Chapter 1

Introduction

1-1 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:

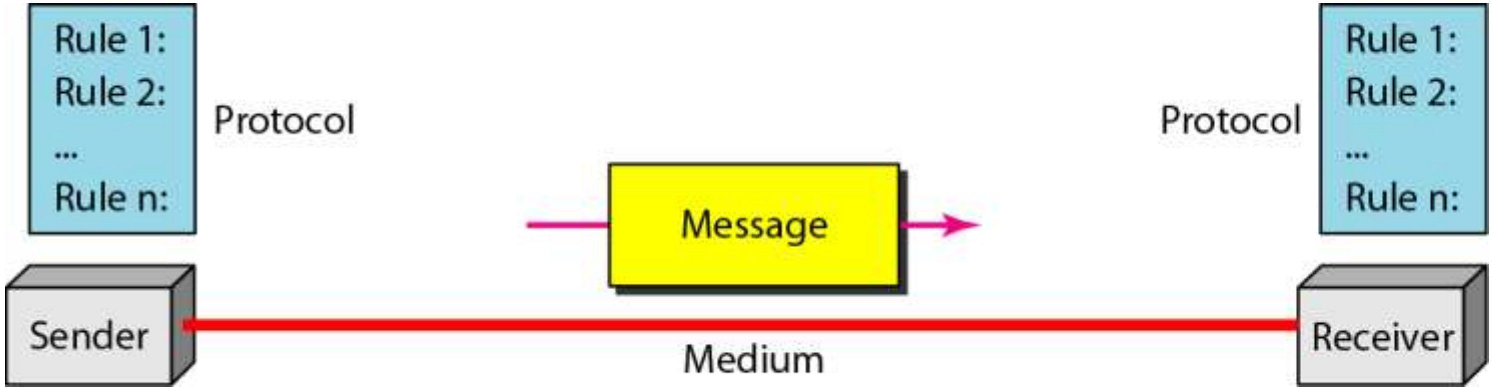
1) Delivery : The system must deliver data to the correct destination.

2) Accuracy : The system must deliver data to the correct destination. 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.

Figure 1.1 *Components of a data communication system*



Components of a data communication system

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.
   - 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.

  ¤ 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

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.

- The monitor can only accept output. Entire-capacity of channel is used to send the data in one direction

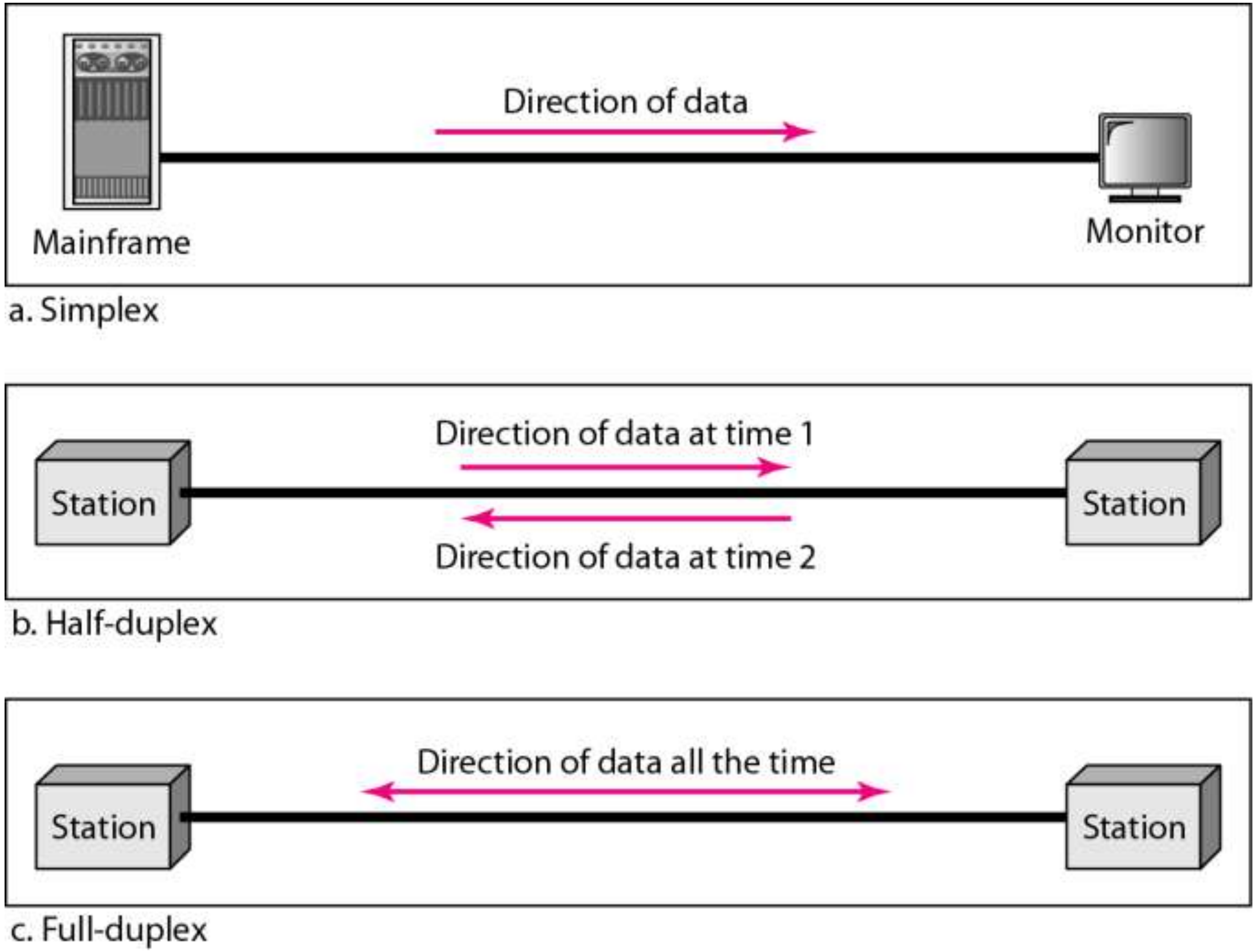
2) Half Duplex :

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

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



1-2 NETWORKS

A network is a set of devices (often referred to as nodes) connected by communication links. A node can be a computer, printer, or any other device capable of sending and/or receiving data generated by other nodes on the network. A link can be a cable, air, optical fiber, or any medium which can transport a signal carrying information.

Topics discussed in this section:

- **Network Criteria**
- **Physical Structures**
- **Categories of Networks**

Network Criteria

- **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:
    - Number of users
    - Type of transmission-medium
    - Efficiency of software Often, performance is evaluated by 2 networking-metrics: i) **throughput and ii) delay**. Good performance can be obtained by achieving higher throughput and smaller delay

Network Criteria

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

Two types of connections

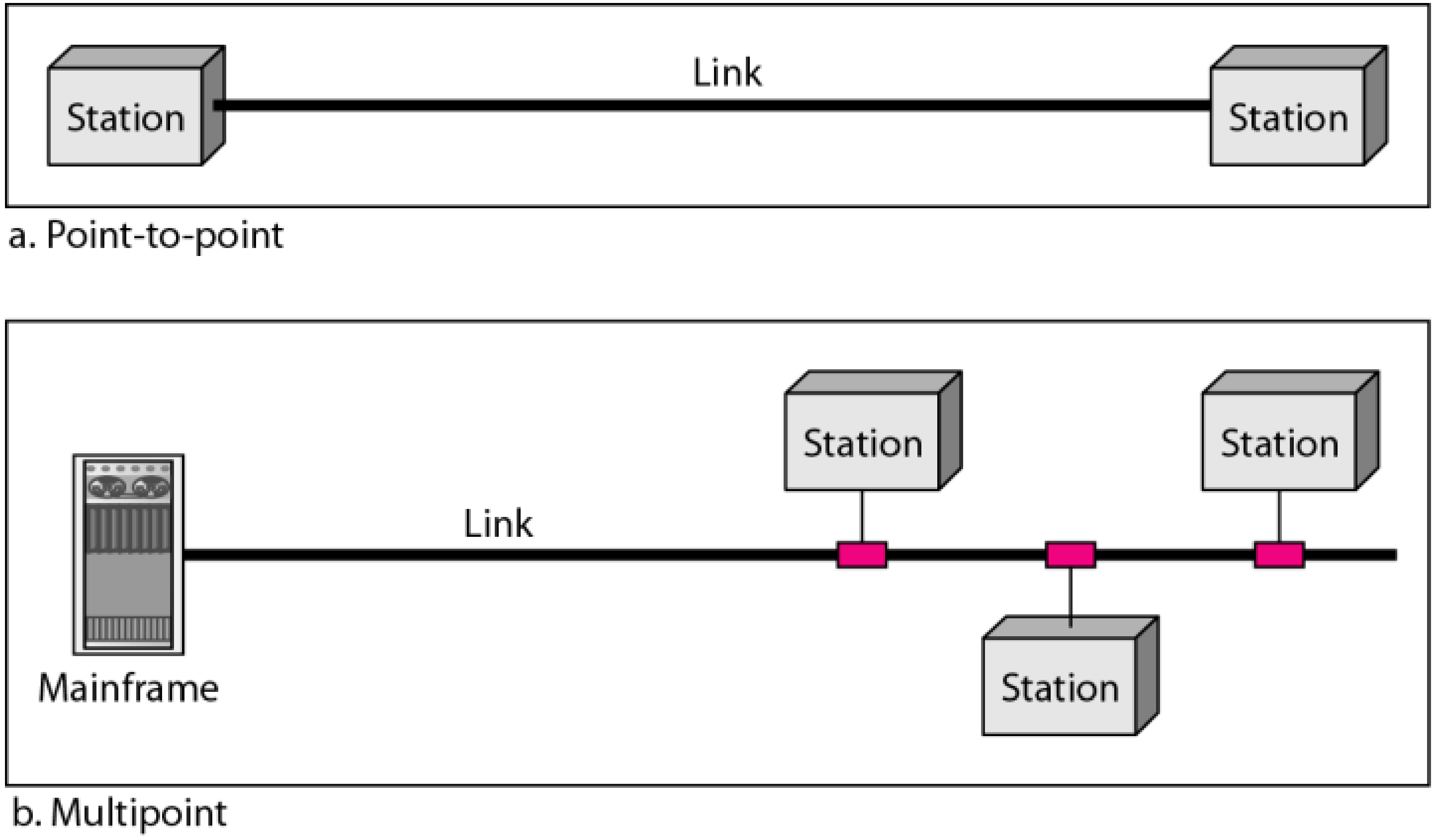
1) Point-to-Point

- Only two devices are connected by a dedicated-link .
- Entire-capacity of the link is reserved for transmission between those two devices.
- For example: Point-to-Point connection b/w remote-control & TV for changing the channels.

2) Multipoint (Multi-Drop)

- **Three or more devices sha**re a single link.
- The capacity of the channel is shared, either spatially or temporally .
- If link is used simultaneously by many devices, then it is spatially shared connection.
- 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*

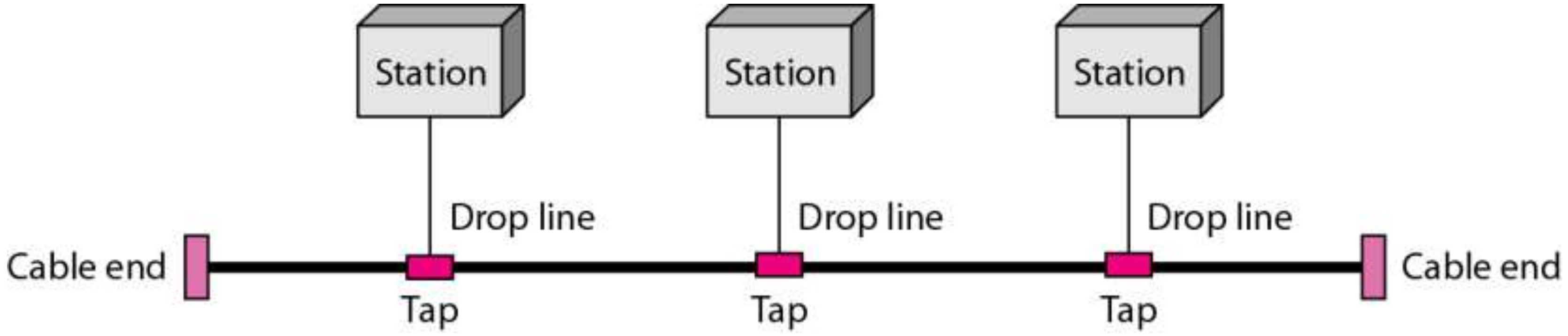


Physical Topology

Bus Topology

- All the devices are connected to the single cable called bus .
- 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.
- Devices are connected to the bus by drop-lines and taps.
- A drop-line is a connection running between the device and the bus.
- As a signal travels along the backbone, some of its energy is transformed into heat.
- Therefore, it becomes weaker and weaker as it travels farther and farther.
- For this reason there is a limit on the number of taps a bus can support and on the distance between those taps.

Figure 1.7 *A bus topology connecting three stations*



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.

Disadvantages:

1) Difficult to detect and troubleshoot fault.

2) A fault/break in the cable stops all transmission.

3) There is a limit on i) Cable length ii) Number of nodes that can be connected.

4) 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** .
- 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.

