send data in bursts and
  - → wait for the acknowledgment of each burst.
- To use the maximum capability of the link
  - → We need to make the burst-size as (2 x bandwidth x delay).
  - → We need to fill up the full-duplex channel (two directions).
- Amount (2x bandwidth x delay) is the number of bits that can be in transition at any time (Fig 3.34).

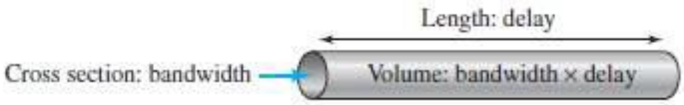


Figure 3.34 *Concept of bandwidth-delay product*

Jitter

- Another performance issue that is related to delay is jitter.
- We can say that jitter is a problem
  - → if different packets of data encounter different delays and
  - → if the application using the data at the receiver site is time-sensitive (for ex: audio/video).
- For example:

If the delay for the first packet is 20ms the delay for the second is 45ms and the delay for the third is 40ms then the real-time application that uses the packets suffers from jitter.

Canara Engineering College

DIGITAL TRANSMISSION

DIGITAL TO DIGITAL CONVERSION

• Data can be analog or digital, so can be the signal that represents it.
• Signal encoding is the conversion from analog/digital data to analog/digital signal.
• The possible encodings are:
1) Digital data to digital signal
2) Digital data to analog signal
3) Analog data to digital signal
4) Analog data to analog signal

LINE CODING

• Line-coding is the process of converting digital-data to digital-signals (Figure 4.1).
• The data may be in the form of text, numbers, graphical images, audio, or video
• The data are stored in computer memory as sequences of bits (0s or 1s).
• Line-coding converts a sequence of bits to a digital-signal.
• At the sender, digital-data is encoded into a digital-signal.
At the receiver, digital-signal is decoded into a digital-data.

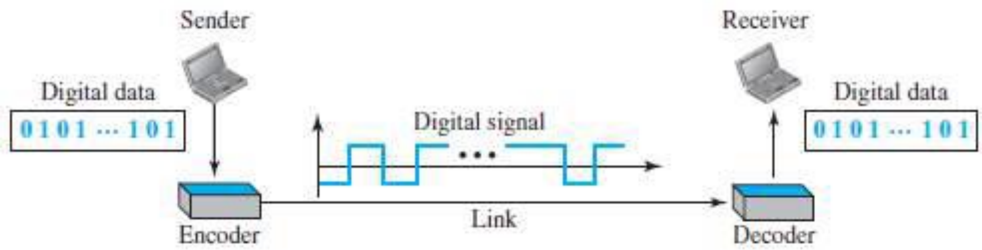


Figure 4.1 *Line coding and decoding*

Characteristics

• Different characteristics of digital signal are
1) Signal Element vs Data Element
2) Data Rate vs Signal Rate
3) Bandwidth
4) Baseline Wandering
5) DC Components
6) Built-in Error Detection
7) Self-synchronization
8) Immunity to Noise and Interference
9) Complexity

Canara Engineering College

Data Element vs. Signal Element

Data Element	Signal Element
A data-element is the smallest entity that can represent a piece of information (Figure 4.2).	A signal-element is shortest unit (timewise) of a digital-signal.
A data-element is the bit.	A signal-element carries data-elements.
Data-elements are being carried.	Signal-elements are the carriers.

- Ratio r is defined as number of data-elements carried by each signal-element.

