Chapter 1: Computer Networks Overview

Let us start our lectures with the following question:
Why wait a week for that letter from Germany to arrive by regular mail when it could appear almost instantaneously through computer networks?

Data communications and networking are changing the way we do business and the way we live. Businesses today rely on computer networks and internetworks (e.g., Internet).

Data communications and ***networking*** enable to exchange data such as text, audio, and video from all points in the world. We want to access the Internet to download and upload information quickly and accurately and at any time.

1.1 Data Communication

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

The term *telecommunication*, which includes telephony, telegraphy, and television, means communication at a distance (tele is Greek for "far").

Data communications are the exchange of data between two devices via some form of transmission medium such as a wire cable. The communicating devices must be part of a communication system which is made up of a combination of **hardware** (physical equipment) and **software** (programs).

The effectiveness of a data communications system depends on four fundamental characteristics:

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

2. **Accuracy**. The system must deliver the data accurately, without errors.

3. **Timeliness**. The system must deliver data in a timely manner. Data delivered late are useless. In the case of video and audio, timely delivery means delivering data as they are produced, without significant delay (*real-time transmission*).

4. **Jitter**. Jitter refers to the variation in the packet arrival time. It is the uneven delay in the delivery of audio or video packets. For example, let us assume that video packets are sent every 3D-ms. If some of the packets arrive with 3D-ms

delay and others with 4D-ms delay, an **uneven quality** in the video is the result.

1.2 Components of data communications system

A data communications system has five components (see Figure 1.1).

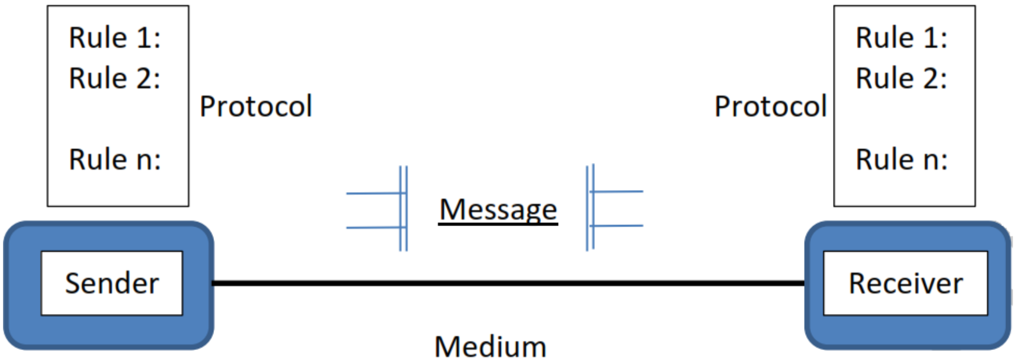


Figure 1.1 *Five components of data communication system*

1. **Message**. It is the information (data) to be communicated. Popular forms of information include text, numbers, pictures, audio, and video.

2. **Sender**. It is the device that sends the data message. It can be a computer, workstation, telephone handset, video camera, and so on.

3. **Receiver**. It is the device that receives the message. It can be a computer, workstation, telephone handset, television, and so on.

4. **Transmission medium**. It is the physical path by which a message travels from sender to receiver. Some examples of transmission media include twisted-pair wire, coaxial cable, fiber-optic cable, and radio waves.

5. **Protocol**. It is a set of rules that govern data communications. It represents an agreement between the communicating devices. Without a protocol, two devices *may be connected but not communicating*, just as a person speaking French cannot be understood by a person who speaks only Japanese.

1.3 Data Representation

Information today comes in different **forms** such as:

Text In data communications, text is represented as a bit pattern, a sequence of bits (0s or 1s). Different sets of bit patterns have been designed to represent text symbols, for example, the ***A**merican **S**tandard **C**ode for **I**nformation **I**nterchange* (**ASCII**) which uses 32 bits to represent a symbol or character used in any language in the world. Each set is called a code, and the process of representing symbols is called coding.

Numbers are also represented by bit patterns. However, a code such as ASCII is not used to represent numbers; the number is directly converted to a binary number.

Images are also represented by bit patterns. In its simplest form, an image is composed of a matrix of pixels (picture elements), where each pixel is a small dot. The size of the pixel depends on the *resolution.*

Audio refers to the recording or broadcasting of sound or music. Audio is by nature different from text, numbers, or images. It is continuous, not discrete.

Video refers to the recording or broadcasting of a picture or movie. Video can either be produced as a continuous entity (e.g., by a TV camera), or it can be a combination of images.

All these information forms **represent the material** that may be sent and received ***in any data communication system.***

1.4 Data Flow

Communication between two devices can be simplex, half-duplex, or full- duplex as shown in Figure 1.2.

Simplex In simplex mode, the communication is unidirectional, as on a one- way street. For example, the keyboard can only introduce input; the monitor can only accept output.

Half-Duplex In half-duplex mode, each station can both transmit and receive, but not at the same time. When one device is sending, the other can only receive, and vice versa, for example, the Walkie-talkies devices.

Full-Duplex In full-duplex mode (also called duplex), both stations can transmit and receive simultaneously, for example, the telephone network.

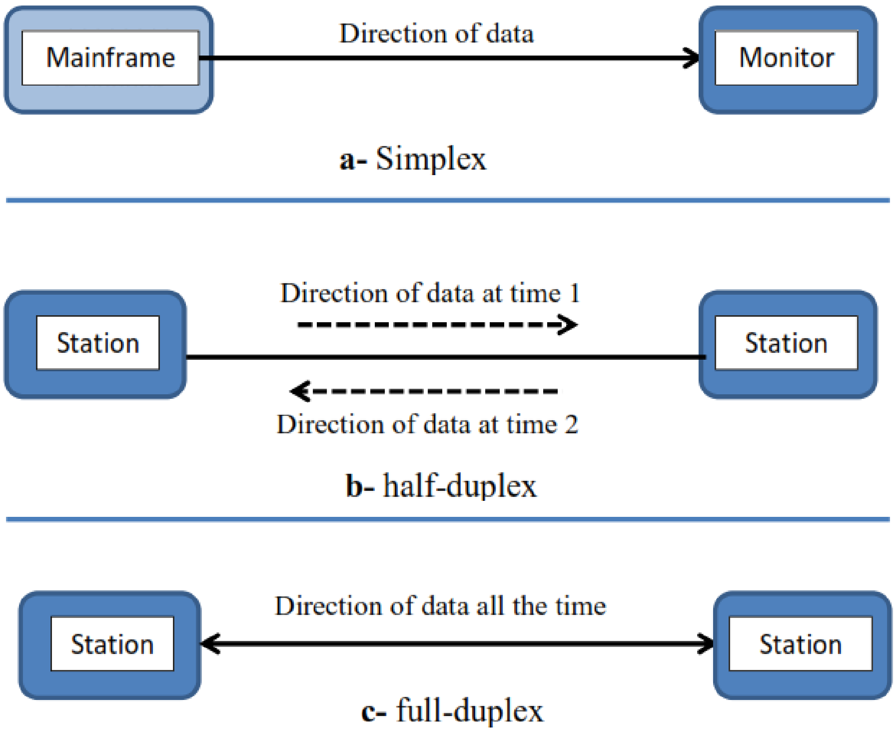


Figure 1.2 *Data flow (simplex, half-duplex, and full-duplex)*

1.5 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.

1.5.1 Distributed Processing

Most networks use distributed processing, in which a task is divided among multiple computers. Instead of one single large machine being responsible for all aspects of a process, separate computers (usually a personal computer or workstation) handle a subset.

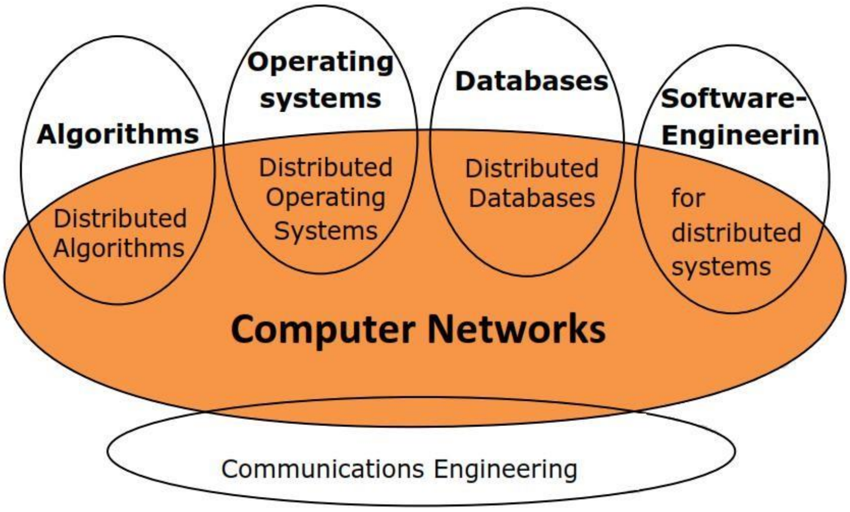


Figure 1.3 Distributed processing in computer networks vs. other Computer Science Classes

1.5.2 Types of Network Connections

There are two possible types of connections:

Point-to-Point This connection provides a dedicated link between two devices. The entire capacity of the link is reserved for transmission between those two devices, (see Figure 1.3a). For example: establishing a point-to-point connection between the remote control and the television's control system.

Multipoint In this connection, more than two specific devices share a single link (see Figure 1.3b). The capacity of the channel is shared, either spatially or temporally. If several devices can use the link simultaneously, it is a ***spatially shared*** connection. If users must take turns, it is a ***timeshared*** connection.

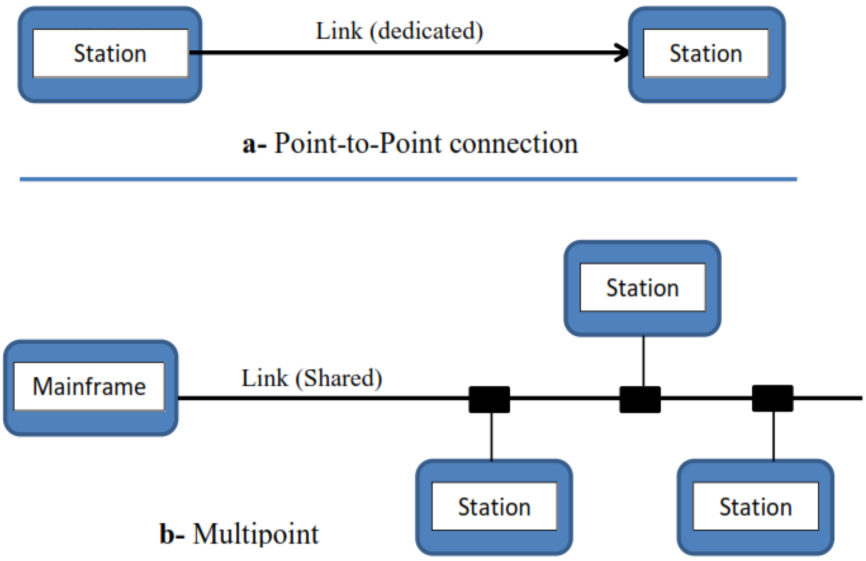


Figure 1.3 *Types of connections: point-to-point and multipoint*

1.5.3 Physical Network Topology

The term ***physical topology*** refers to the way in which a network is arranged physically. There are four possible basic topologies:

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

***Advantages*:**
1) Eliminating the traffic problems.
2) If one link becomes unusable, it does not fall the entire system.
3) High privacy or security.