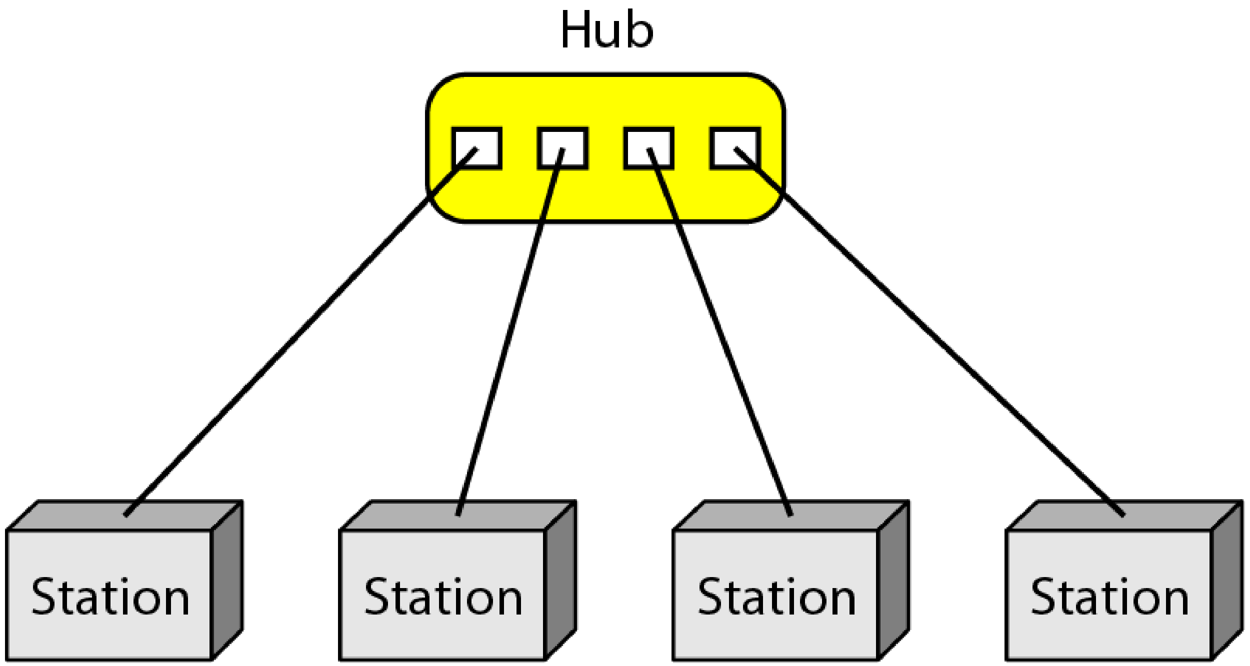
Advantages:

- 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:

- Single point of failure: If the hub goes down, the whole network is dead.
- Cable length required is the more compared to bus/ring topologies.
- Number of nodes in network depends on capacity of hub.

Figure 1.6 *A star topology connecting four stations*



Ring Topology

- Each device is connected to the next, forming a ring
- 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.
- Easy installation and reconfiguration. :To add/delete a device, requires changing only 2 connections.

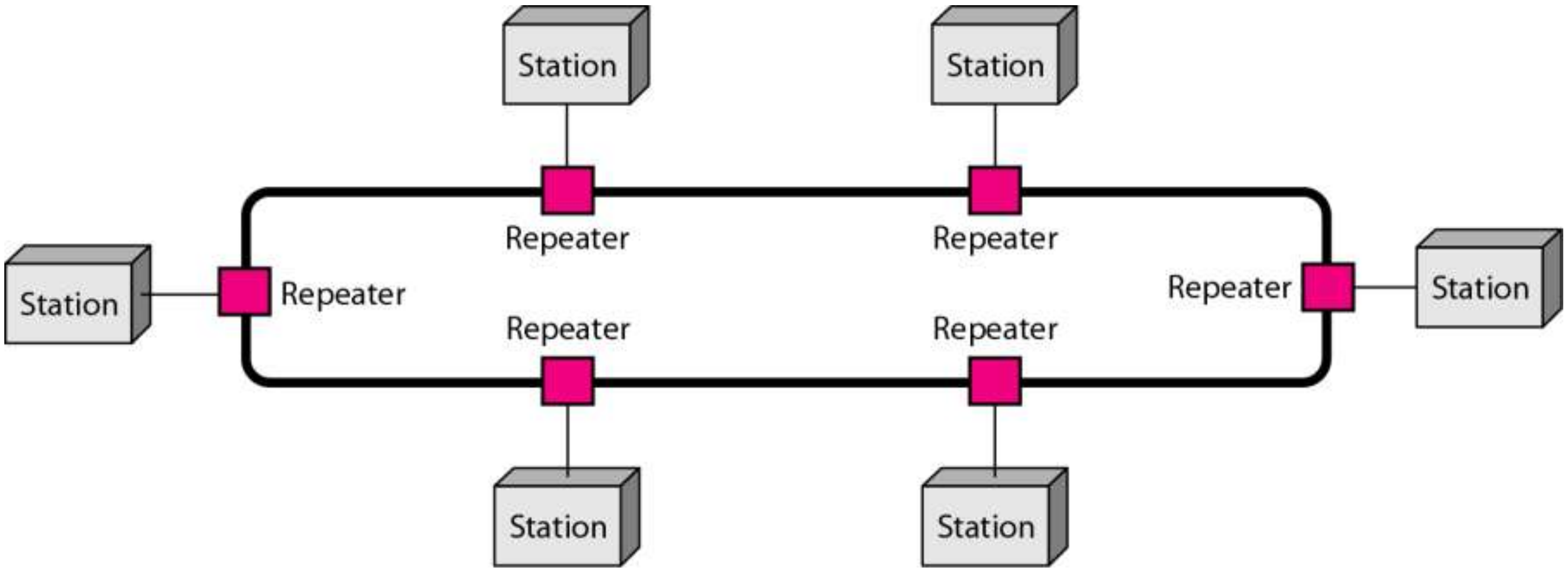
Advantages

- 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.
- Congestion reduced: Because all the traffic flows in only one direction

Disadvantages:

- Unidirectional traffic.
- A fault in the ring/device stops all transmission.
- There is a limit on
  - Cable length &
  - Number of nodes that can be connected.
- Slower: Each data must pass through all the devices between source and destination.

Figure 1.8 *A ring topology connecting six stations*



Mesh Topology

- All the devices are connected to each other .
- 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.

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.

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

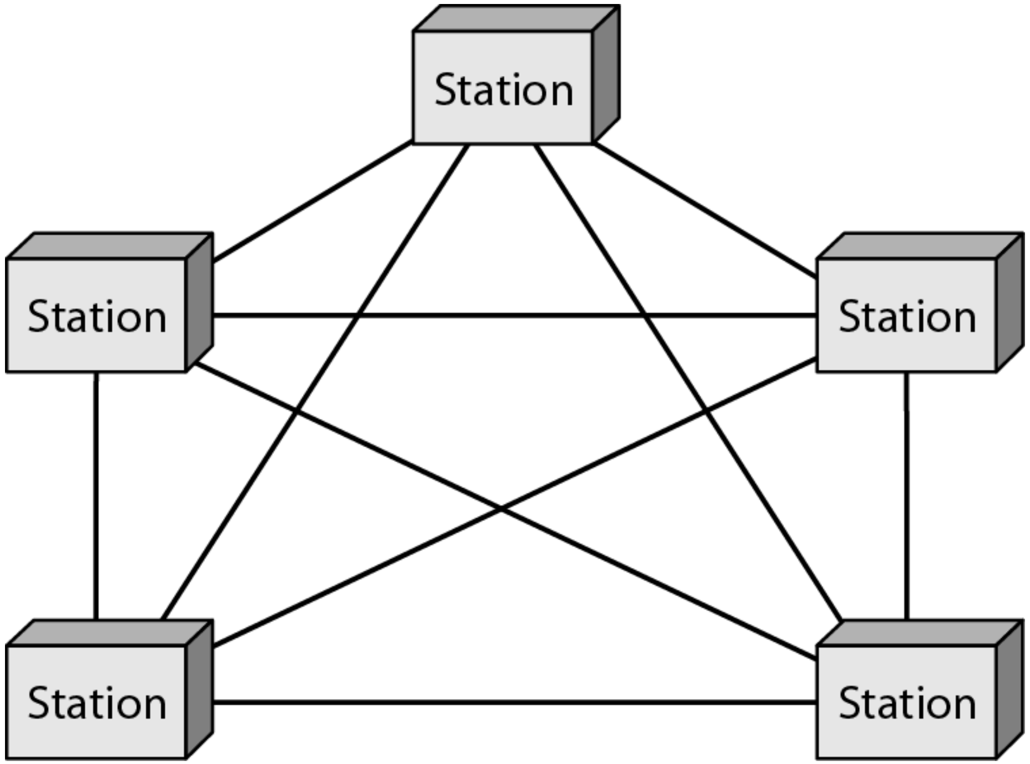
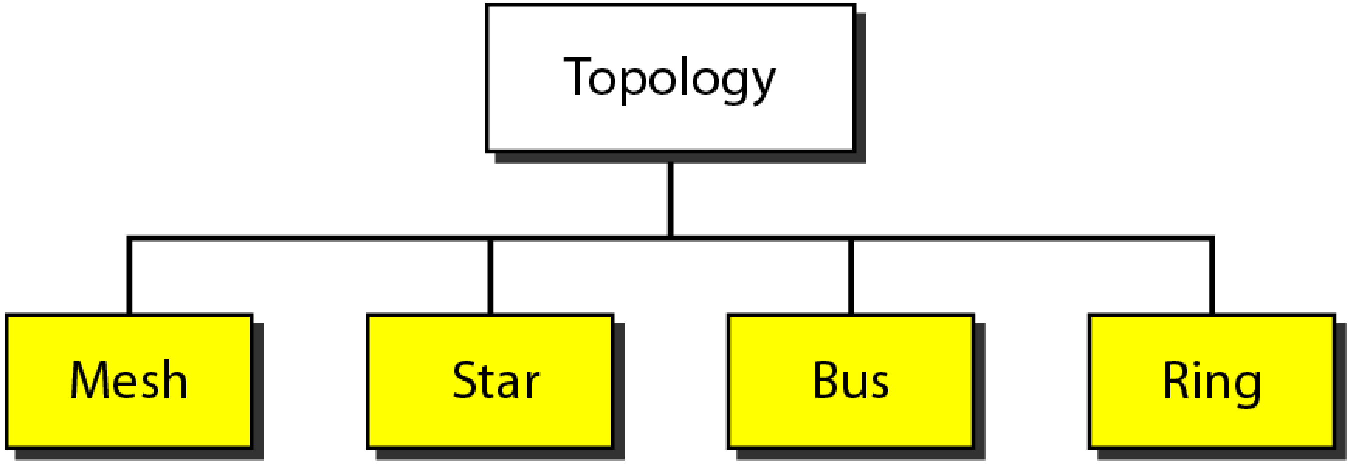


Figure 1.4 *Categories of topology*



Network Types

LAN

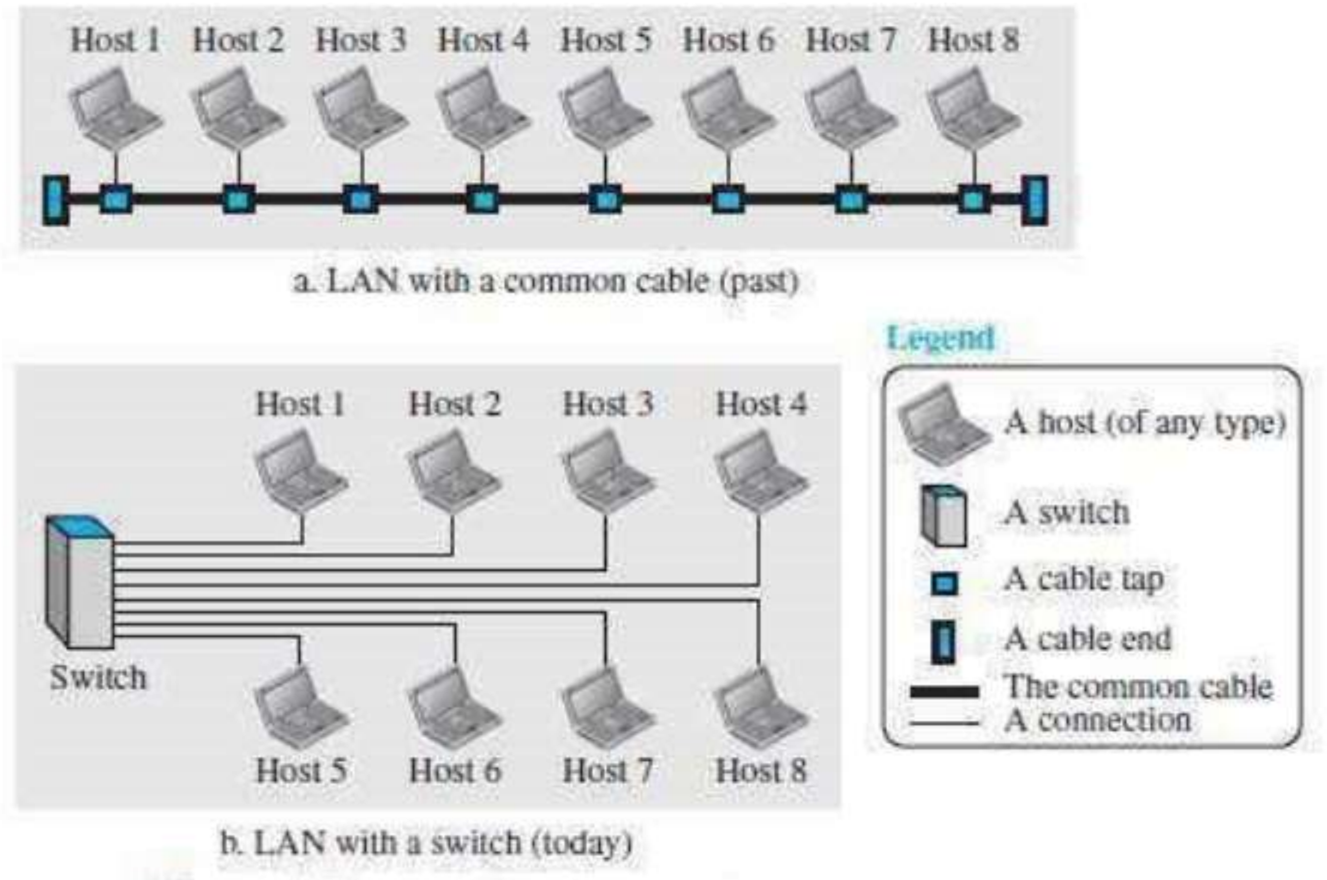
- LAN is used to connect computers in a single office, building or campus
- LAN is usually privately owned network.
- A LAN can be simple or complex.
- Simple: LAN may contain 2 PCs and a printer.
- Complex: LAN can extend throughout a company.
- Each host in a LAN has an address that uniquely defines 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
  - recognize the destination address of the packet &
  - guide the packet to its destination.
- The switch
  - reduces the traffic in the LAN &
  - allows more than one pair to communicate with each other at the same time.

Advantages

1) Resource Sharing

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


Host 1
Host 2
Host 3
Host 4
Host 5
Host 6
Host 7
Host 8
a. LAN with a common cable (past)
Legend
A host (of any type)
A switch
A cable tap
A cable end
The common cable
A connection
Switch
b. LAN with a switch (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.
- WAN is created & run by communication companies (Ex: BSNL, Airtel) → leased by an organization that uses it.