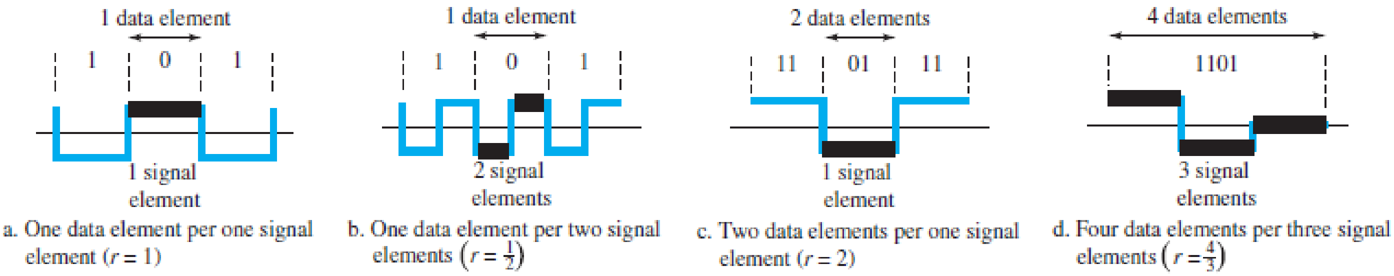


Figure 4.2 *Signal element versus data element*

Data Rate vs. Signal Rate

Data Rate	Signal Rate
The data-rate defines the number of data-elements (bits) sent in 1 sec.	The signal-rate is the number of signal-elements sent in 1 sec.
The unit is bits per second (bps).	The unit is the baud.
The data-rate is sometimes called the bit-rate.	The signal-rate is sometimes called the pulse rate, the modulation rate, or the baud rate
Goal in data-communications: increase the data-rate.	Goal in data-communications: decrease the signal-rate.
Increasing the data-rate increases the speed of transmission.	Decreasing the signal-rate decreases the bandwidth requirement.

- The relationship between data-rate and signal-rate is given by

$$S_{ave} = c \times N \times (1/r) \quad \text{baud}$$

where N = data-rate (in bps)
c = case factor, which varies for each case
S = number of signal-elements and
r = previously defined factor.

This relationship depends on
→ value of r.
→ data pattern.
(If we have a data pattern of all 1s or all 0s, the signal-rate may be different from a data pattern of alternating Os and 1s).

Example 4.1

A signal is carrying data in which one data element is encoded as one signal element ($r = 1$). If the bit rate is 100 kbps, what is the average value of the baud rate if c is between 0 and 1?

Solution

We assume that the average value of c is 1/2. The baud rate is then

$$S = c \times N \times (1 / r) = 1/2 \times 100{,}000 \times (1/1) = 50{,}000 = 50 \text{ kbaud}$$

Bandwidth

- Digital signal that carries information is non-periodic.
- The bandwidth of a non-periodic signal is continuous with an infinite range. However, most digital-signals we encounter in real life have a bandwidth with finite values.
- The effective bandwidth is finite.
- The baud rate, not the bit-rate, determines the required bandwidth for a digital-signal.
- More changes in the signal mean injecting more frequencies into the signal. (Frequency means change and change means frequency.)
- The bandwidth refers to range of frequencies used for transmitting a signal.
- Relationship b/w baud rate (signal-rate) and the bandwidth (range of frequencies) is given as

$$B_{min} = c \times N \times (1 / r)$$

where N = data-rate (in bps)
c = case factor, which varies for each case
r = previously defined factor
B_{min} = minimum bandwidth

Baseline Wandering

- While decoding, the receiver calculates a running-average of the received signal-power. This average is called the baseline.
- The incoming signal-power is estimated against this baseline to determine the value of the data-element.
- A long string of 0s or 1s can cause a drift in the baseline (baseline wandering). Thus, make it difficult for the receiver to decode correctly.
- A good line-coding scheme needs to prevent baseline wandering.

DC Components

- When the voltage-level in a digital-signal is constant for a while, the spectrum creates very low frequencies.
- These frequencies around zero are called DC (direct-current) components.
- DC components present problems for a system that cannot pass low frequencies.
- For example: Telephone line cannot pass frequencies below 200 Hz.
- For Telephone systems, we need a scheme with no DC component.

Canara Engineering College

Built-in Error Detection

Built-in error-detecting capability has to be provided to detect the errors that occurred during transmission.

Self Synchronization

- To correctly interpret the signals received from the sender, the receiver's bit intervals must correspond exactly to the sender's bit intervals.
- If the receiver clock is faster or slower, the bit intervals are not matched and the receiver might misinterpret the signals.

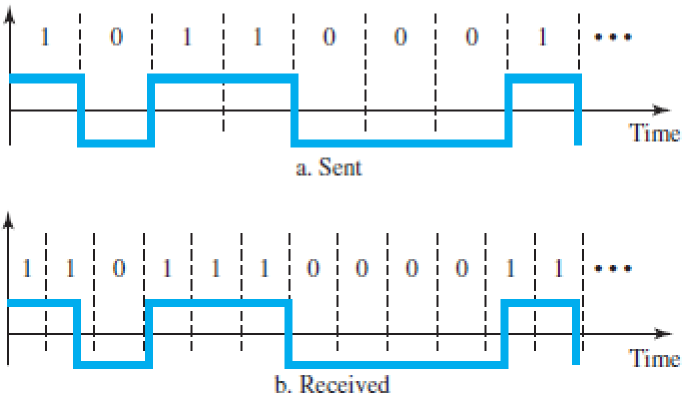


Figure 4.3 *Effect of lack of synchronization*

- As shown in figure 4.3, we have a situation where the receiver has shorter bit duration. The sender sends 10110001, while the receiver receives 110111000011.
- A self-synchronizing digital-signal includes timing-information in the data being transmitted.
  ¤ This can be achieved if there are transitions in the signal that alert the receiver to the beginning, middle, or end of the pulse.
  ¤ If the receiver's clock is out-of-synchronization, these points can reset the clock.