Disadvantages:
1) Amount of cabling is required.
2) Number of I/O ports (e.g., LAN Cards) is required.

Numbers of cables and hosts

- Number of cables: $n(n-1)/2$
- Number of ports: $n(n-1)$ **n**: number of hosts

2. Star Topology In a star topology, each device has a dedicated point-to-point link only to a central controller, usually called a ***hub***. For example, the star topology is used in local-area networks (**LANs**).

***Advantages*:**
1) Each device needs only one link and one I/O port to connect it.
2) Easy to install and reconfigure.

Disadvantages:
Single point dependency. If the hub goes down, the whole network is dead.

Numbers of cables and hosts

- Number of cables: $1 \times n$
- Number of ports: $1 \times n$ **n**: number of hosts

Example: Assume that we have 5 hosts in Mesh and Star topologies. What are the numbers of cables and ports are needed?

For Mesh Topology

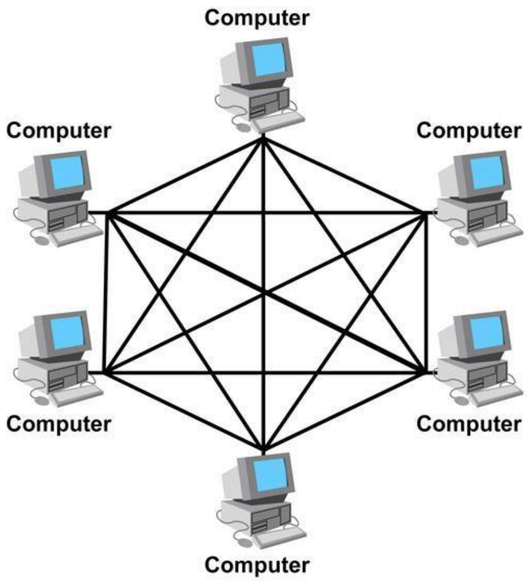
- Number of cables: $\boldsymbol{n\ (n-1)\ /\ 2}$

  $5\ (5 - 1)\ /\ 2 = 10$ cables
- Number of ports: $\boldsymbol{n\ (n-1)}$

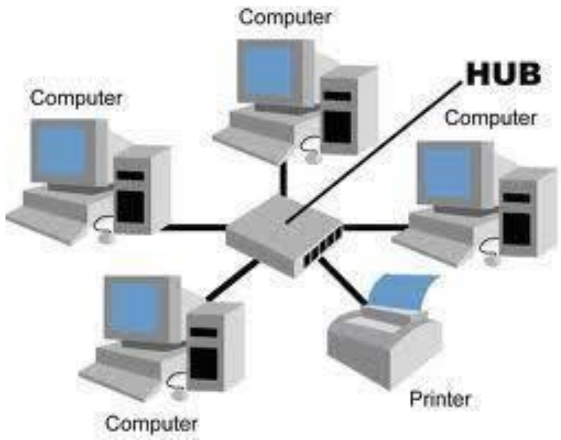
  $5\ (5 - 1) = 20$ ports

For Star Topology

- Number of cables: $\mathbf{1 \times} \boldsymbol{n} = 5$ cables
- Number of ports: $\mathbf{1 \times} \boldsymbol{n} = 5$ ports



Mesh network topology



Star network topology

3. Bus Topology This topology is multipoint. One long cable acts as a backbone to link all the devices in a network. Nodes are connected to the bus cable by ***drop lines*** and ***taps***.

<u>Advantages:</u>

1) Ease of installation.

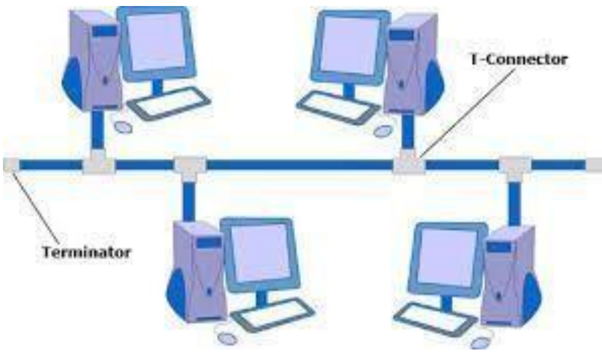
2) Less cabling than mesh or star topologies.

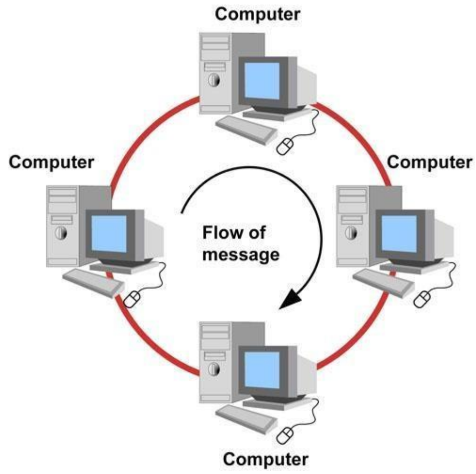
<u>Disadvantages:</u> difficult reconnection and fault isolation.

4. Ring Topology Each device has a dedicated point-to-point connection with only the two devices on either side of it. A signal is passed along the ring in one direction, from device to device, until it reaches its destination.

Advantages:
Easy to install and reconfigure. Each device is linked to only its immediate neighbours (either physically or logically).

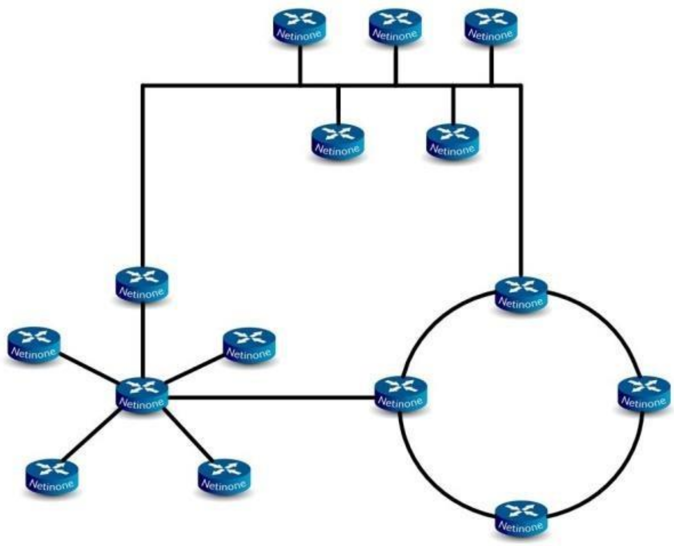
Disadvantages:
Unidirectional traffic i.e., a break in the ring (such as a disabled station) can disable the entire network.





Bus network topology *Ring network topology*

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



1.5.4 Categories of Networks

The category of networks is determined by its ***size*** and ***connection type.*** Figure 1.5 classifies the networks by their rough physical *size*.

Distance	Location	Network category
1 m	Square meter	Personal area network
10 m	Room	Local area network
100 m	Building	Local area network
1 km	Campus	Local area network
10 km	City	Metropolitan area network
100 km	Country	Wide area network
1000 km	Continent	Wide area network
10,000 km	Planet	The Internet

Figure 1.5 *Classification of networks by size*

1. Personal Area Networks (PANs) let devices communicate over the range of a person. A common example is a wireless network that connects a computer with its peripherals: monitor, keyboard, mouse, and printer. In the simplest form, Figure 1.6 shows a Bluetooth network uses the master-slave paradigm.

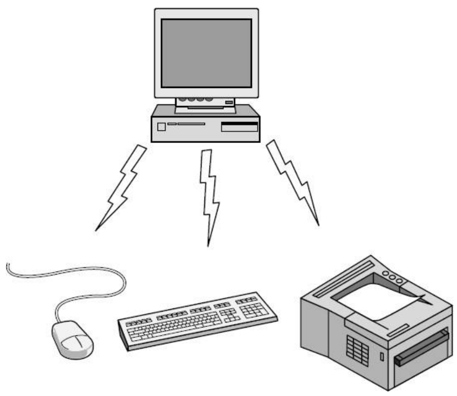


Figure 1.6 *Bluetooth PAN Configuration*

2. Local area networks (LANs) link the devices in a single office, building, or campus (see Figure 1.7). LANs can be as simple as two PCs and a printer in someone's home office; or it can extend throughout a company, they are called ***enterprise networks***.

Wireless LANs are very popular these days in which every computer has a radio modem (e.g., wireless LAN card) to communicate with other computers. Each computer talks to an *Access Point* or *wireless router* (see Figure 1.7(a)) to relay data packets between the wireless computers and also with the Internet. There is a standard for wireless LANs called ***IEEE 802.11***, popularly known as ***WiFi***. It runs at speeds from **11 to hundreds of Mbps.**

Wired LANs use a range of different transmission technologies. Most of them use copper wires, but some use optical fibre. Typically, wired LANs run at **speeds of 100 Mbps to 1 Gbps**, have low delay (microseconds or nanoseconds), and make very few errors. Compared to wireless networks, wired LANs exceed them in all dimensions of performance. ***IEEE 802.3***, popularly called ***Ethernet***, is the most common type of wired LAN. Figure 1.7(b) shows a sample topology of switched Ethernet.

Note: The 1 Mbps (Megabits/sec) is 1,000,000 bits/sec, and the 1 Gbps (gigabits/sec) is 1,000,000,000 bits/sec.)

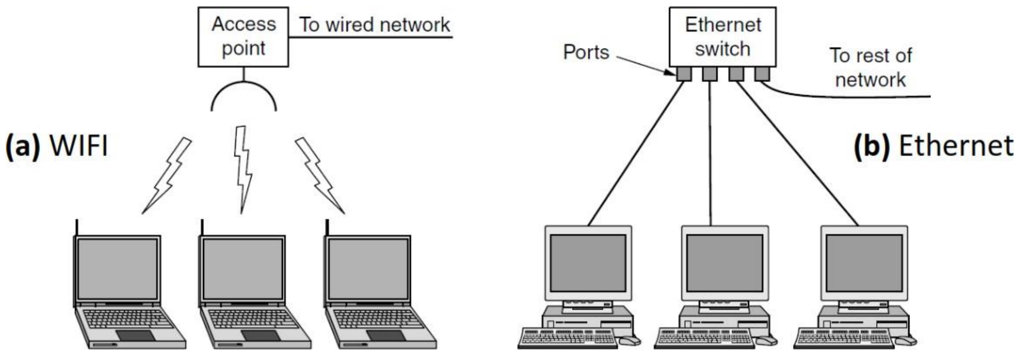


Figure 1.7 *Wireless and wired LANs*

3. Metropolitan Area Network (MAN) covers a city. The best-known examples of MANs are the cable television networks available in many cities. In Figure 1.8 we see both television signals and Internet being fed into the centralized cable head-end for subsequent distribution to people's homes.