A WAN can be of 2 types:

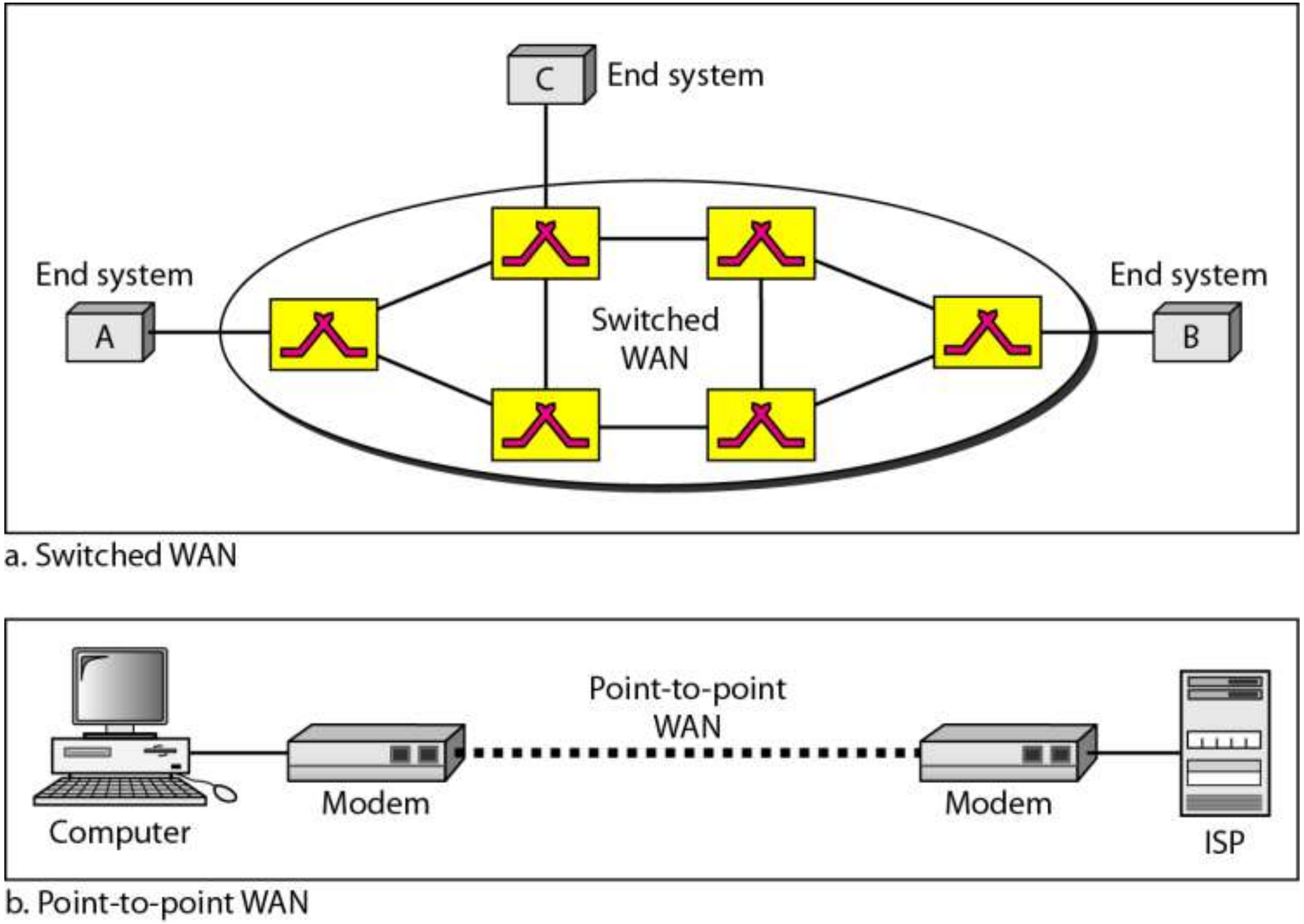
- **Point-to-Point WAN**

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

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

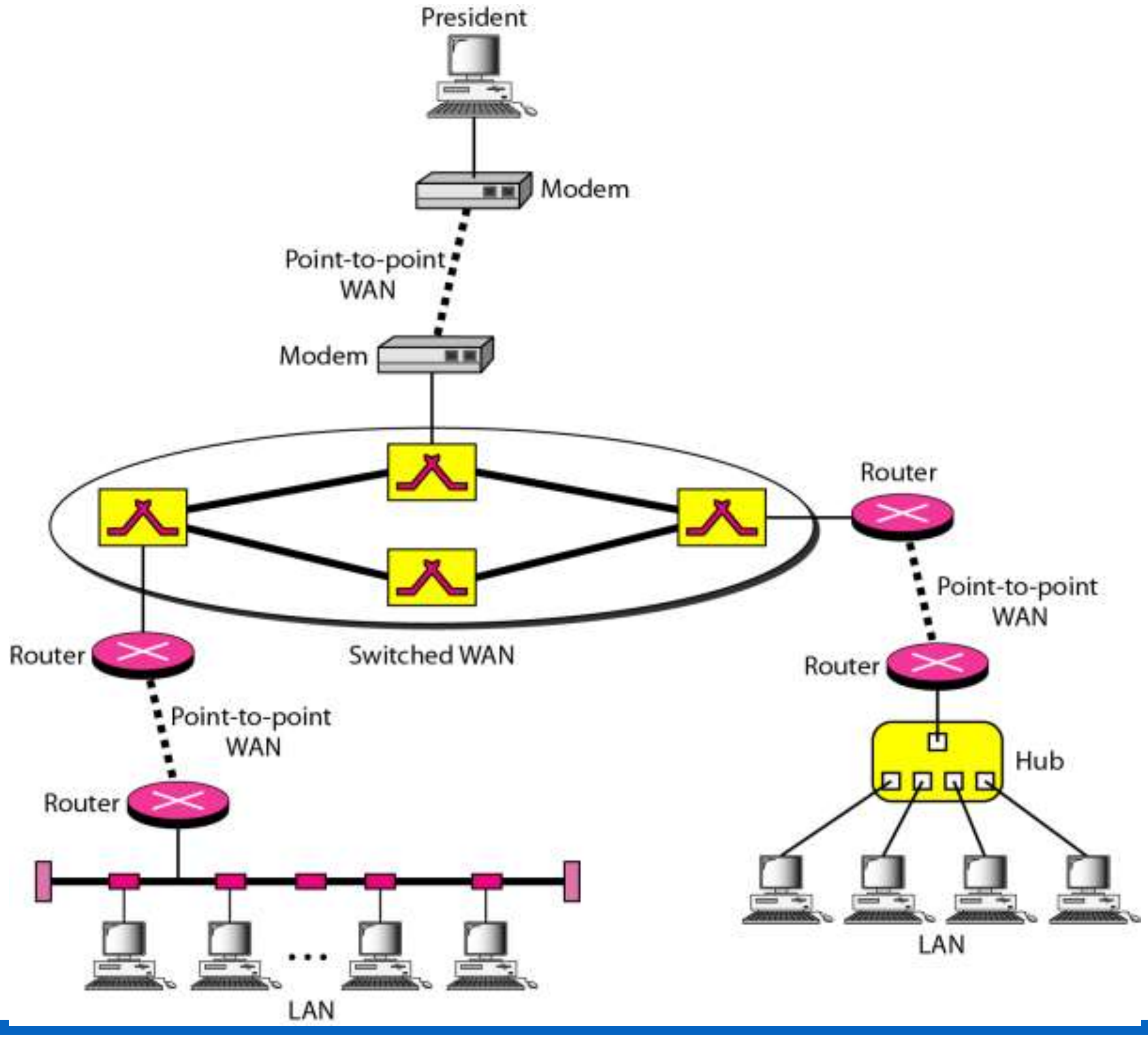


Internetwork

- A network of networks is called an internet. (Internet inter-network)
- For example (Figure 1.11): Assume that an organization has two offices,
  - First office is on the east coast &
  - 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)
- 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 1.11 *An internetwork made of two LANs and one point-to-point WAN*

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



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.
- As shown in Figure, 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.

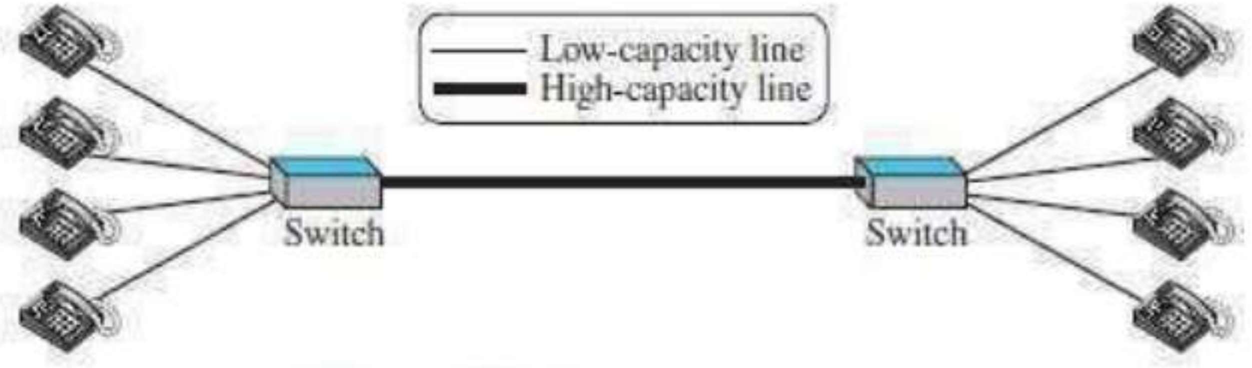


Figure 1.13 *A circuit-switched network*

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

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.
- As shown in Figure 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.

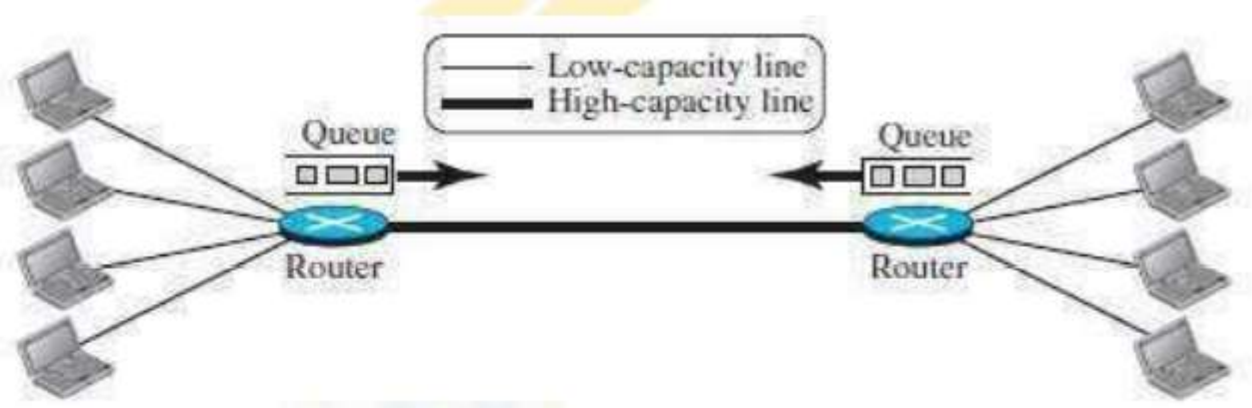


Figure 1.14 *A packet-switched network*

The Internet Today

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

Internet is made up of

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.

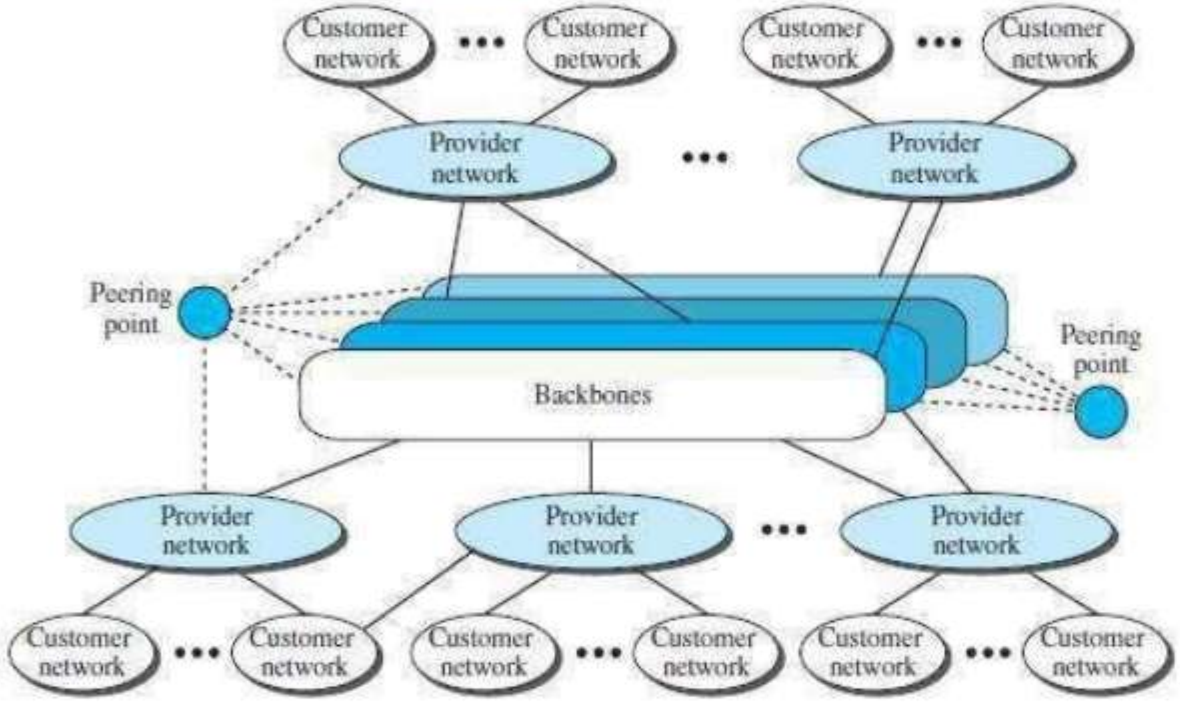


Figure 1.15 *The Internet today*

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

    A) Dial-up service

    B)DSL Service

2) Using Cable Networks

3)Using Wireless Networks

4) Direct Connection to the Internet

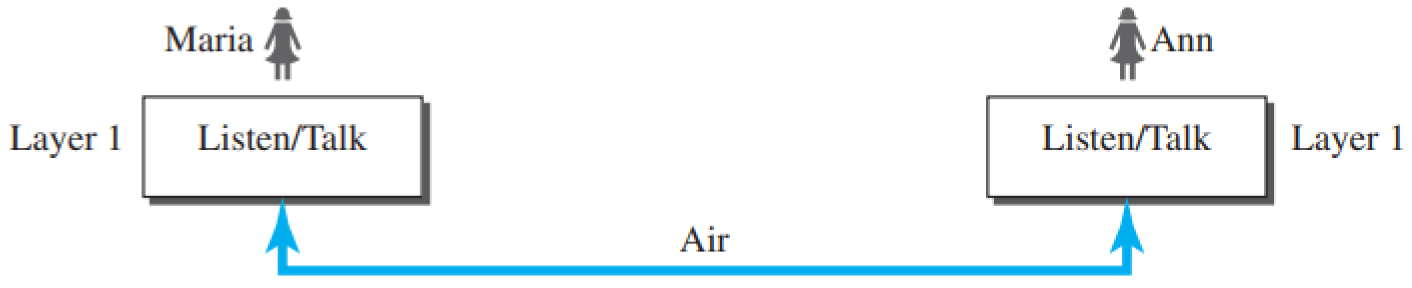
PROTOCOL LAYERING

- 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.