Immunity to Noise & Interference

- The code should be immune to noise and other interferences.

Complexity

- A complex scheme is more costly to implement than a simple one. For ex: A scheme that uses 4 signal-levels is more difficult to interpret than one that uses only 2 levels.

Example 4.3

In a digital transmission, the receiver clock is 0.1 percent faster than the sender clock. How many extra bits per second does the receiver receive if the data rate is 1 kbps? How many if the data rate is 1 Mbps?

Solution

At 1 kbps, the receiver receives 1001 bps instead of 1000 bps.

Canara Engineering College

1000 bits sent	→	1001 bits received	→	1 extra bps

At 1 Mbps, the receiver receives 1,001,000 bps instead of 1,000,000 bps.

1,000,000 bits sent	→	1,001,000 bits received	→	1000 extra bps

LINE CODING SCHEMES

• The Line Coding schemes are classified into 3 broad categories (Figure 4.4):

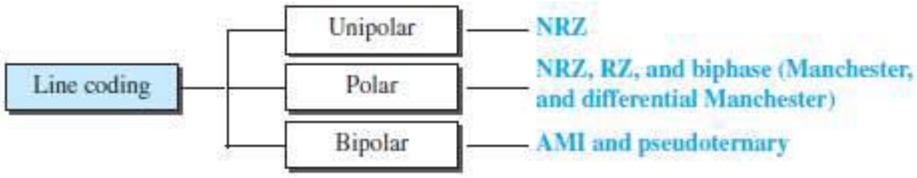


Figure 4.4 *Line coding schemes*

Unipolar Scheme

• All signal levels are either above or below the time axis.

NRZ (Non-Return-to-Zero)

- The positive voltage defines bit 1 and the zero voltage defines bit 0 (Figure 4.5).
- It is called NRZ because the signal does not return to 0 at the middle of the bit.

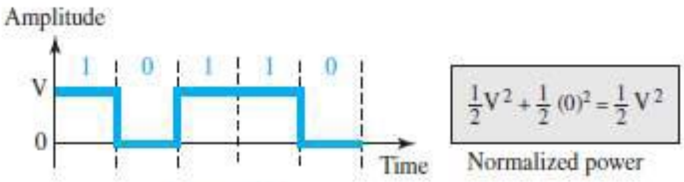


Figure 4.5 *Unipolar NRZ scheme*

Disadvantages:

1) Compared to polar scheme, this scheme is very costly.
2) Also, the normalized power is double that for polar NRZ.
3) Not suitable for transmission over channels with poor performance around zero frequency.

(Normalized power -power needed to send 1 bit per unit line resistance)

Polar Schemes

• The voltages are on the both sides of the time axis.

• Polar NRZ scheme can be implemented with two voltages (V).

For example: -V for bit 1

+V for bit 0.

Non-Return-to-Zero (NRZ)

- We use 2 levels of voltage amplitude.
- Two versions of polar NRZ (Figure 4.6):

i) NRZ-L (NRZ-Level)

¤ The level of the voltage determines the value of the bit.
¤ For example: i) Voltage-level for 0 can be positive and
ii) Voltage-level for 1 can be negative.

ii) NRZ-I (NRZ-Invert)

¤ The change or lack of change in the level of the voltage determines the value of the bit.
¤ If there is no change, the bit is 0;
If there is a change, the bit is 1.