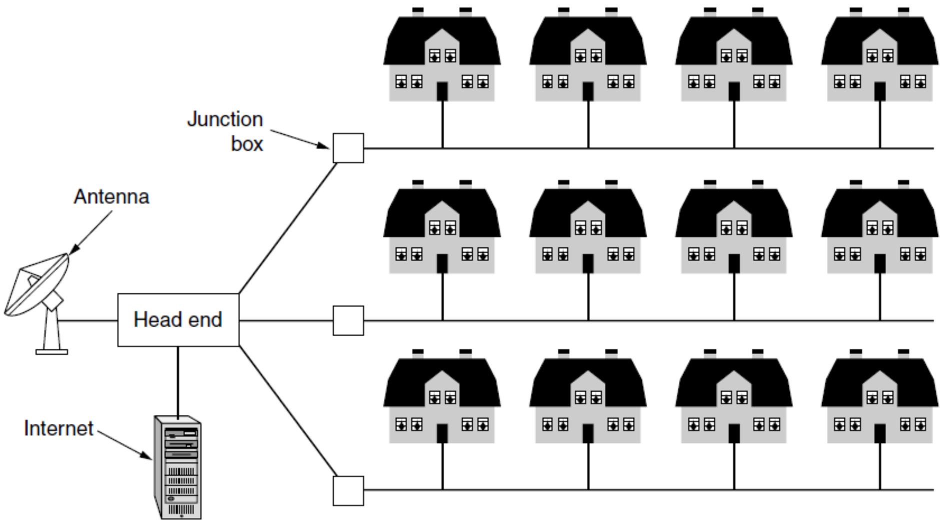


Figure 1.8 A metropolitan area network based on cable TV

4. A wide area network (WAN) provides long-distance transmission of data over large geographic areas that may comprise a country, a continent, or even the whole world.

The WAN in Figure 1.9 is a network in Australia that connects offices in Perth, Melbourne, and Brisbane. Each of these offices contains computer machines intended for running user (i.e., application) programs. Each of these machines are called host. The rest of the network that connects these hosts is then called the **communication subnet**, or just **subnet** for short.

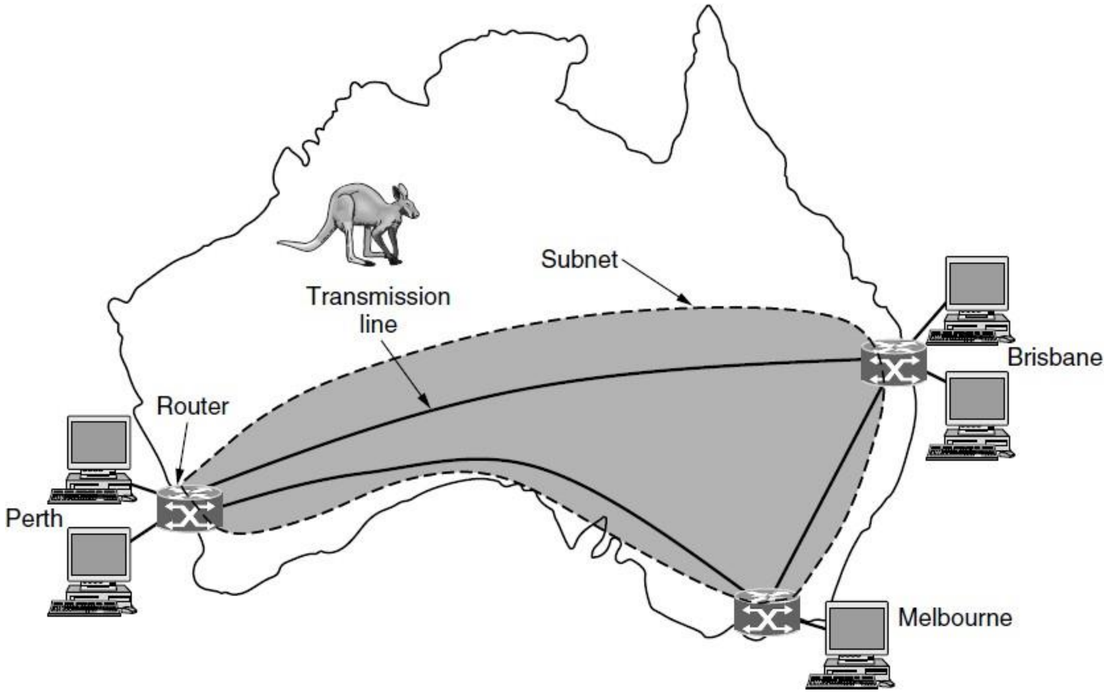


Figure 1.9 *WAN that connects three branch offices in Australia*

5. Interconnection of Networks: Internetwork

Today, it is very rare to see a LAN, a MAN, or a WAN in isolation; they are connected to one another. When two or more networks with different hardware and software are connected, they become an ***internetwork***, or ***internet***. (i.e., it is a collection of interconnected networks). These terms is used in a generic sense, in contrast to the worldwide **Internet** (*which is one specific **internet***), which we will always capitalize.

6. Internet

The Internet is a communication system that has brought a wealth of information to our fingertips and organized it for our use. The Internet today is not a simple hierarchical structure. It is made up of many wide- and local-area networks joined by connecting lines and switching devices (see Figure 1.10).

The Internet uses ISP networks to connect enterprise networks, home networks, and many other networks.

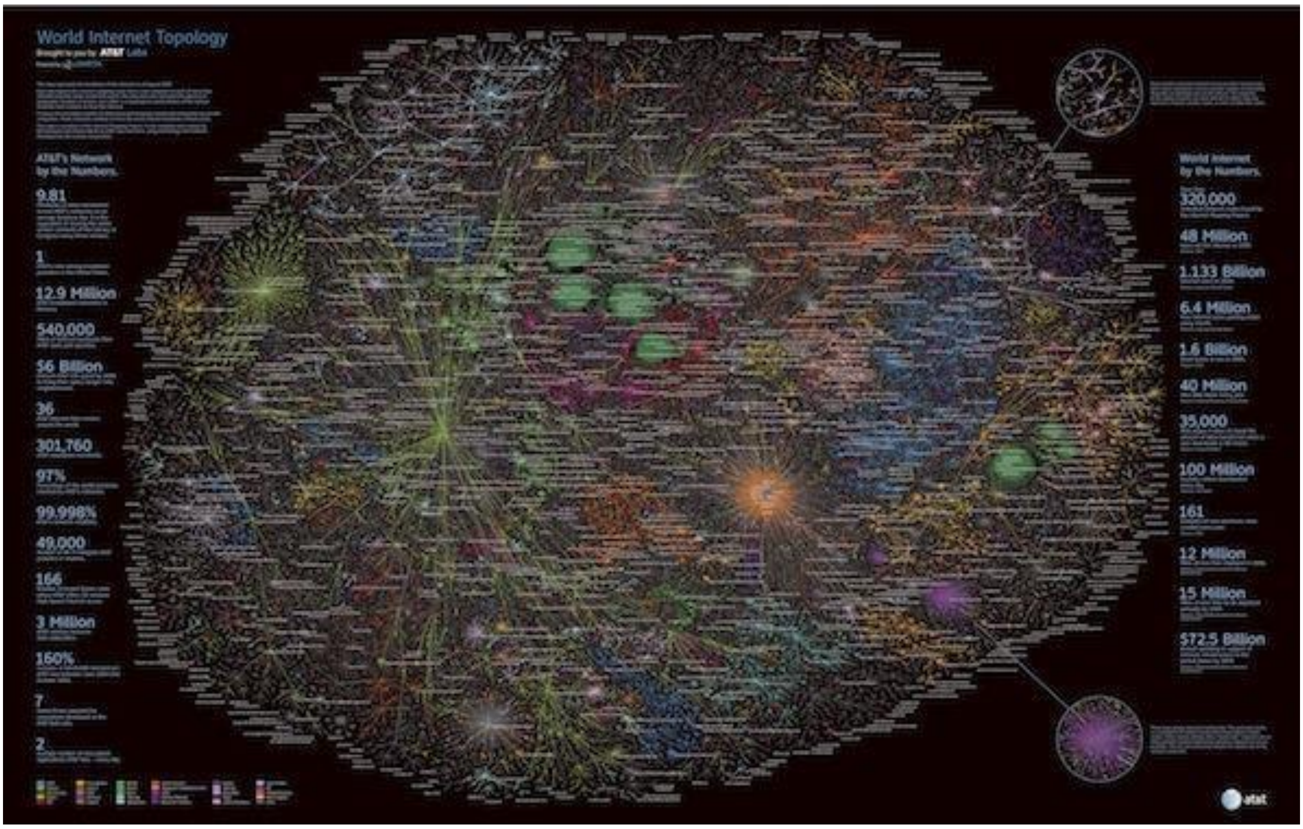
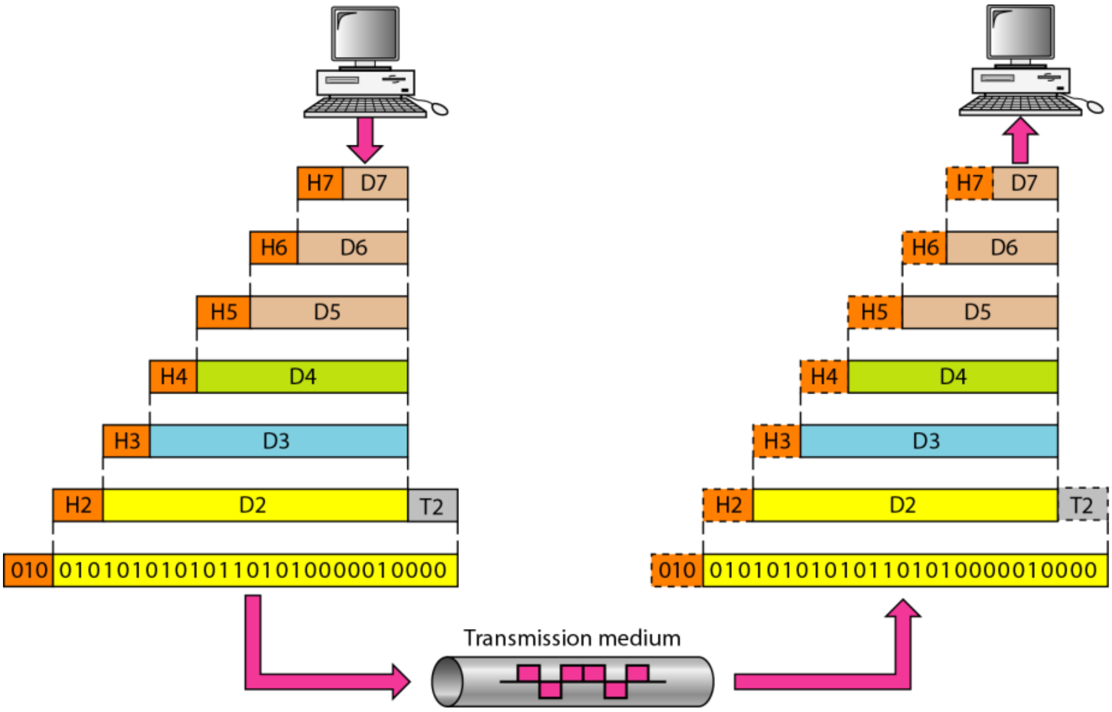


Figure 1.10 *Internet Map*

Chapter 2: Network Standards and OSI Model



2.1 Network Criteria

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

a) Performance can be ***measured*** in many ways, including *transit time* and *response time*. **Transit time** is the amount of time required for a message to travel from one device to another. **Response time** is the elapsed time between an inquiry and a response.