First Scenario

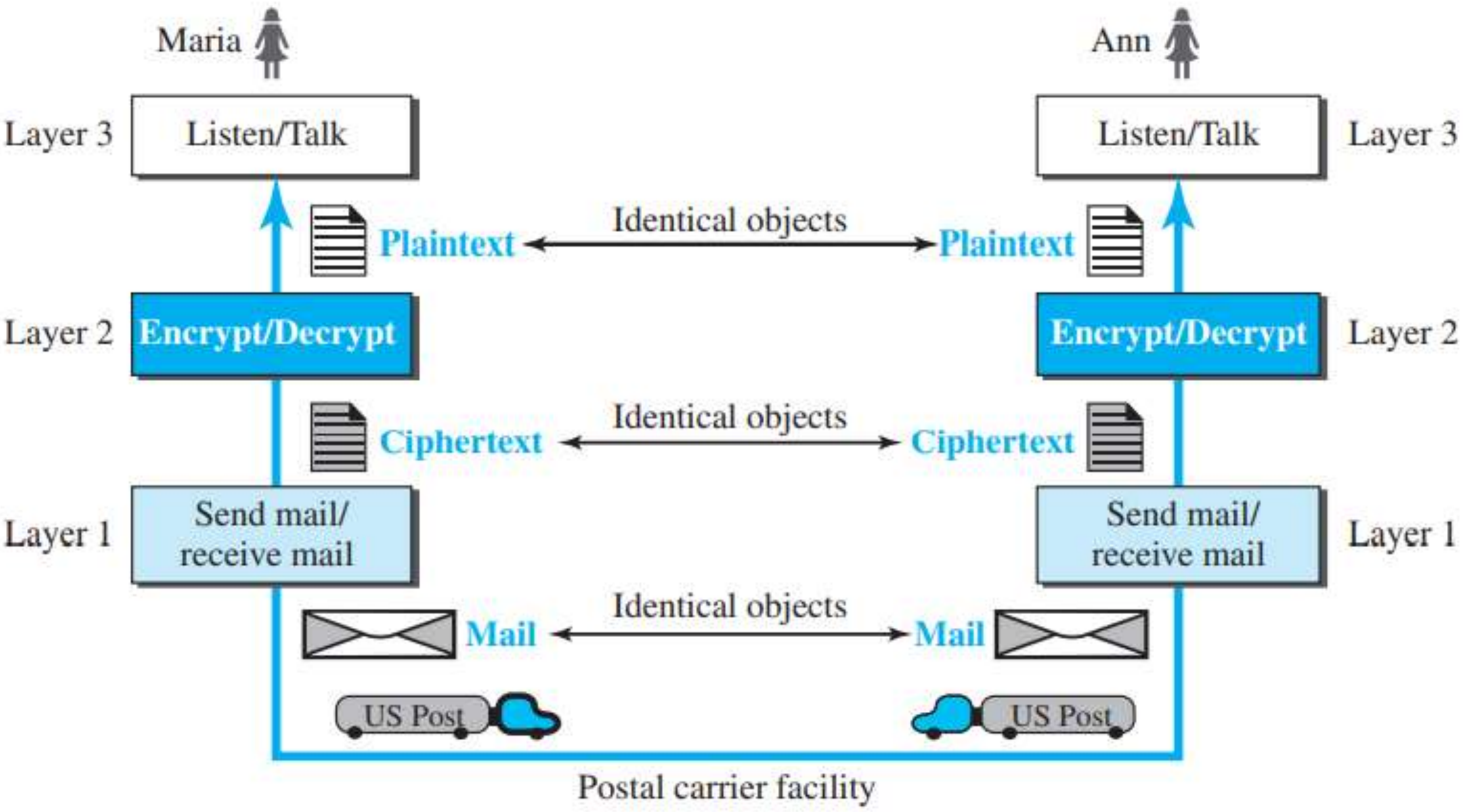
- In the first scenario, communication is so simple that it can occur in only one layer (Figure 2.1).
- Assume Maria and Ann are neighbors with a lot of common ideas.
- Communication between Maria and Ann takes place in one layer, face to face, in the same language

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.



Maria
Ann
Layer 3
Listen/Talk
Listen/Talk
Layer 3
Plaintext
Identical objects
Plaintext
Layer 2
Encrypt/Decrypt
Encrypt/Decrypt
Layer 2
Ciphertext
Identical objects
Ciphertext
Layer 1
Send mail/
receive mail
Send mail/
receive mail
Layer 1
Mail
Identical objects
Mail
US Post
US Post
Postal carrier facility

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.

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).

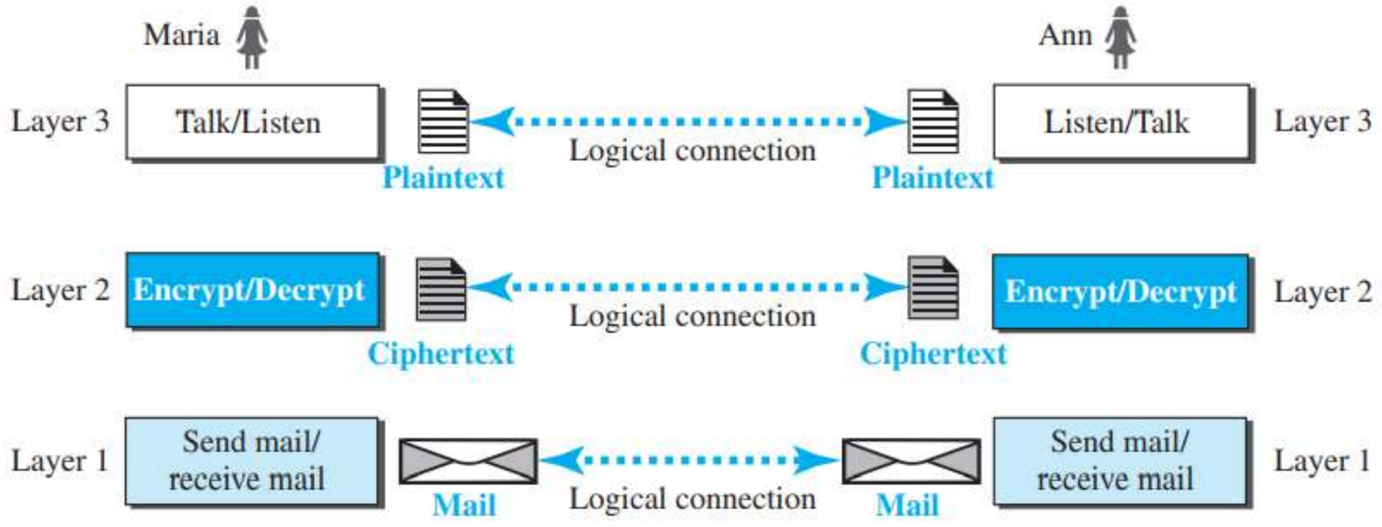
2) ***Second Principle***

- 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.

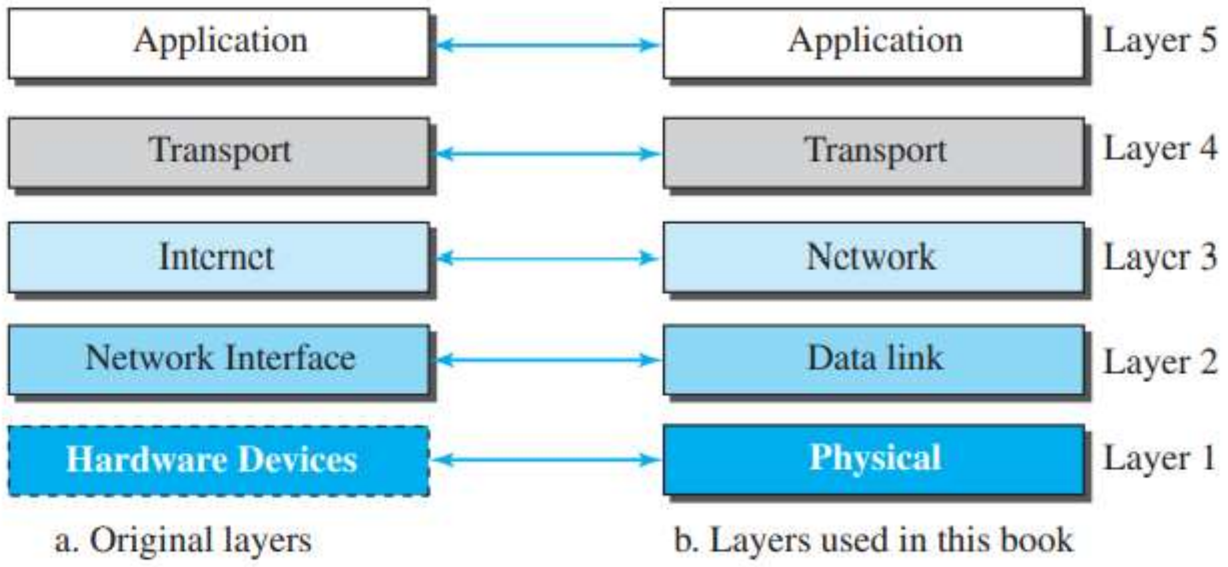
- There is a logical connection at each layer through which 2 end systems can send the object created from that layer.

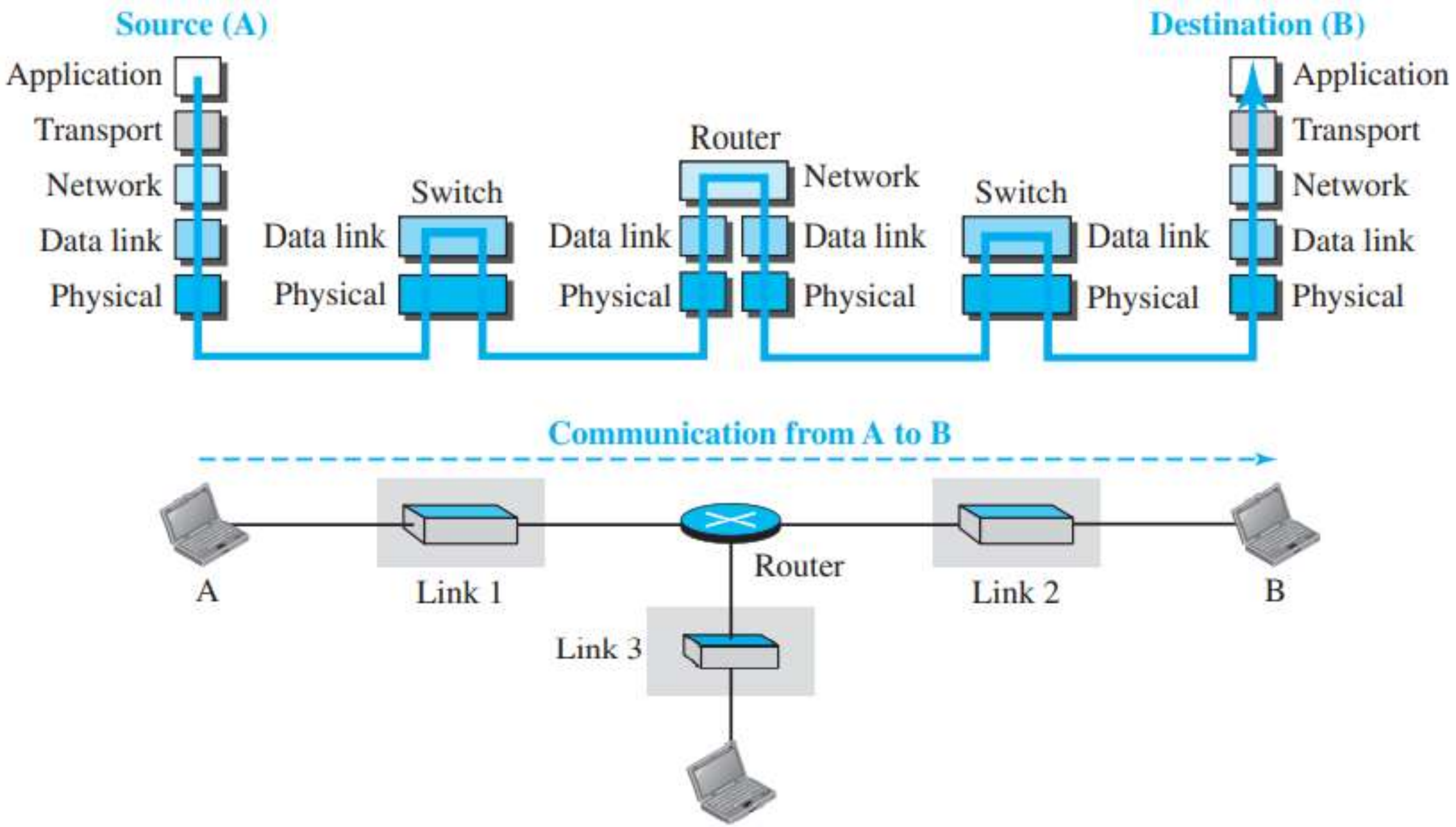

Maria
Ann
Layer 3
Talk/Listen
Plaintext
Logical connection
Plaintext
Listen/Talk
Layer 3
Layer 2
Encrypt/Decrypt
Ciphertext
Logical connection
Ciphertext
Encrypt/Decrypt
Layer 2
Layer 1
Send mail/
receive mail
Mail
Logical connection
Mail
Send mail/
receive mail
Layer 1