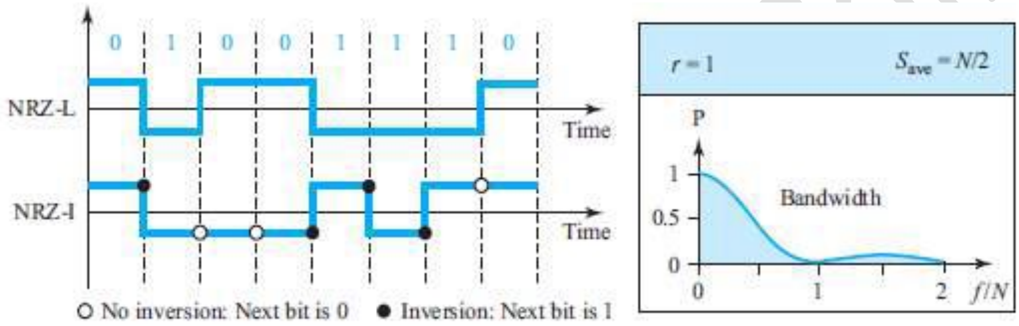


Figure 4.6 *Polar NRZ-L and NRZ-I schemes*

Disadvantages:
1) **Baseline wandering** is a problem for both variations (NRZ-L NRZ-I).
i) In NRZ-L, if there is a long sequence of 0s or 1s, the average signal-power becomes skewed.
The receiver might have difficulty discerning the bit value.
ii) In NRZ-I, this problem occurs only for a long sequence of 0s.
If we eliminate the long sequence of 0s, we can avoid baseline wandering.

2) The **synchronization problem** also exists in both schemes.
→ A long sequence of 0s can cause a problem in both schemes.
→ A long sequence of 1s can cause a problem in only NRZ-L.

3) In NRZ-L, problem occurs when there is a sudden **change of polarity** in the system.
For example:
In twisted-pair cable, a change in the polarity of the wire results in
→ all 0s interpreted as 1s and
→ all 1s interpreted as 0s.
¤ NRZ-I does not have this problem.
¤ Both schemes have an average signal-rate of N/2 Bd.
4) NRZ-L and NRZ-I both have a **DC component problem**.

Example 4.4

A system is using NRZ-I to transfer 10-Mbps data. What are the average signal rate and minimum bandwidth?

Solution

The average signal rate is $S = N/2 = 500$ kbaud. The minimum bandwidth for this average baud rate is $B_{min} = S = 500$ kHz.

Return-to-Zero (RZ)

- In NRZ encoding, problem occurs when the sender-clock and receiver-clock are not synchronized.
- Solution: Use return-to-zero (RZ) scheme (Figure 4.7).
- RZ scheme uses 3 voltages: positive, negative, and zero.
- There is always a transition at the middle of the bit. Either
  i) from high to zero (for 1) or
  ii) from low to zero (for 0)

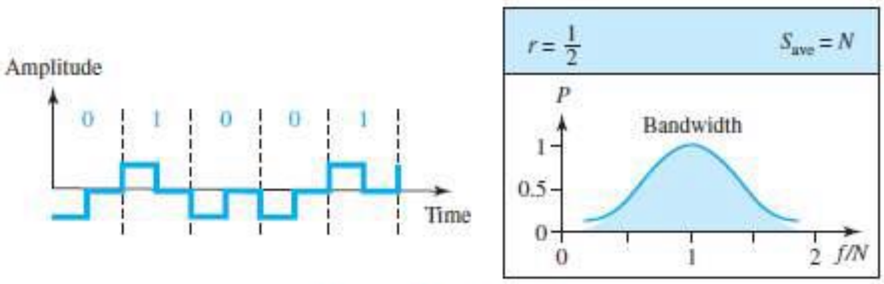


Figure 4.7 *Polar RZ scheme*

Disadvantages:
1) RZ encoding requires 2 signal-changes to encode a bit and occupies greater bandwidth.
2) Complexity: RZ uses 3 levels of voltage, which is more complex to create and detect.
3) Problem occurs when there is a sudden change of polarity in the system. This result in
  → all 0s interpreted as 1s &
  → all 1s interpreted as 0s.

Biphase: Manchester & Differential Manchester

i) Manchester Encoding

- This is a combination of NRZ-L & RZ schemes (RZ-transition at the middle of the bit).
- There is always a transition at the middle of the bit. Either
  i) from high to low (for 0) or
  ii) from low to high (for 1).
- It uses only two voltage levels (Figure 4.8).
- The duration of the bit is divided into 2 halves.
- The voltage
  → remains at one level during the first half &

Canara Engineering College

→ moves to the other level in the second half.

- The transition at the middle of the bit provides synchronization.

ii) Differential Manchester

- This is a combination of NRZ-I and RZ schemes.
- There is always a transition at the middle of the bit, but the bit-values are determined at the beginning of the bit.
- If the next bit is 0, there is a transition. If the next bit is 1, there is none.