The performance of a network **depends** on a number of ***factors***:
1) Number of users.
2) Type of transmission medium.
4) Capabilities of the connected hardware.
5) Efficiency of the software.

Performance is often ***evaluated*** by two networking ***metrics***: *throughput* and *delay*. We often need more *throughput* and *less delay.*

b) Reliability network reliability is measured by the:
1) Accuracy of delivery.
2) Frequency of failure.
3) The time it takes a link to recover from a failure.
4) Network's robustness in a catastrophe.

c) Security Network security issues include:
1) Protecting data from *unauthorized access*
2) Protecting data from *damage and change.*
3) Implementing policies for recovery from data losses.

2.2. Protocols and Standards

A) Protocol

A protocol defines ***what*** is communicated, ***how*** it is communicated, and ***when*** it is communicated. The *key elements* of a protocol are syntax, semantics, and timing.

Syntax It refers to the *structure or format* of the data, meaning the order in which they are presented.

<u>*For example*</u>, a simple protocol might expect the first 8 bits of data to be the address of the sender, the second 8 bits to be the address of the receiver, and the rest of the stream to be the message itself.

Semantics It refers to the meaning of each section of bits. How is a particular pattern to be interpreted, and what action is to be taken based on that interpretation? *For example*, does an address identify the route to be taken or the final destination of the message?

B) Standards

Standards are essential in creating an open and competitive market for equipment manufacturers and in guaranteeing national and international telecommunications.

Standards Creation Committees

Most data telecommunications rely primarily on the standards published by the following committees:

- *International Organization for Standardization* (**ISO**). The ISO is a multinational. It is active in the fields of scientific, technological, and economic activity.

- *American National Standards Institute* (**ANSI**). The ANSI is a completely private in USA. However, all ANSI activities are undertaken.

- *Institute of Electrical and Electronics Engineers* (**IEEE**). This institute is the largest professional engineering society in the world. It aims to advance theory, creativity, and product quality in the fields of electrical engineering, electronics, and radio.

- *Electronic Industries Association* (**EIA**). In the field of *information technology* (IT), the EIA has made significant contributions by defining *physical connection interfaces* and *electronic signalling specifications* for data communication.

2.3 Network Models

The layered model that dominated data communications and networking literature before 1990 was the *Open Systems Interconnection* (**OSI**) model. Everyone believed that the OSI model would become the ultimate standard for data communications, but this did not happen. *The **TCP/IP** protocol suite became the dominant commercial architecture* **because** it was used and tested extensively in the Internet; ***the OSI model was never fully implemented***. Figure 2.1 shows the layers included in the TCP/IP and OSI models and also the main protocols and services provided by each layer.

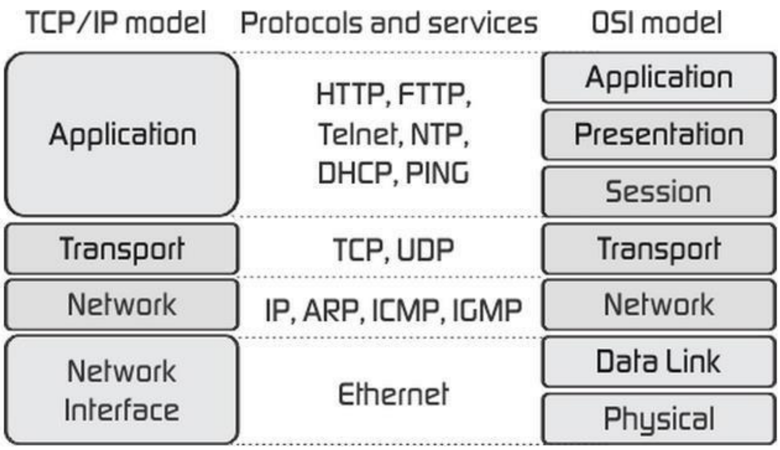


Figure 2.1 The TCP/IP and OSI models for data communications and networking

2.3.1 Layered Task

We use the concept of layers in our daily life. As an example, let us consider two friends who communicate through postal mail. The process of sending a letter to a friend would be complex if there were no services available from the post office. Figure 2.2 shows the steps in this task.

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

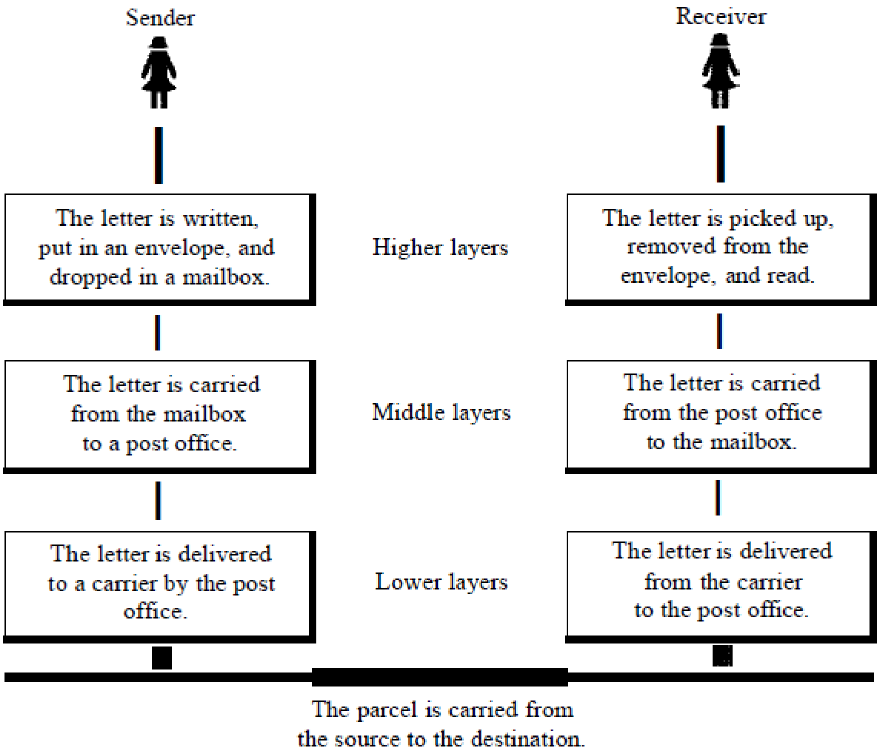


Figure 2.2 *Tasks involved in sending a letter*

2.3.2 The OSI Model

The ***Open Systems Interconnection*** (OSI) model is introduced in the late 1970s by the ***International Standards Organization*** (ISO).
(**Note:** ISO is the organization. OSI is the model.)

The OSI model is a layered model for the design and understands of network systems that allows communication between all types of computer systems.

OSI consists of seven separate but related layers, each of which defines a part of the process of moving information across a network (*see the right side of Figure 2.1*).

Figure 2.3 shows the layers involved when a message is sent ***for example*** from device **A** to device **B**. As the message travels from A to B, it may pass through many intermediate nodes, **called routers**. These *intermediate nodes usually involve only the first three layers* of the OSI model.

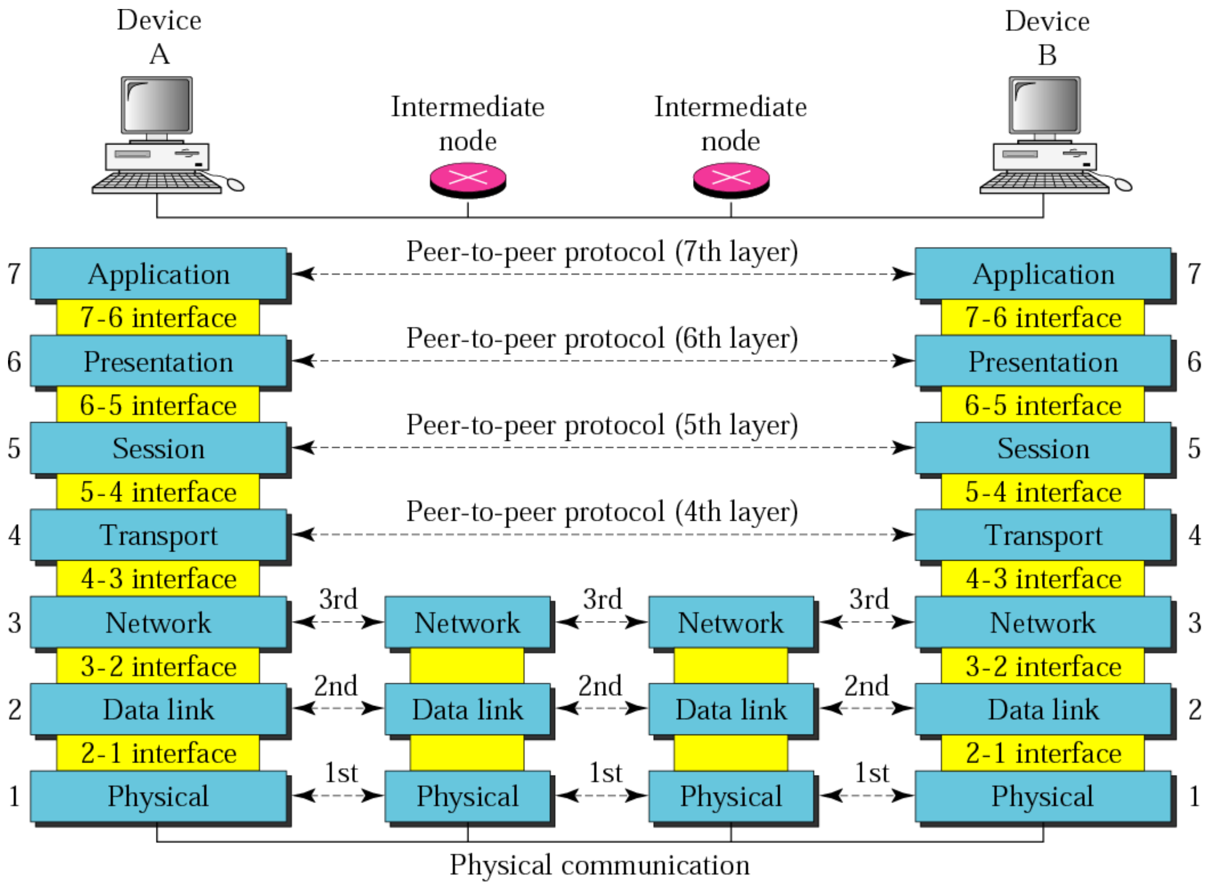


Figure 2.3 *The interaction between layers in the OSI model*

Each layer defines a family of functions distinct from those of the other layers. Layer 3, for example, uses the services provided by layer 2 and provides services for layer 4.

The processes on each machine that communicate at a given layer are called ***peer-to-peer processes***. Communication between machines is therefore a peer- to-peer process using the protocols appropriate to a given layer.