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*




Source (A)
Application
Transport
Network
Data link
Physical
Switch
Data link
Physical
Router
Data link
Physical
Network
Data link
Physical
Switch
Data link
Physical
Destination (B)
Application
Transport
Network
Data link
Physical
Communication from A to B
A
Link 1
Router
Link 2
B
Link 3

Layered Architecture

- Let us assume that computer A communicates with computer B ,
- As the Figure 2.5 shows, we have five communicating devices:
- 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

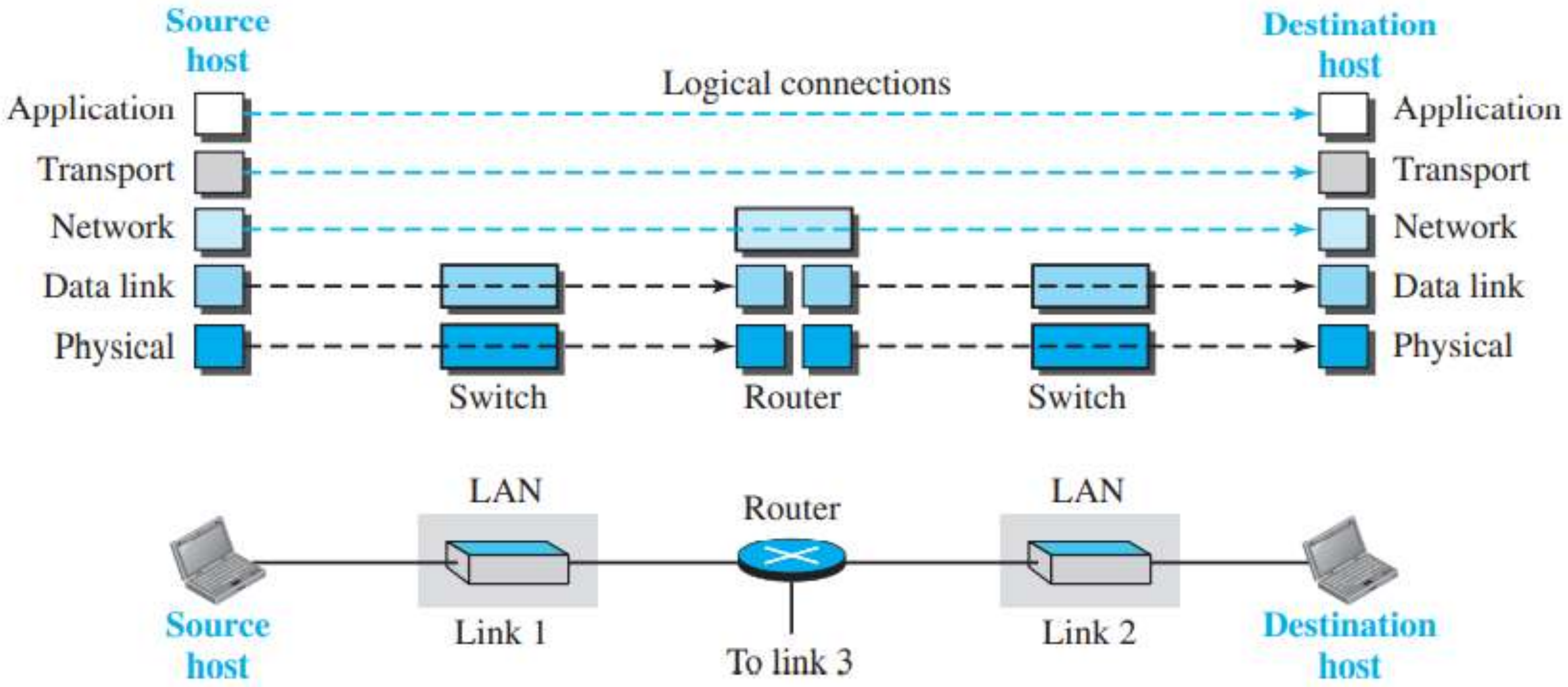
→ creates a message in the application layer and

1) → sends the message down the layers so that it is physically sent to the destination host.

- The destination host

1) → receives the message at the physical layer and
2) → 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.



Source host
Destination host
Logical connections
Application
Transport
Network
Data link
Physical
Application
Transport
Network
Data link
Physical
Switch
Router
Switch
LAN
Router
LAN
Link 1
To link 3
Link 2
Source host
Destination host

Layers in the TCP/IP Protocol Suite

- 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.
- Identical objects exist between two hops.
- Because router may fragment the packet at the network layer and send more packets than received .
- 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 frame from the data-link layer & → sends 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.
- Each protocol may provide a different service.
- Some protocols provide complete error detection and correction; some protocols provide only error correction

The 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)

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 connectionless & unreliable protocol.
- i) Connection-less means : there is no connection setup b/w the sender and the receiver.
- ii) Unreliable protocol means
  - 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.

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) 2) UDP (User Datagram Protocol) &
3) 3) SCTP (Stream Control Transmission Protocol

TCP

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

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

SCTP

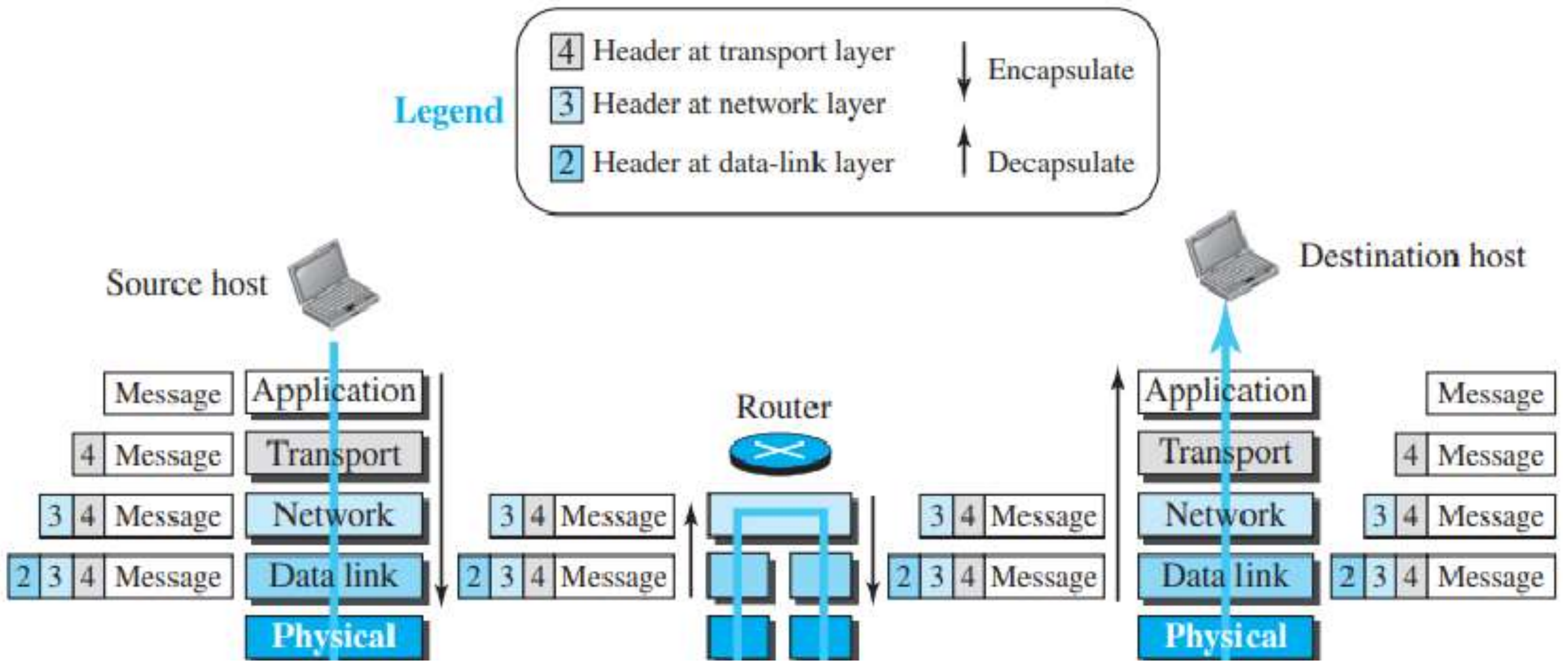
- SCTP provides support for newer applications such as voice over the Internet. It combines the best features of UDP and TCP.

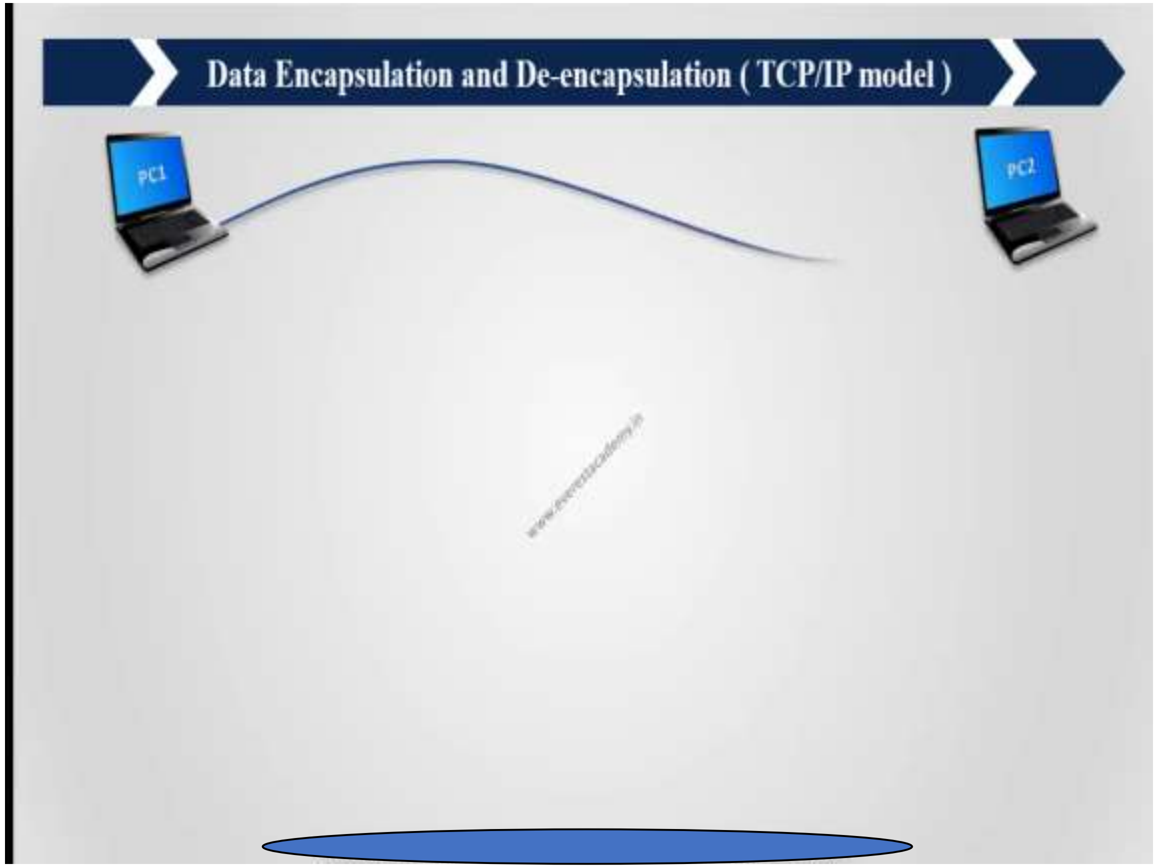
Application Layer

- 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 is used to transport email between a source and destination.
- 2) TELNET is used for accessing a site remotely.
- 3) FTP is used for transferring files from one host to another.
- 4) DNS is used to find the IP address of a computer.
- 5) SNMP is used to manage the Internet at global and local levels.
- 6) HTTP is used for accessing the World Wide Web (WWW).

(FTP File Transfer Protocol)(SMTP Simple Mail Transfer Protocol) (DNS Domain Name System HTTP Hyper Text Transfer Protocol) (SNMP Simple Network Management Protocol TELNET Terminal Network)

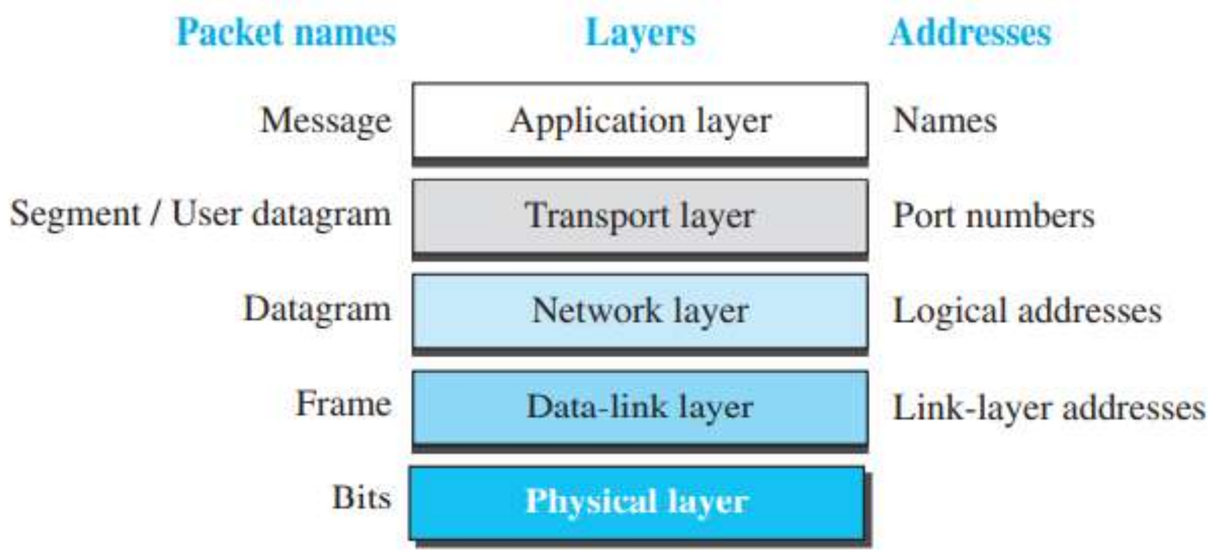
Encapsulation and Decapsulation




Data Encapsulation and De-encapsulation (TCP/IP model)
PC1
PC2
www.everestacademy.in