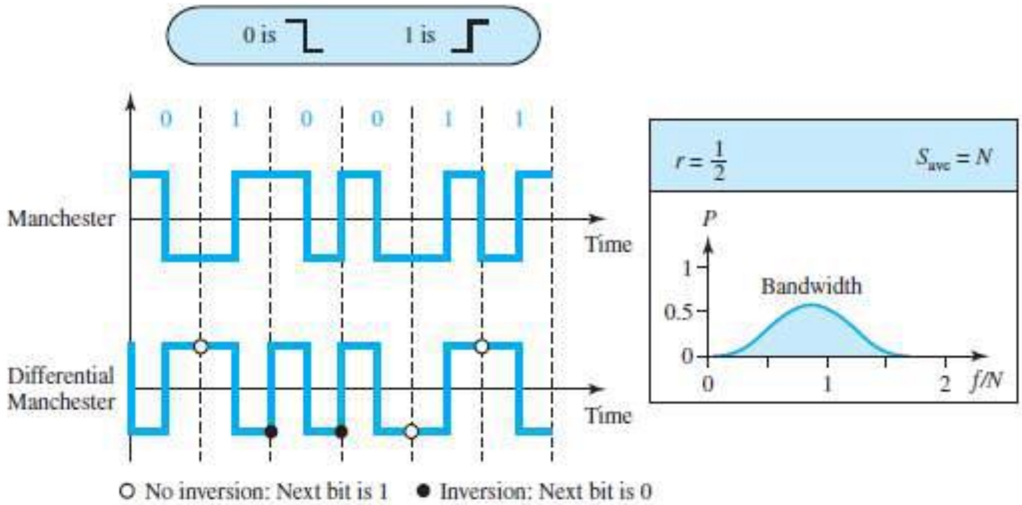


Figure 4.8 *Polar biphase: Manchester and differential Manchester schemes*

Advantages:

1) The Manchester scheme overcomes problems associated with NRZ-L.
Differential Manchester overcomes problems associated with NRZ-I.
2) There is no baseline wandering.
3) There is no DC component '.' each bit has a positive & negative voltage contribution.

Disadvantage:

1) Signal-rate: Signal-rate for Manchester & diff. Manchester is double that for NRZ.

Bipolar Schemes (or Multilevel Binary)

• This coding scheme uses 3 voltage levels (Figure 4.9):
  i) positive
  ii) negative &
  iii) zero.

• Two variations of bipolar encoding:
i) AMI (Alternate Mark Inversion)
ii) Pseudoternary

i) AMI

- Binary 0 is represented by a neutral 0 voltage (AMI Alternate 1 Inversion).
- Binary 1s are represented by alternating positive and negative voltages.

ii) Pseudoternary

- Binary 1 is represented by a neutral 0 voltage.
- Binary 0s are represented by alternating positive and negative voltages.

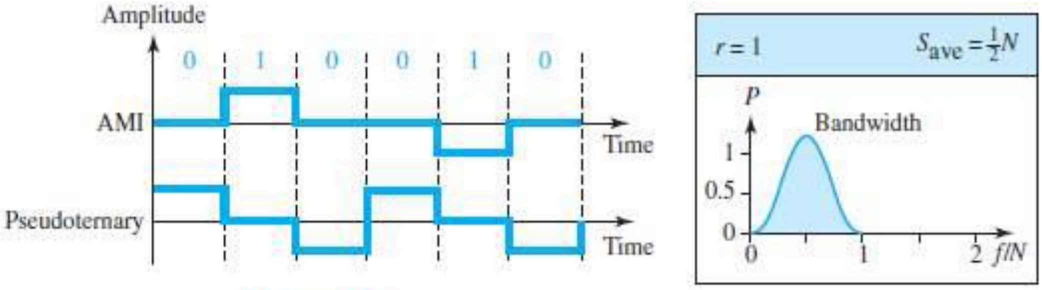


Figure 4.9 *Bipolar schemes: AMI and pseudoternary*

Advantages:
1) The bipolar scheme has the same signal-rate as NRZ.
2) There is no DC component '.' each bit has a positive & negative voltage contribution.
3) The concentration of the energy is around frequency N/2.
Disadvantage:
1) AMI has a synchronization problem when a long sequence of 0s is present in the data.

Table 4.1 *Summary of line coding schemes*

Category	Scheme	Bandwidth (average)	Characteristics
Unipolar	NRZ	$B = N/2$	Costly, no self-synchronization if long 0s or 1s, DC
Polar	NRZ-L	$B = N/2$	No self-synchronization if long 0s or 1s, DC
	NRZ-I	$B = N/2$	No self-synchronization for long 0s, DC
	Biphase	$B = N$	Self-synchronization, no DC, high bandwidth
Bipolar	AMI	$B = N/2$	No self-synchronization for long 0s, DC

Canara Engineering College