The **interfaces** between layers are to define the *information and services* that each layer must provide for the layer above it.

The seven layers are belonging to three subgroups. Layers 1, 2, and 3 are the network support layers; they deal with the physical aspects of moving data from one device to another (such as electrical specifications, physical connections, and physical addressing). Layers 5, 6, and 7 are the user support layers; they allow interoperability among unrelated software systems. Layer 4, the transport layer, links the two subgroups.

2.3.3 Layers in the OSI Model

In the following we describe the functions of each layer in the OSI model.

1. **Physical Layer**

The physical layer coordinates the functions required to carry a bit stream over a physical medium (see Figure 2.4).

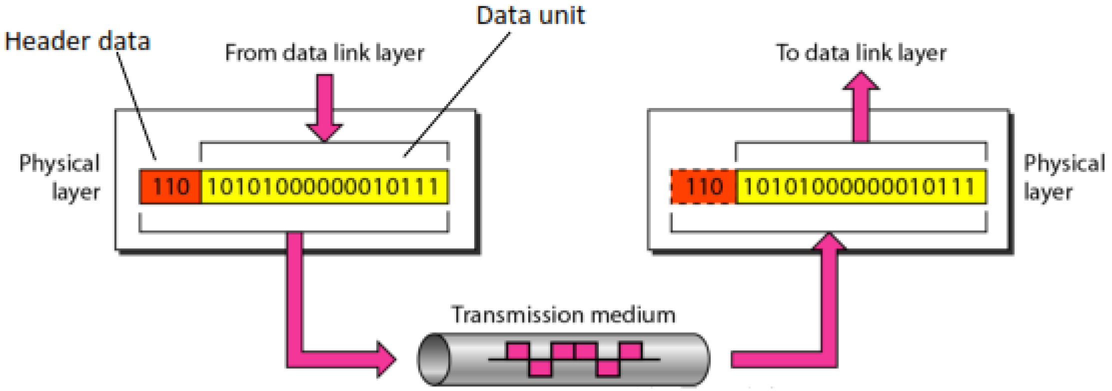


Figure 2.4 *Physical layer*

The physical layer is concerned with the following:

- ➢ Physical characteristics of interfaces and medium.
- ➢ Representation of bits (sequence of 0s or 1s) with defining the type of encoding (how 0s and 1s are changed to signals).
- ➢ Data rate or the transmission rate (the number of bits sent each second)

- ➢ Synchronization of bits. The sender and receiver not only must use the same bit rate but also must be synchronized at the bit level.
- ➢ Line configuration (point-to-point or multipoint configuration).
- ➢ Physical topology (mesh, star, bus, etc.).
- ➢ Transmission mode (simplex, half-duplex, or full-duplex).

2. Data Link Layer

The data link layer makes the physical layer appear error-free to the upper layer (network layer). Figure 2.5 shows the relationship of the data link layer to the network and physical layers.

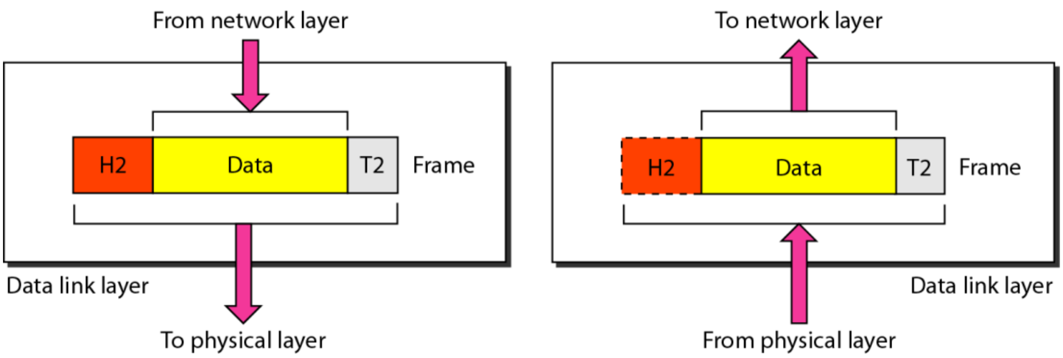


Figure 2.5 Data Link layer

The data link layer is responsible for moving frames from one hop (node) to the next. Other responsibilities of the data link layer include the following:

- ➢ Framing. The data link layer divides the stream of bits received from the network layer into manageable data units called **frames** (see Figure 2.5).
- ➢ Physical addressing. The data link layer adds a header to the frame to define the sender and/or receiver devices of the frame. The devises are defined by the physical address (or MAC address).
- ➢ Flow control. The data link layer ensures the rate at which data are produced in the sender and arrived in the receiver.
- ➢ Error control. The data link layer adds mechanisms to detect and retransmit damaged or lost frames. Error control is normally achieved through a **trailer** added to the end of the frame (see the **T2** in Figure 2.5).
- ➢ Access control. The data link layer controls the access to the link when two or more devices are connected to the same link.

3. Network Layer

The network layer is responsible for the **source-to-destination delivery** of a **packet** (see Figure 2.6), possibly across multiple networks (links). Whereas the data link layer oversees the delivery of the packet between two systems *on the same* network (links).

Other responsibilities of the network layer include the following:

- Logical addressing. The network layer adds a header to the packet coming from the upper layer that, among other things, includes the logical addresses (**IP address**) of the sender and receiver. We discuss logical addresses in a next lecture.

- Routing. When independent networks or links are connected to create internetworks, the connecting devices (called routers or switches) route or switch the packets to their final destination. One of the functions of the network layer is to provide this mechanism.

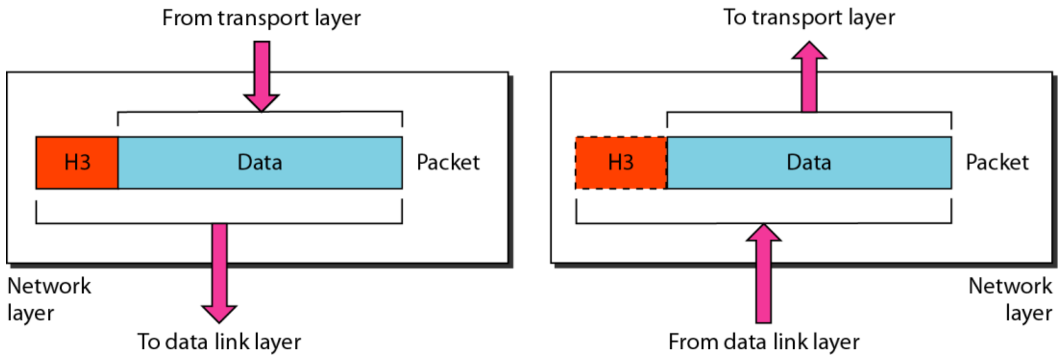


Figure 2.6 Network layer

4. <u>Transport Layer</u>

The transport layer is responsible for **process-to-process delivery** of the entire message. A process is an application program running on a host. Whereas the network layer oversees source-to-destination delivery of individual packets.

Other responsibilities of the transport layer include the following:

- Service-point addressing. Computers often run several programs at the same time. The transport layer header includes a type of address called a service- point address (or **port address**) *in order to deliver the entire message to the correct process on that computer*.
- Segmentation and reassembly. A message is divided into transmittable segments, with each segment containing a sequence number. These numbers enable the transport layer to reassemble the message correctly at the destination (see Figure 2.7).

➢ Flow control. Like the data link layer, the transport layer is responsible for flow control. However, flow control at this layer is performed end to end rather than across a single link.

➢ Error control. Like the data link layer, the transport layer is responsible for error control. However, error control at this layer is performed process-to-process rather than across a single link.

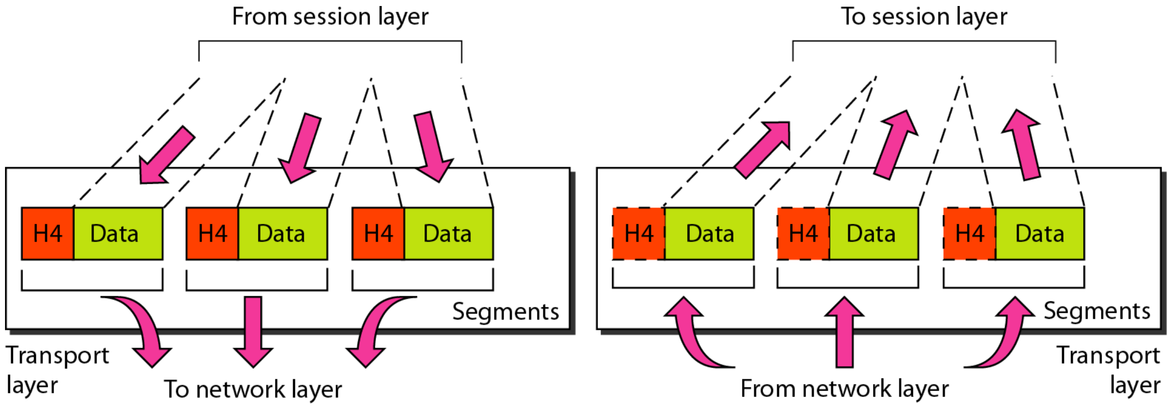


Figure 2.7 Transport layer

5. <u>Session Layer</u>

The services provided by the first three layers (physical, data link, and network) are not sufficient for some processes. The session layer is responsible for dialog control and synchronization.

Specific responsibilities of the session layer include the following:

➢ Dialog control. The session layer allows the communication between two processes to take place in either half-duplex (one way at a time) or fullduplex (two ways at a time) mode.

➢ Synchronization. The session layer allows a process to add checkpoints, or synchronization points, to a stream of data (see Figure 2.8). For example, if a system is sending a file of 2000 pages, it is advisable to insert checkpoints after every 100 pages.

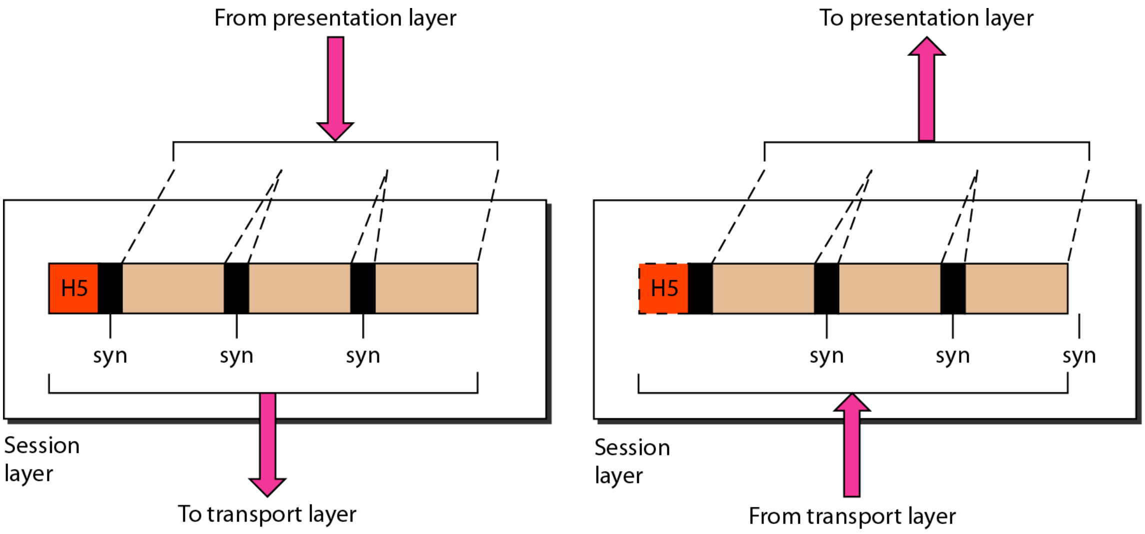


Figure 2.8 Session layer

6. Presentation Layer

The presentation layer is concerned with the syntax and semantics of the information exchanged between two systems (see Figure 2.9).