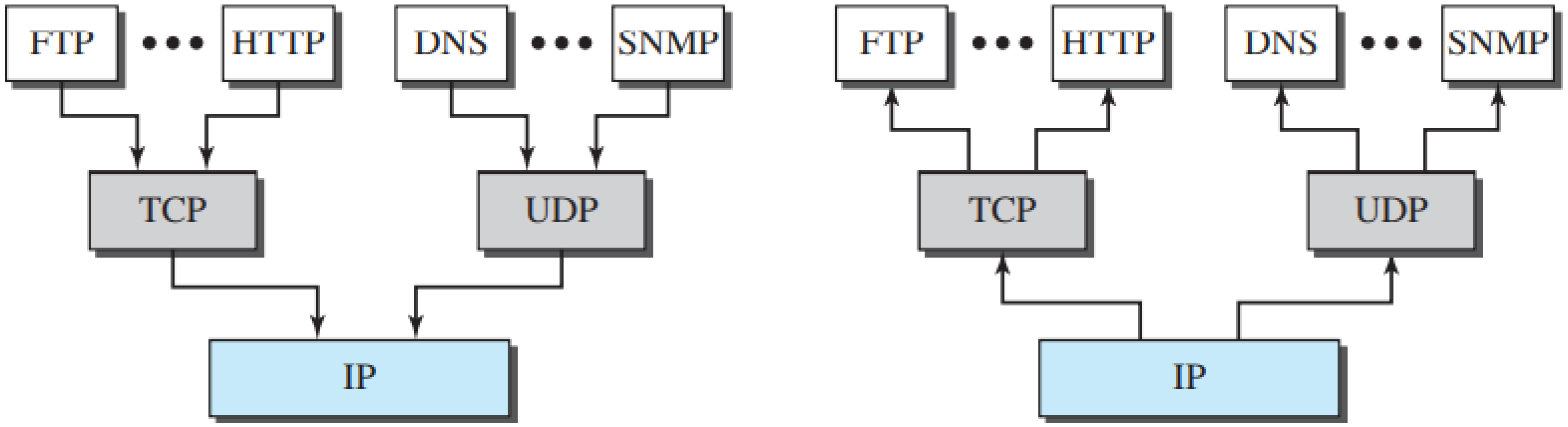
Addressing

Figure 2.9 Addressing in the TCP/IP protocol suite



Multiplexing and Demultiplexing

Figure 2.10 *Multiplexing and demultiplexing*

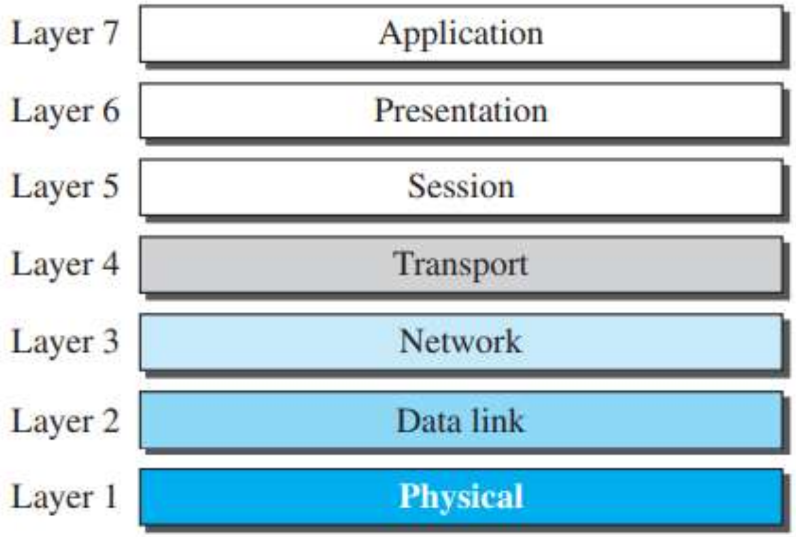


a. Multiplexing at source

b. Demultiplexing at destination

THE OSI MODEL
(Open Systems Interconnection)

Figure 2.11 The OSI model



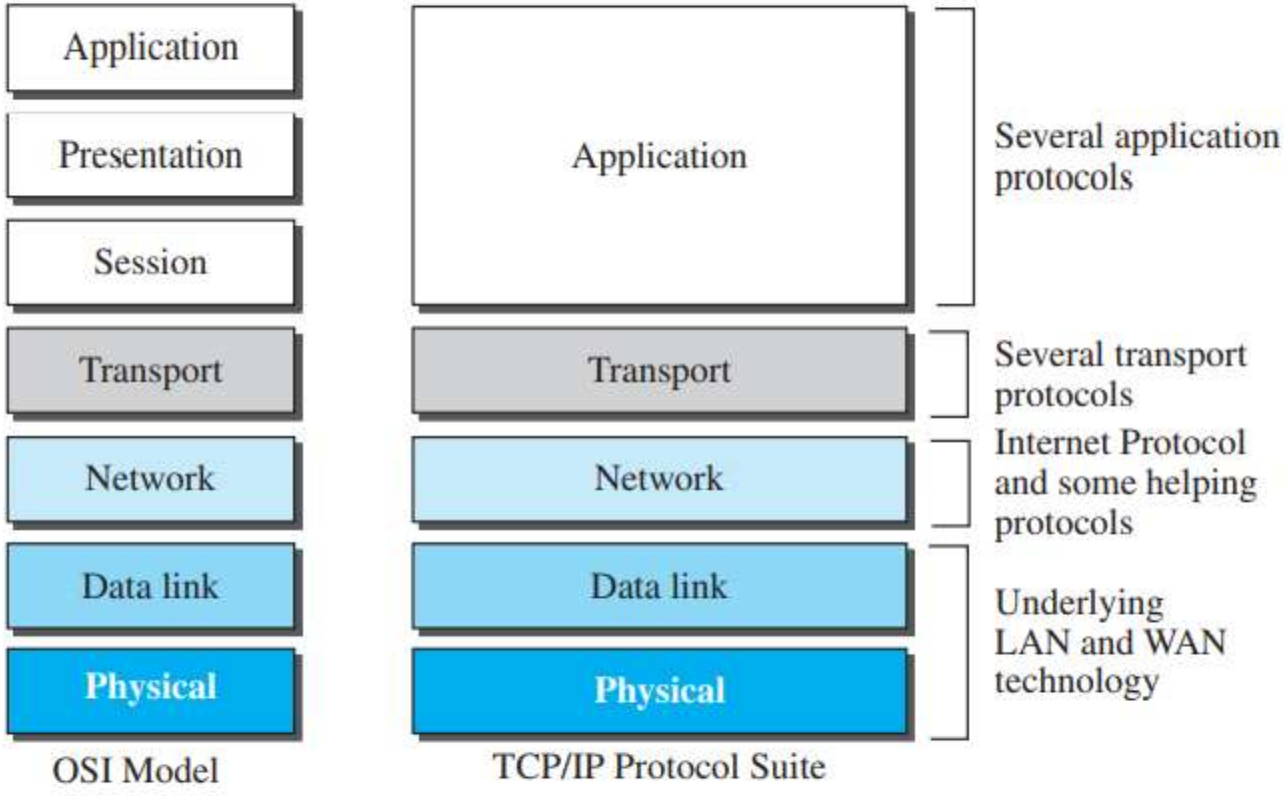
The session layer

- The session layer (layer 5) is responsible for establishing, managing, synchronizing and terminating sessions between end-user application processes.

Presentation layer

- Presentation layer format(**ASCII → EBCDIC**) and encrypts data to be sent across the network.
- This layer takes care that the data is sent in such a way that the receiver will understand the information (data) and will be able to use the data efficiently and effectively.
- Some examples of presentation layer protocols are **SSL, HTTP/ HTML (agent), FTP (server),**

OSI versus TCP/IP



Introduction to Physical Layer

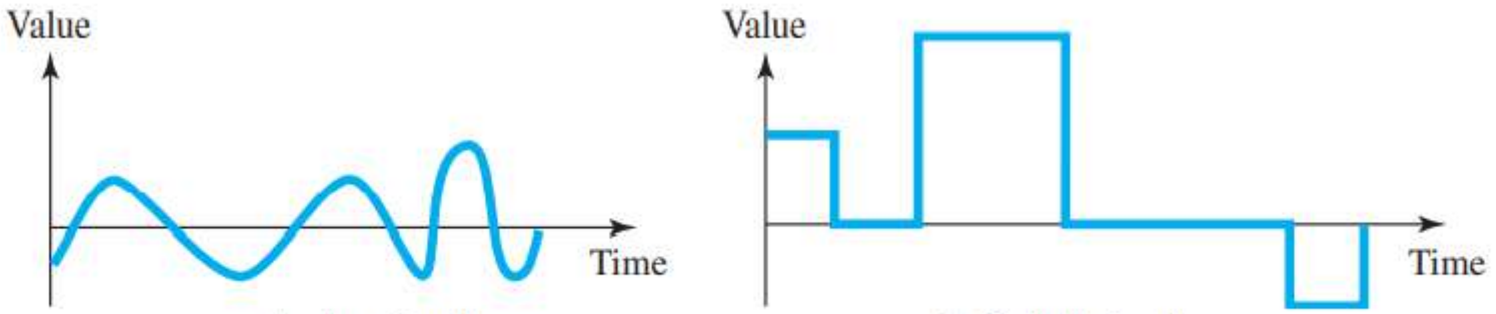
DATA AND SIGNALS

DATA AND SIGNALS

Analog & Digital Data

- To be transmitted, data must be transformed to electromagnetic-signals.
- Data can be either analog or digital.
- Analog Data refers to information that is continuous. For example: The sounds made by a human voice.
- Digital Data refers to information that has discrete states. For example: Data are stored in computer-memory in the form of 0s and 1s.

Figure 3.2 *Comparison of analog and digital signals*



a. Analog signal

b. Digital signal

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.

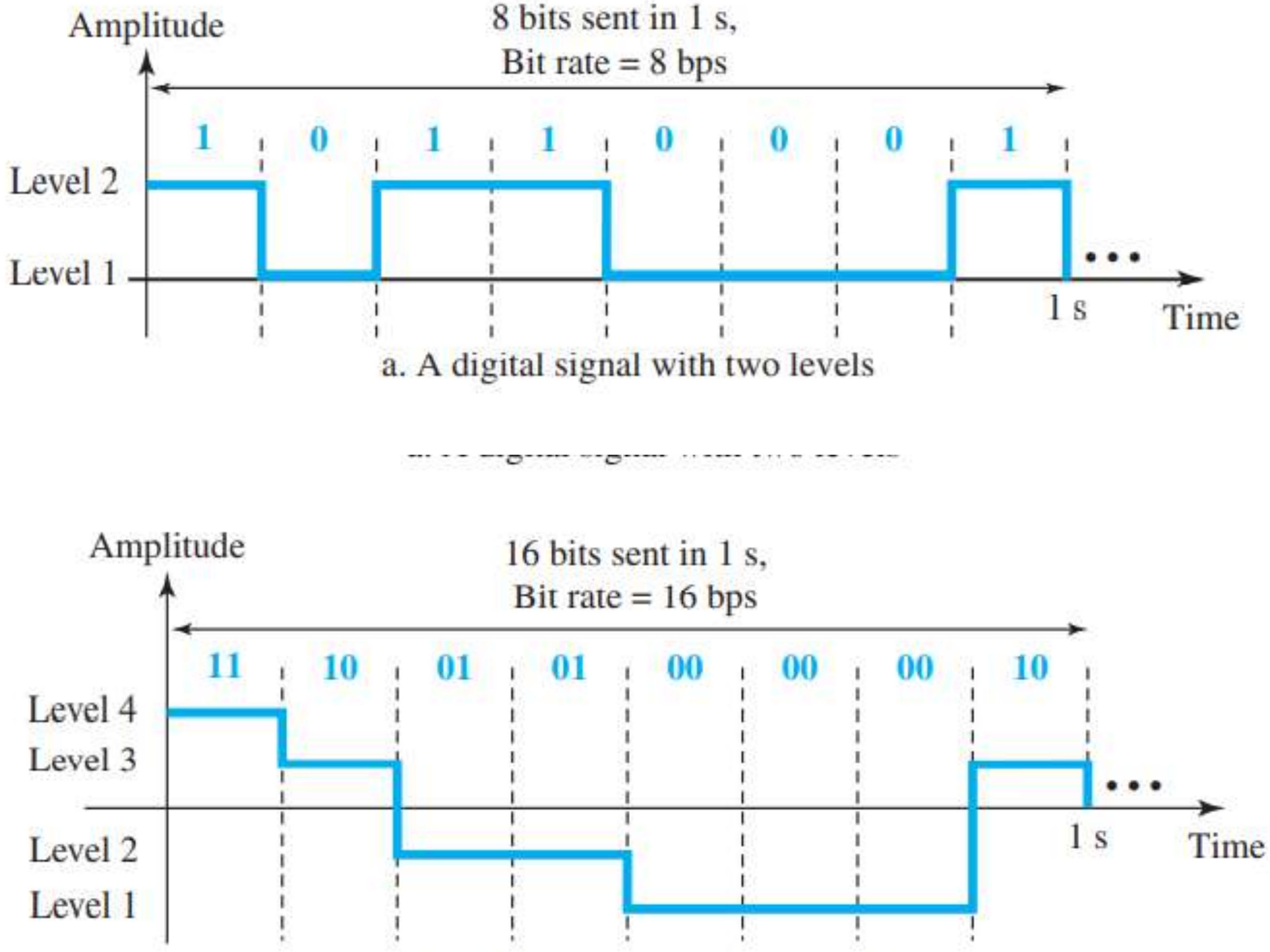
2) Non-Periodic Signal Signals which do not repeat itself after a fixed time period are called Non-Periodic Signals.

In data communications, we commonly use periodic analog signals and nonperiodic digital 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).

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



a. A digital signal with two levels

b. A digital signal with four levels

Example 3.16

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).