Specific responsibilities of the presentation layer include the following:

- Translation. Because different computers use different encoding systems, the presentation layer is responsible for interoperability between these different encoding methods.
- Encryption. To carry sensitive information, a system must be able to ensure privacy.
- Compression. Data compression **reduces** the number of bits contained in the information. Data compression becomes particularly important in the data transmission.

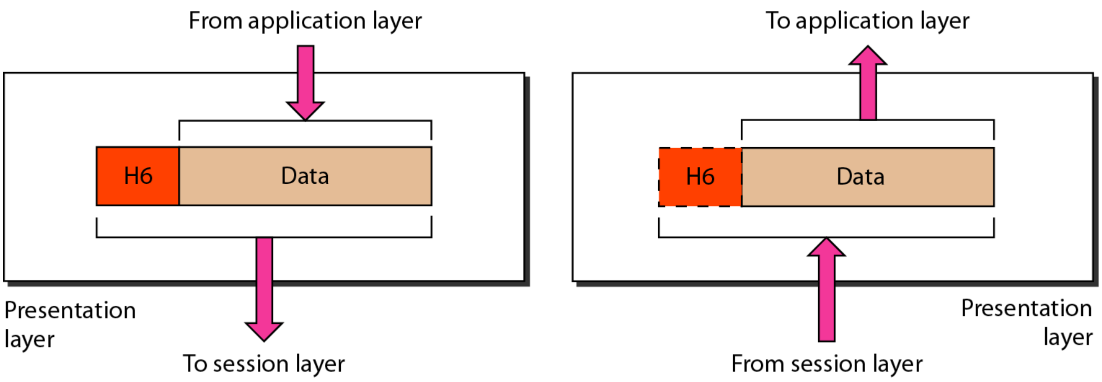


Figure 2.9 Presentation layer

7. Application Layer

The application layer is responsible for providing services to the user. It enables the user, whether human or software, to access the network. It provides user interfaces and support for services such as electronic mail, remote file access and transfer, and other types of distributed information services.

Figure 2.10 shows only three application services available on a user computer: *XAOO* (message-handling services), X.500 (directory services), and file transfer, access, and management (FTAM). The user in this example employs *XAOO* to send an e-mail message.

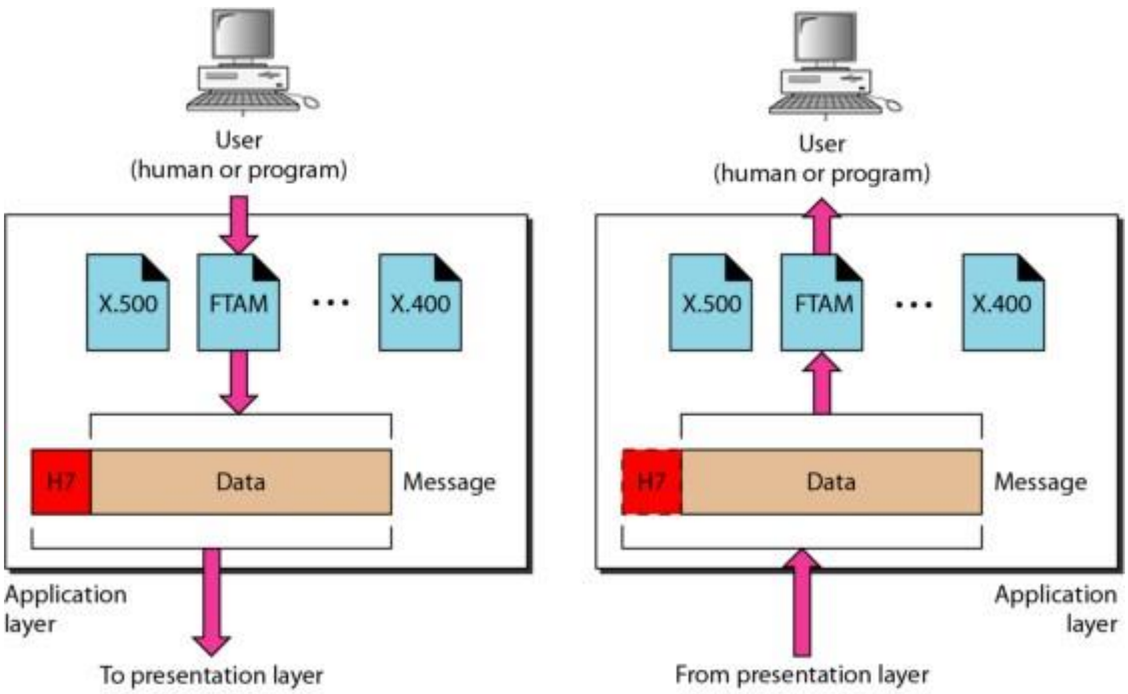


Figure 2.10 Application layer

Summary of OSI Layers

Figure 2.11 shows a summary of duties for each layer in the OSI model.

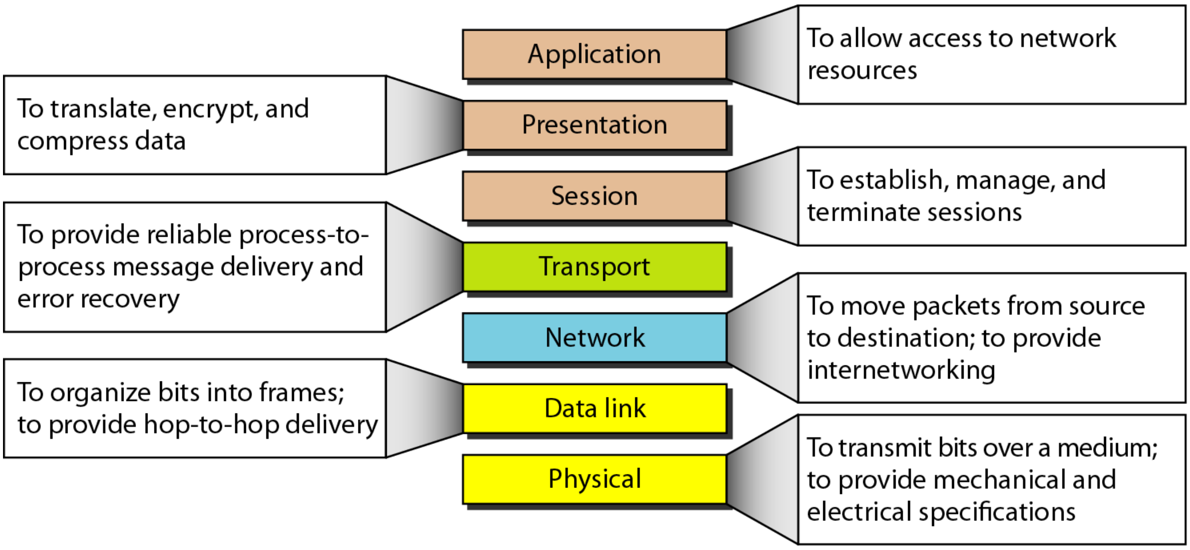
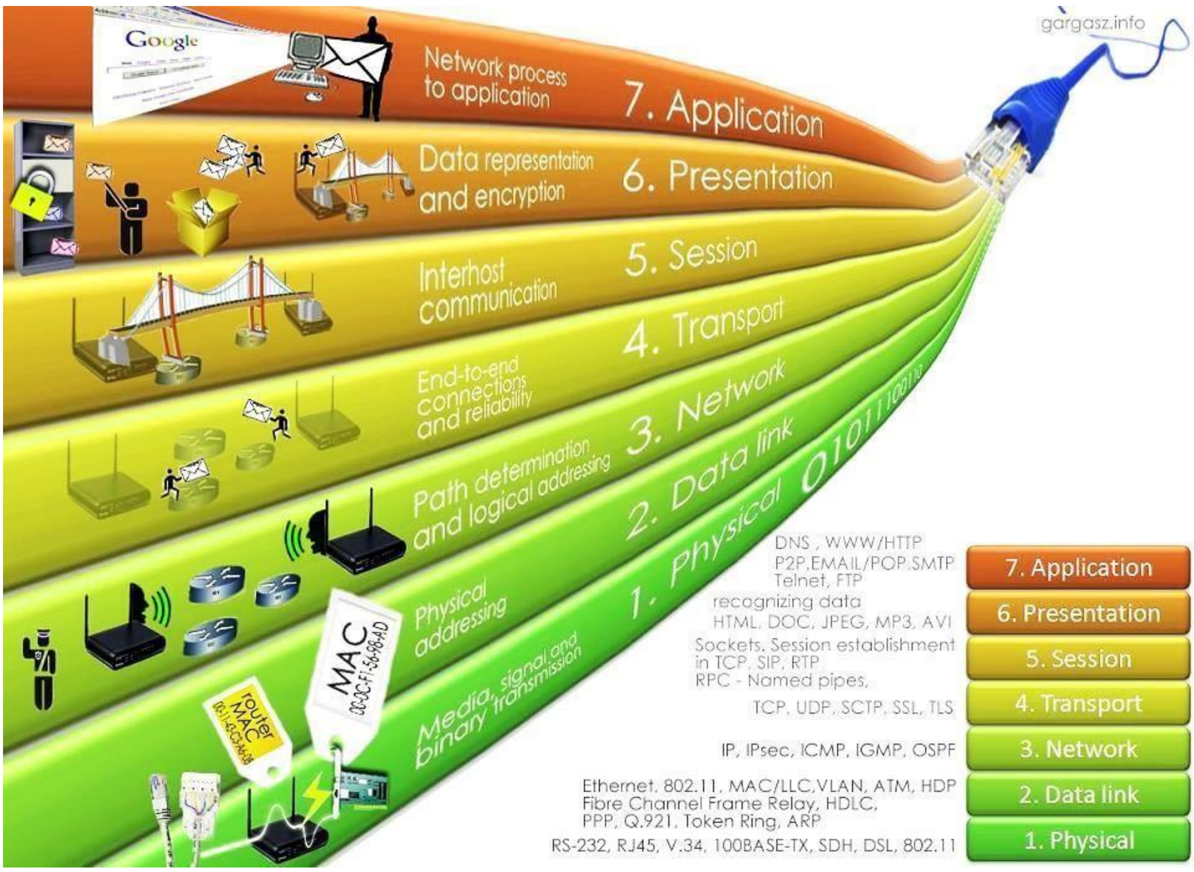
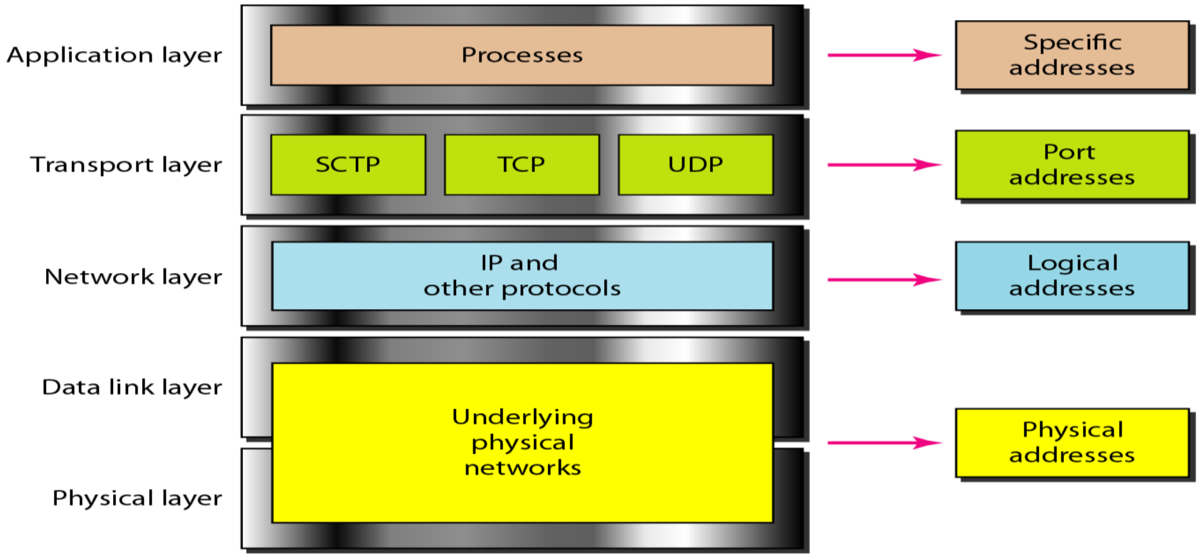