Example 3.18

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

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

Example 3.19

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

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

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);

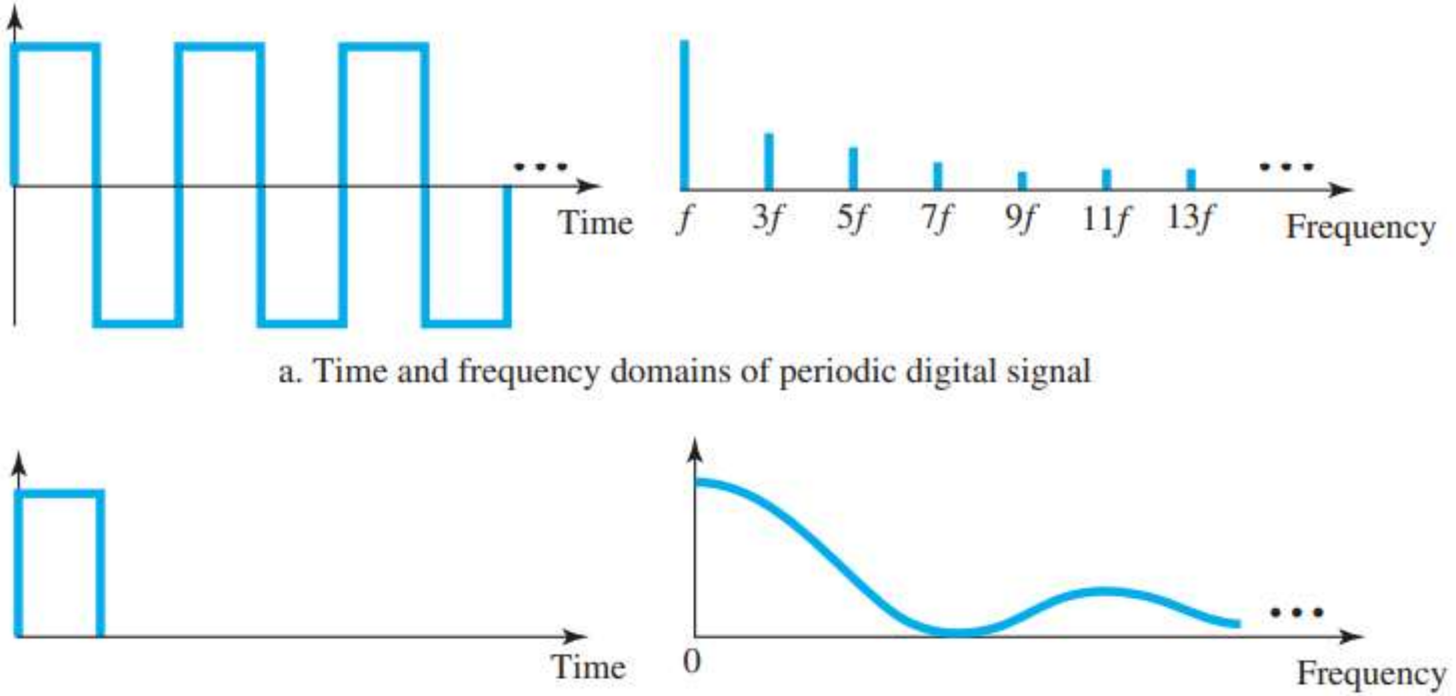
2) A horizontal line in the time domain means a frequency of zero (no change in time).

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*



a. Time and frequency domains of periodic digital signal

b. Time and frequency domains of nonperiodic digital signal

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).
- 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 3.19 *Baseband transmission*

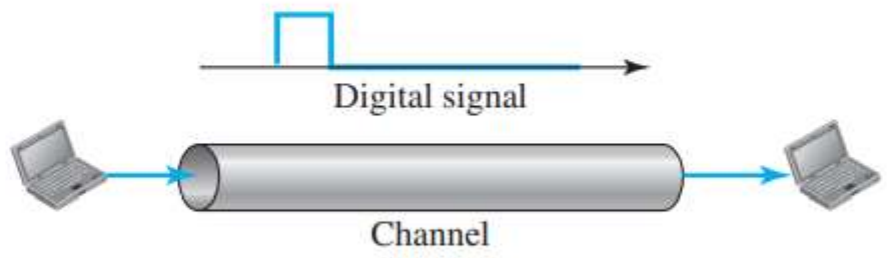
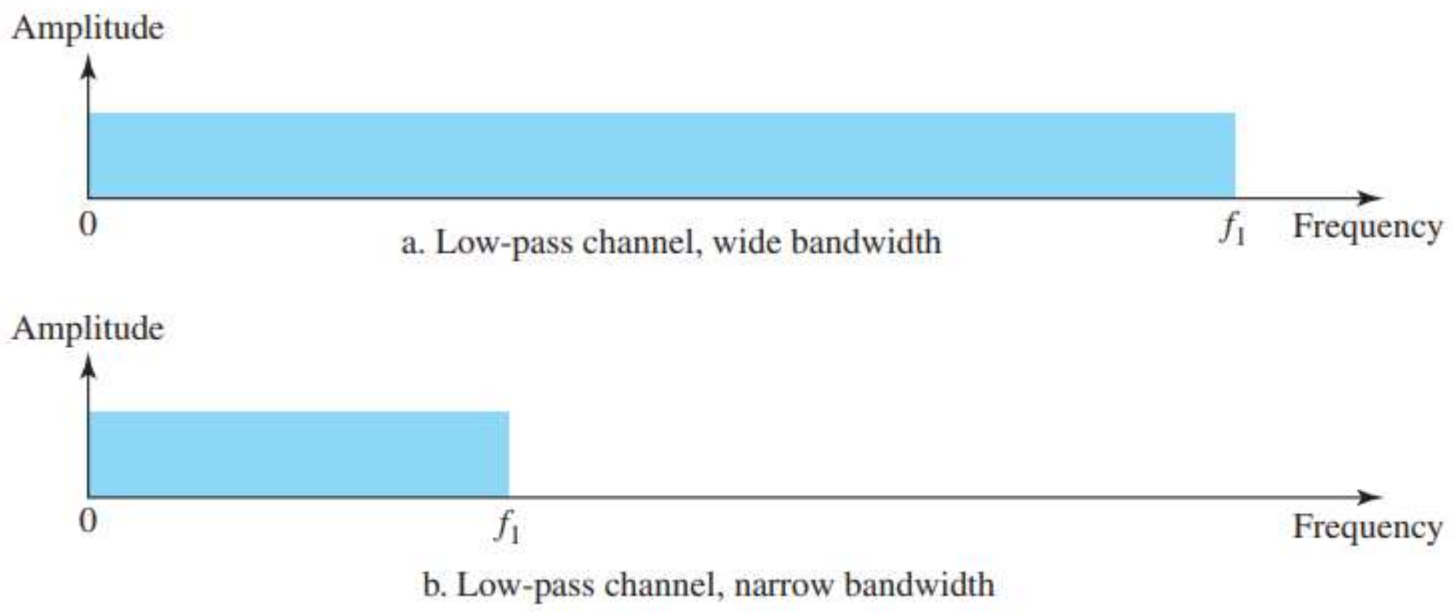


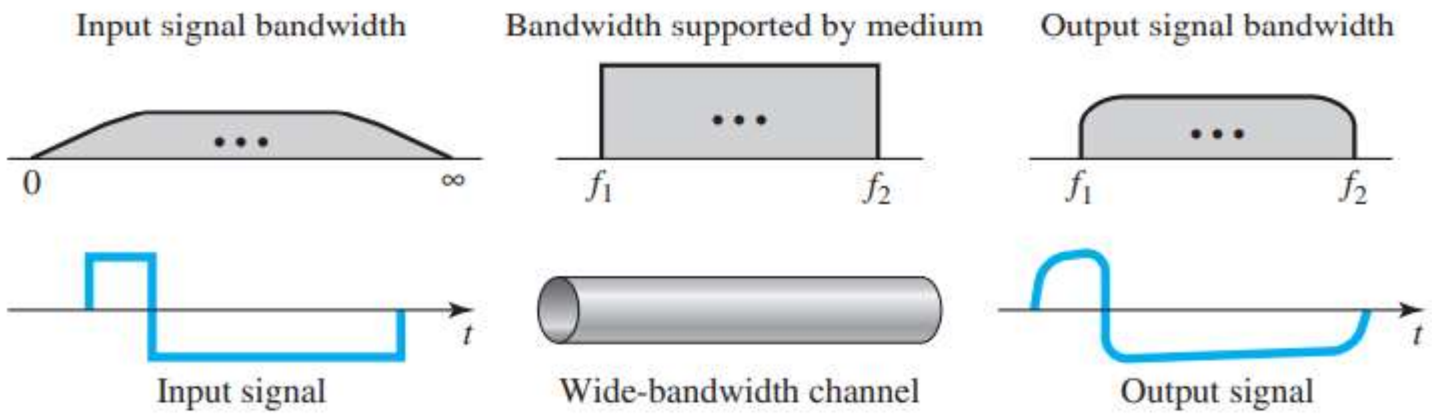
Figure 3.20 *Bandwidths of two low-pass channels*



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*



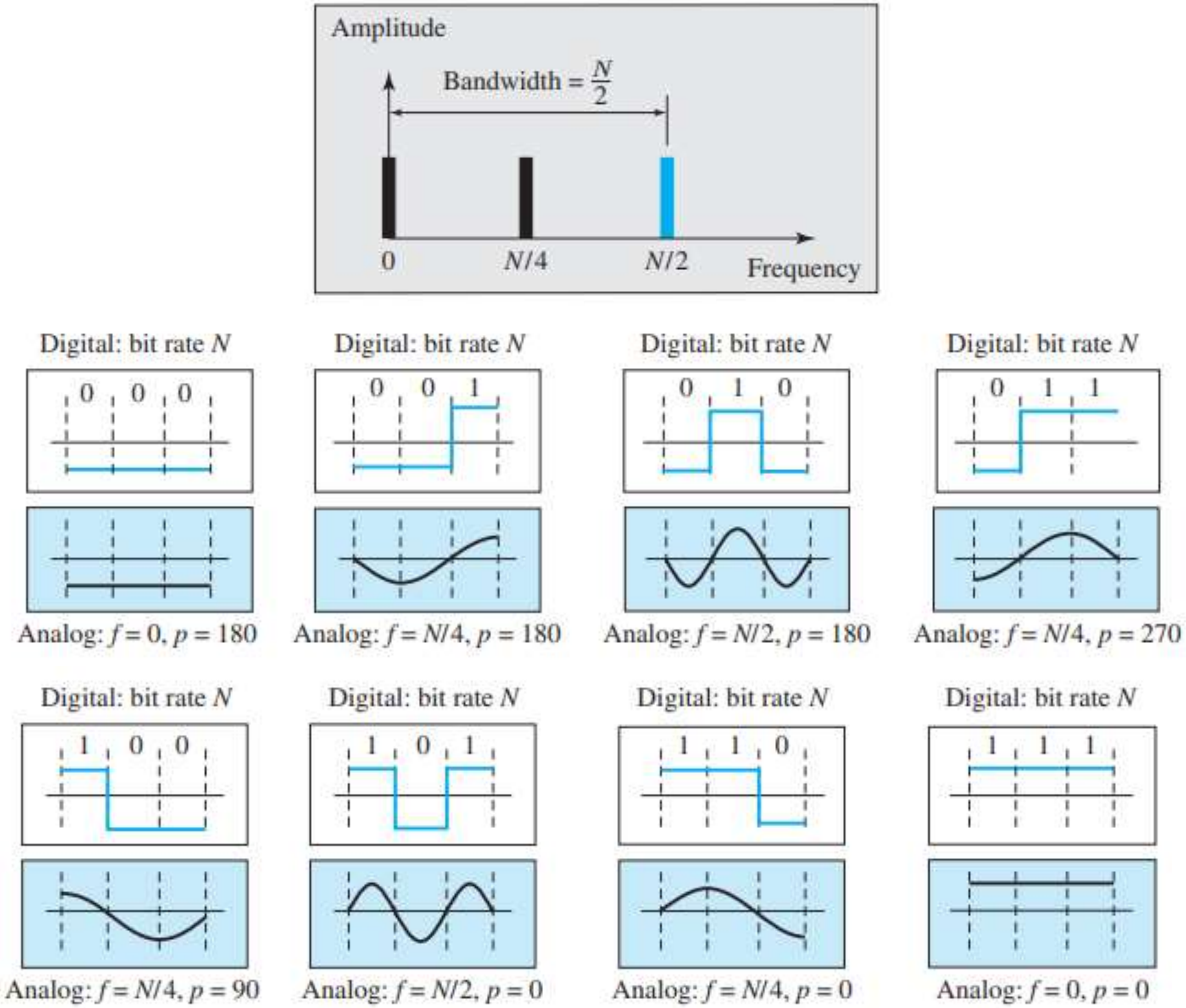
Low-Pass Channel with Limited Bandwidth

- 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.

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

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



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.

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

- **Example 3.22** 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. b. A better result can be achieved by using the first and the third harmonics with the required bandwidth B = 3 × 500 kHz = 1.5 MHz.