Figure 2.11 Summary of OSI layers



هذه الصفحة للاطلاع فقط

Chapter 3: TCP/IP Protocol Suite and addressing



3.1 TCP/IP Protocol Suite

The TCP/IP protocol suite is made of five layers: physical, data link, network, transport, and application. The first four layers provide physical standards, network interfaces, internetworking, and transport functions that correspond to the first four layers of the OSI model (see Figure 3.1). The three topmost layers in the OSI model, however, are represented in TCP/IP by a single layer called the *application layer*.

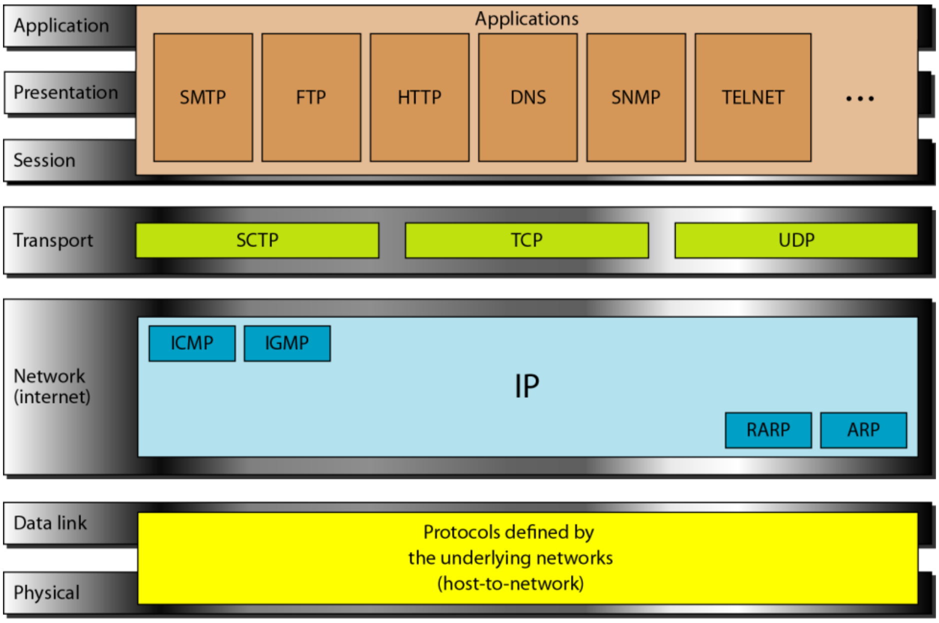


Figure 3.1 TCP/IP and OSI Model

1. Physical and Data Link Layers

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

2. Network Layer

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

The **Internetworking Protocol (IP)** is the transmission mechanism used by the TCP/IP protocols. IP transports data in packets called ***datagrams***, each of which is transported separately. Datagrams can travel along different routes and can arrive out of sequence or be duplicated.

The **Address Resolution Protocol (ARP)** is used to associate a logical address with a physical address. ARP is used to find the physical address of the node when its Internet address is known.

The **Reverse Address Resolution Protocol (RARP)** allows a host to discover its Internet address when it knows only its physical address. It is used when a computer is connected to a network for **the first time**.

The **Internet Control Message Protocol (ICMP)** is a mechanism used by hosts and gateways to send notification of datagram problems back to the sender.

The **Internet Group Message Protocol (IGMP)** is used to facilitate the simultaneous transmission of a message to a group of recipients.

3.Transport Layer

The main protocols in transport layer are the TCP and UDP. IP protocol in network layer is a **source-to-destination** protocol, meaning that it can deliver a packet from one physical device to another. Whereas, UDP and TCP are transport level protocols responsible for delivery of a message from a process (running program) to another process (**process-to-process** protocols).

The **User Datagram Protocol (UDP)** is the simpler of the two standard TCP/IP transport protocols. It is a process-to-process protocol that adds only **port addresses, checksum error control, and length information** to the data from the upper layer.

The **Transmission Control Protocol (TCP)** is a **reliable stream** transport protocol. The term ***stream***, in this context, means **connection-oriented**: A connection must be established between both ends of a transmission before either can transmit data.

At the sending end of each transmission, TCP divides a stream of data into smaller units called ***segments***. Each segment includes a sequence number for reordering after receipt. At the receiving end, TCP collects each datagram as it comes in and reorders the transmission based on sequence numbers.

The Stream Control Transmission Protocol (SCTP) provides support for newer applications such as voice over the Internet.

4. Application Layer

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

3.2 Addressing

Four levels of addresses are used in an internet employing the TCP/IP protocols: **physical (link) addresses**, **logical (IP) addresses**, **port addresses**, and **specific addresses** (see Figure 3.2).

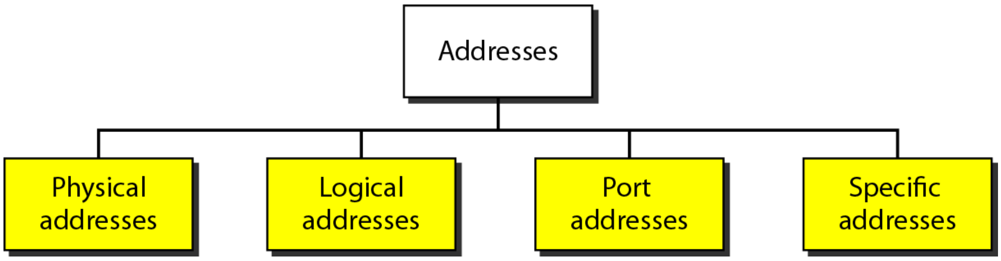


Figure 3.2 Addressing in TCP/IP

Each address is related to a specific layer in the TCP/IP architecture, as shown in Figure 3.3.

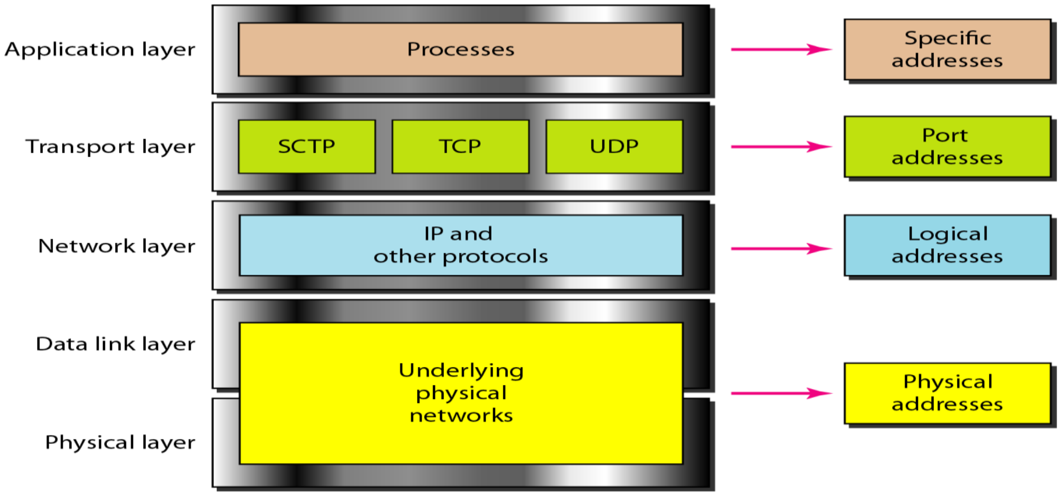


Figure 3.3 Relationship of layers and addresses in TCP/IP

3.2.1 Physical Addresses

The physical address, also known as the link address, is the address of a node as defined by its LAN or WAN. It is included in the frame used by the data link layer. It is the lowest-level address.

Example 1

In Figure 3.4 a node with physical address 10 sends a frame to a node with physical address 87. The two nodes are connected by a link (bus topology LAN). At the data link layer, this frame contains physical (link) addresses in the header. These are the only addresses needed. The data link layer at the sender receives data from an upper layer. It encapsulates the data in a frame, adding a header and a trailer.

Encapsulation means that a packet (header, data and maybe trailer) at a specific level is encapsulated in one whole packet.