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

Example 3.23 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*

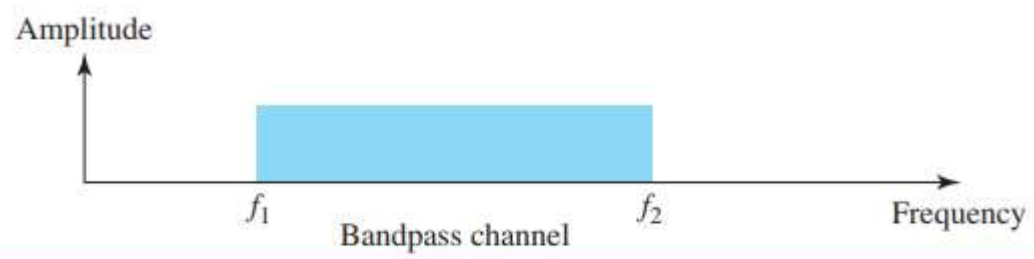
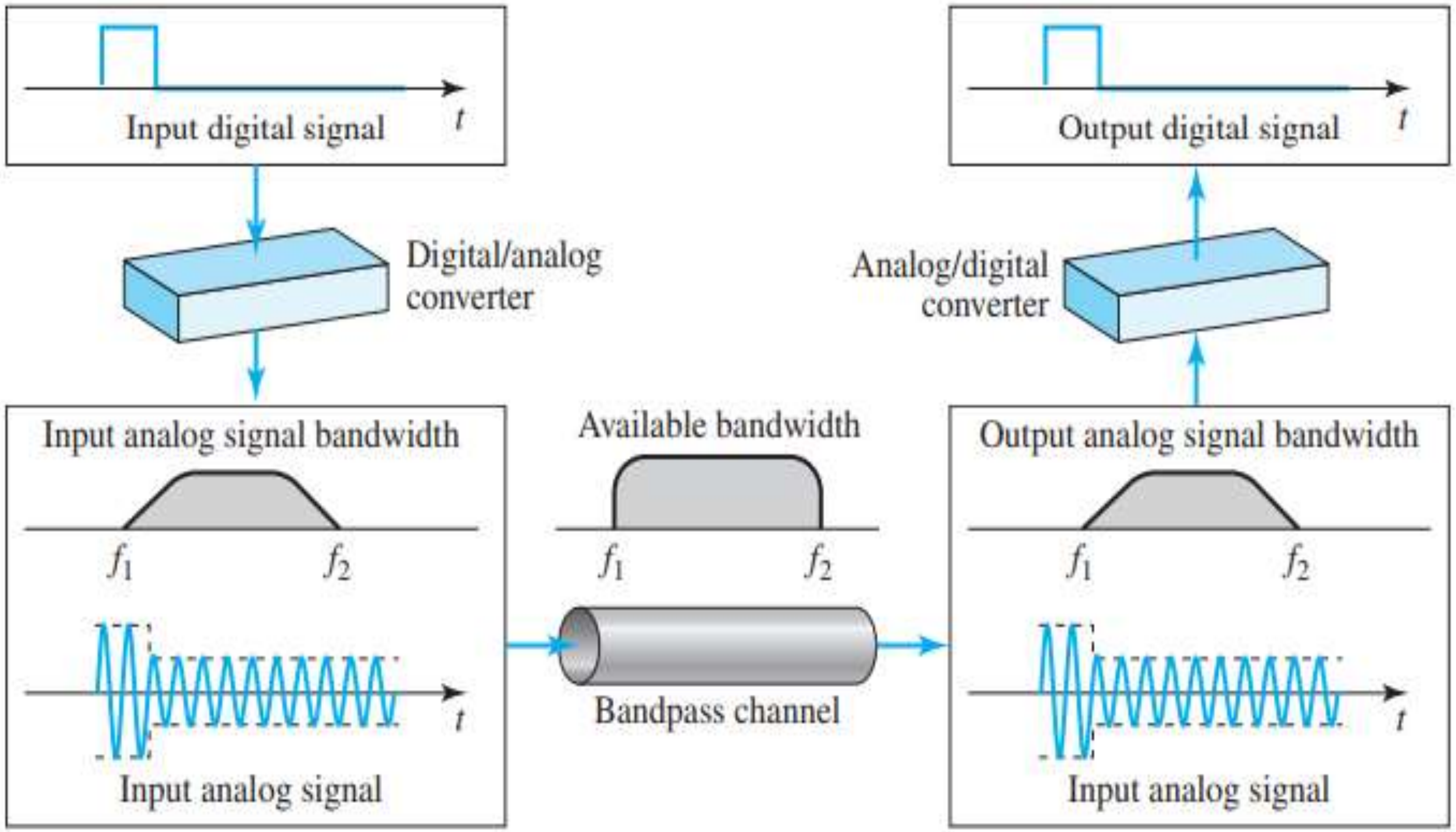
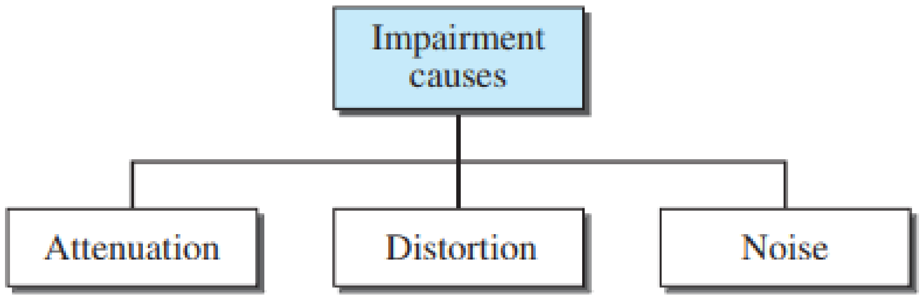


Figure 3.25 *Modulation of a digital signal for transmission on a bandpass channel*



TRANSMISSION IMPAIRMENT

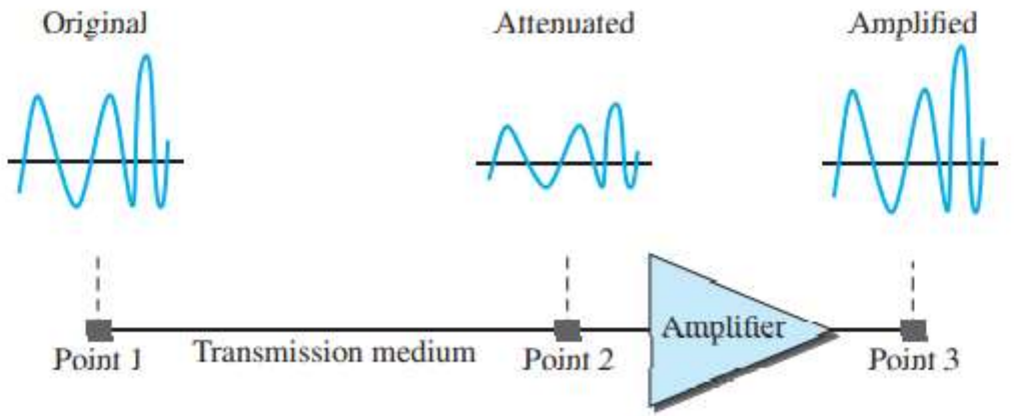
Figure 3.26 *Causes of impairment*



Attenuation:

- Attenuation means a loss of energy
- When a signal, simple or composite, travels through a medium, it loses some of its energy
- 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 two signals or one signal at two different points

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

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

Example 3.26

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

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

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

Example 3.27

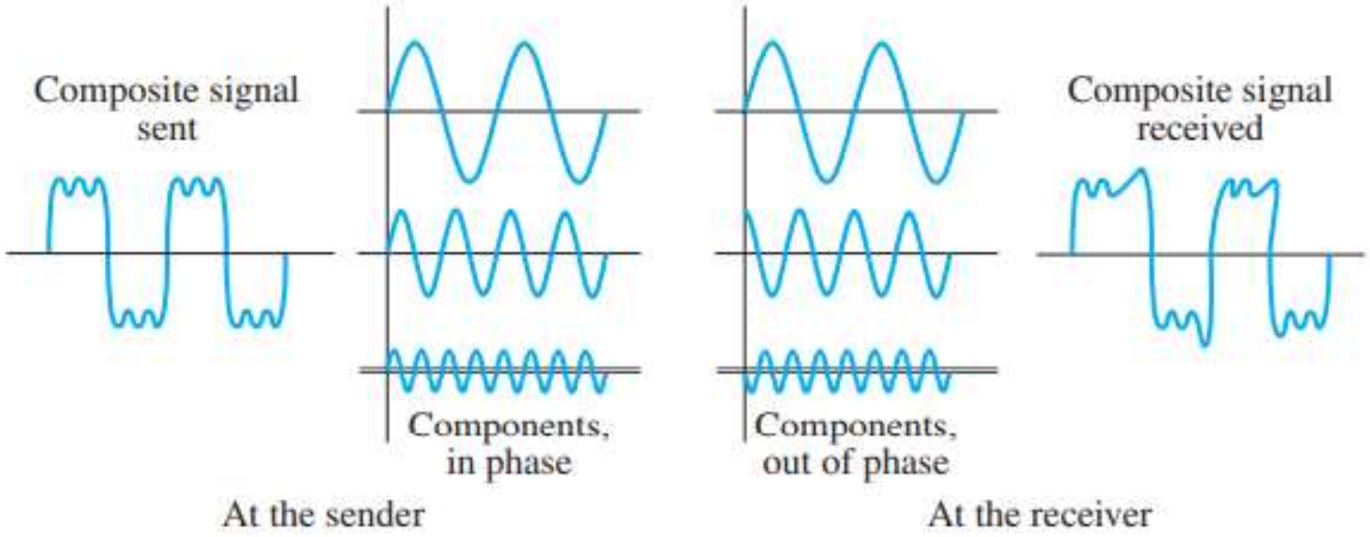
A signal travels through an amplifier, and its power is increased 10 times. This means that $P_2 = 10P_1$. In this case, the amplification (gain of power) can be calculated as

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

Distortion

- Distortion means that the signal changes its form or shape
- Each signal component has its own propagation speed through a medium and, therefore, its own delay in arriving at the final destination

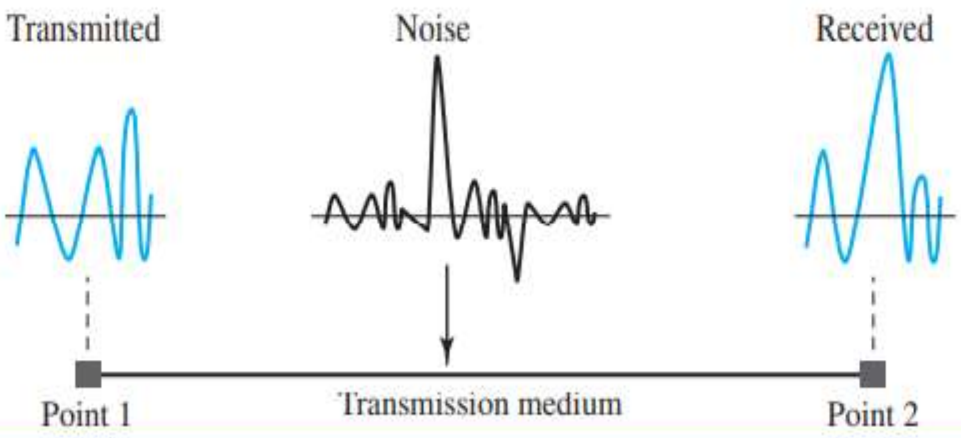
Figure 3.29 *Distortion*



Noise

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

Figure 3.30 *Noise*

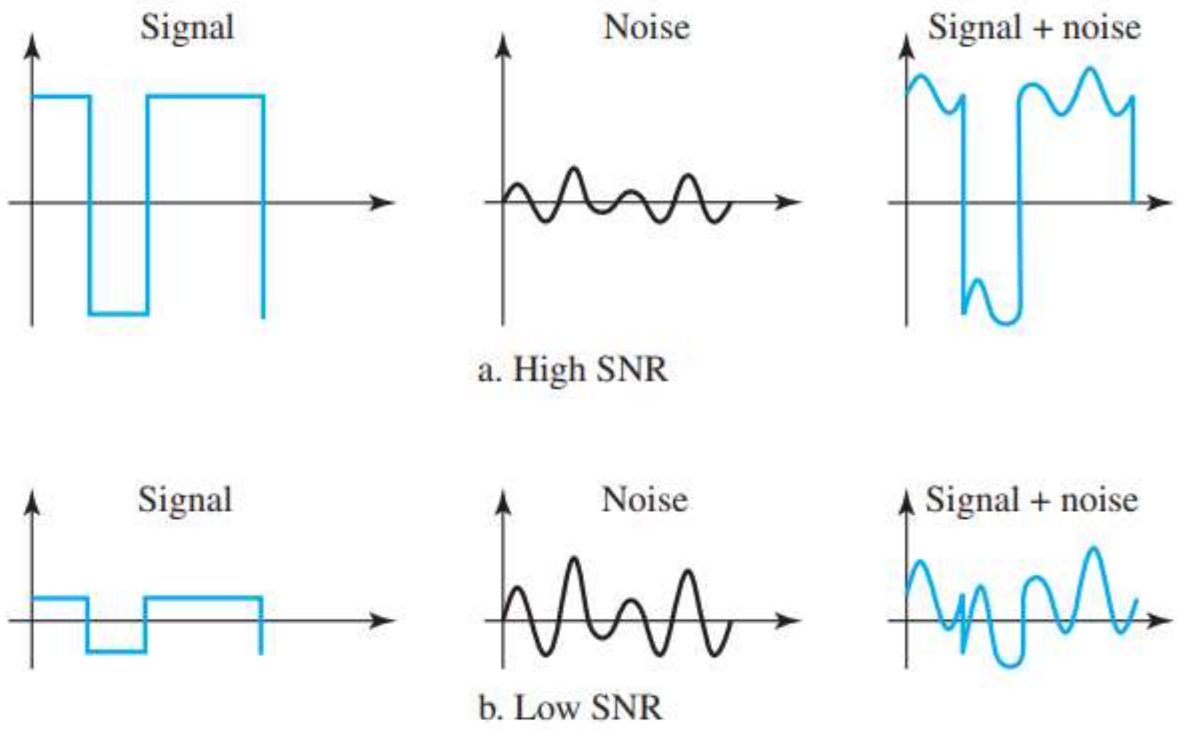


Signal-to-Noise Ratio (SNR)

As we will see later, to find the theoretical bit rate limit, we need to know the ratio of the signal power to the noise power. The **signal-to-noise ratio** is defined as

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

Figure 3.31 *Two cases of SNR: a high SNR and a low 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:

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

Example 3.32

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

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

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

DATA RATE LIMITS

- Data rate depends on three factors:
  - The bandwidth available
  - The level of the signals we use
  - The quality of the channel (the level of noise)

- **Noiseless Channel: Nyquist Bit Rate**

$$\text{BitRate} = 2 \times \text{bandwidth} \times \log_2 L$$

- In this formula, bandwidth is the bandwidth of the channel, L is the number of signal levels used to represent data, and BitRate is the bit rate in bits per second.
- According to the formula, we might think that, given a specific bandwidth, we can have any bit rate we want by increasing the number of signal levels.
- If the number of levels in a signal is just 2, the receiver can easily distinguish between a 0 and 1.
- If the level of a signal is 64, the receiver must be very sophisticated to distinguish between 64 different levels.

Increasing the levels of a signal may reduce 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

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

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

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

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:

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

Noisy Channel: Shannon Capacity

- In 1944, Claude Shannon introduced a formula, called the Shannon capacity, to determine the theoretical highest data rate for a noisy channel:

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

- Shannon formula there is no indication of the signal level, which means that no matter how many levels we have, we cannot achieve a data rate higher than the capacity of the channel.

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

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

This means that the capacity of this channel is zero regardless of the bandwidth. In other words, we cannot receive any data through this channel.

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

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

PERFORMANCE

- Bandwidth
  - Bandwidth in Hertz
  - Bandwidth in Bits per Seconds

In networking, we use the term *bandwidth* in two contexts.

- The first, *bandwidth in hertz,* refers to the range of frequencies in a composite signal or the range of frequencies that a channel can pass.
- The second, *bandwidth in bits per second,* refers to the speed of bit transmission in a channel or link.