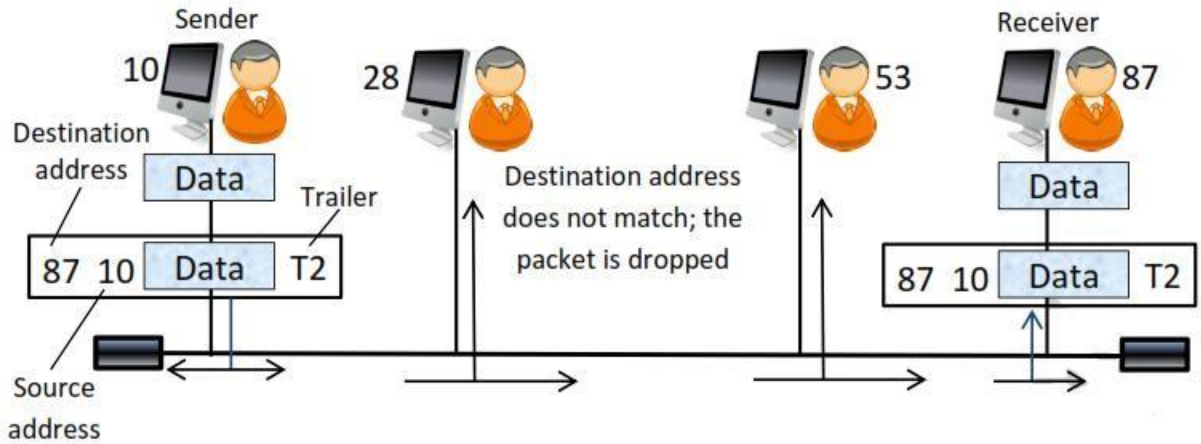


Figure 3.4 Example of physical addresses

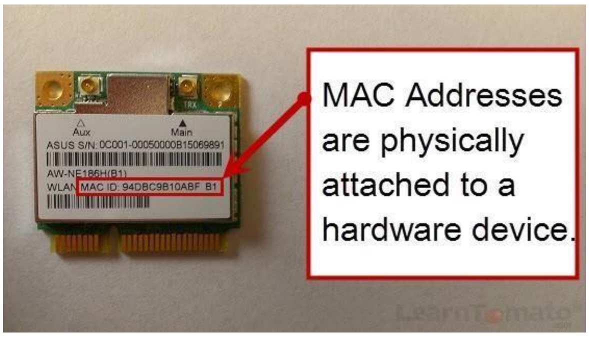
Example 2

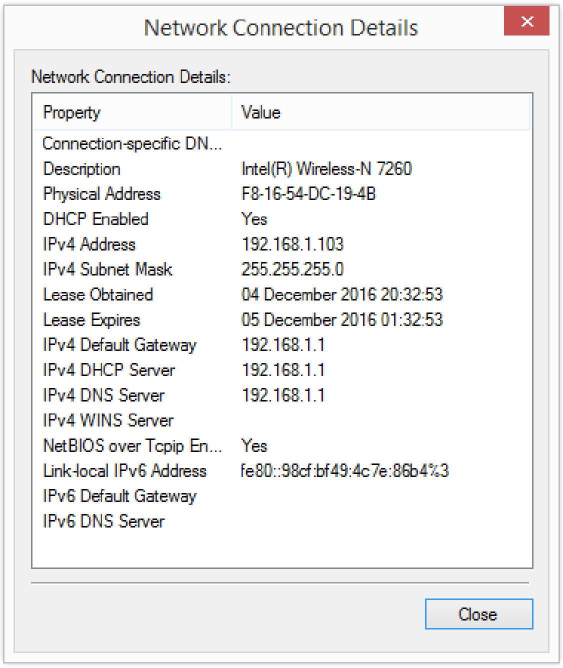
Most local-area networks use a **48-bit (6-byte)** physical address written as **12 hexadecimal digits**; every byte (2 hexadecimal digits) is separated by a colon, as shown below:

07:01:02:01:2C:4B

A 6-byte (12 hexadecimal digits) physical address

The physical address is imprinted on the network interface card (NIC) as shown in the left picture and can be displayed on any computer as shown in the right screenshot.





A screenshot that shows a physical address and other connection details

3.2.2 Logical Addresses

Physical addresses are not enough in an internetwork environment. A universal (or logical) addressing system is needed in which each host can be identified uniquely, regardless of the underlying physical network.

A logical address in the Internet is currently a 32-bit address that can uniquely define a host connected to the Internet. No two publicly addressed hosts on the Internet can have the same IP address.

192.168.1.100

IP (or logical) address, Version 4, Class C

Example 3

Figure 3.5 shows a part of an internet with two routers connecting three LANs. Each device (computer or router) has a pair of addresses (logical and physical). The computer with **logical address A** and **physical address 10** needs to send a packet to the computer with **logical address P** and **physical address 95**.

The sender encapsulates its data in a packet at the network layer and adds two logical addresses (A and P). The network layer, however, needs to find the physical address of the next hop before the packet can be delivered. The network layer consults its **routing table** and finds the logical address of the next hop (router 1) to be F. The **ARP** finds the physical address of router 1 that corresponds to the logical address of 20. Now the network layer passes this address to the data link layer, which in turn encapsulates the packet with physical destination address 20 and physical source address 10.

Since the logical destination address does not match the router's logical address, the router 1 knows that the packet needs to be forwarded to router 2. When the frame reaches the destination, the packet is de-capsulated. The destination logical address P matches the logical address of the computer. The data are de-capsulated from the packet and delivered to the upper layer.

Note: Although physical addresses will **change from hop to hop**, logical addresses **remain the same** from the source to destination (but there are some exceptions to this rule).

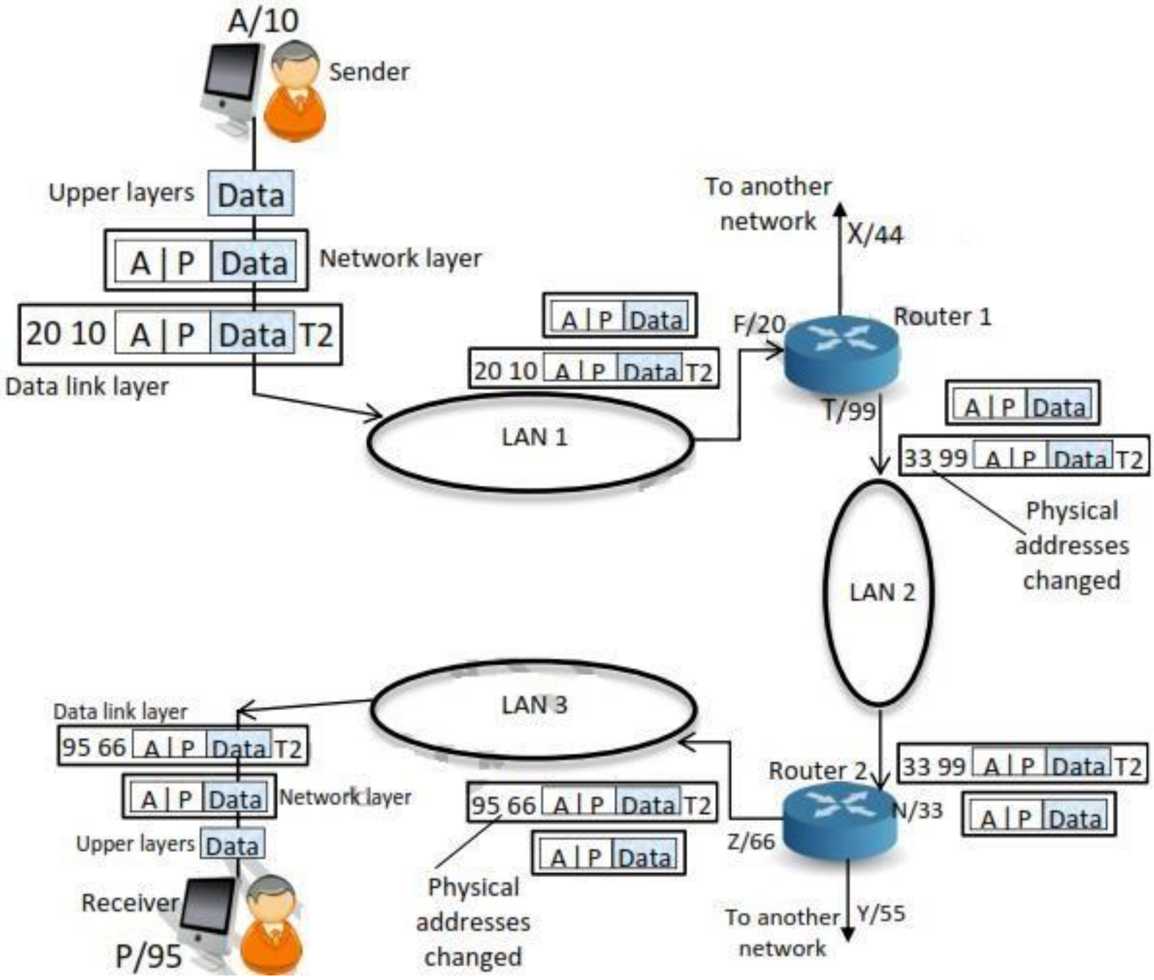


Figure 3.5 Example of logical (IP) addresses

3.2.3 Port Addresses

Today, computers are devices that can run multiple processes at the same time. The end objective of Internet communication is a process communicating with another process. Therefore, we need a method to label the different processes. In the TCP/IP architecture, the label assigned to a process is called a **port address**. A port address in TCP/IP is 16 bits in length.

Example 4

Figure 3.6 shows two computers communicating via the Internet. The sending computer is running three processes at this time with port addresses (a), (b), and (c). The receiving computer is running two processes at this time with port addresses (j) and (k). Process (a) in the sending computer needs to communicate with process (j) in the receiving computer. Note that although both computers are using the same application, FTP, for example, the port addresses are different **because** one is a **client program** and the other is a **server program.**

Note: Although **physical addresses** change from hop to hop, **logical and port addresses** remain the same from the source to destination (there are some exceptions to this rule).

3.2.4 Specific Addresses

Some applications have user-friendly addresses that are designed for that specific address, for example:

- The e-mail address (for example, someone@gmail.com) that defines the recipient of an e-mail.
- Universal Resource Locator (**URL**) (for example, www.google.com) that is used to find a document on the *World Wide Web* (WWW).

The e-mail and URL addresses **are changed** automatically to the corresponding **port** and **logical addresses** by the sending computer (**by using the DNS**: *Domain Name System*).

For example, Figure 3.6 shows a screenshot of the command prompt window in which the 172.217.17.196 is the logical address of the URL address: www.google.com.

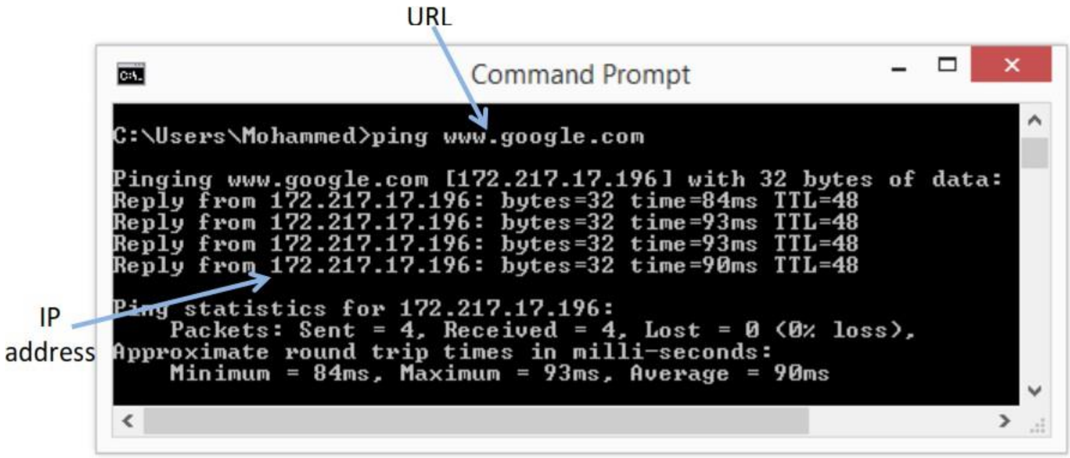
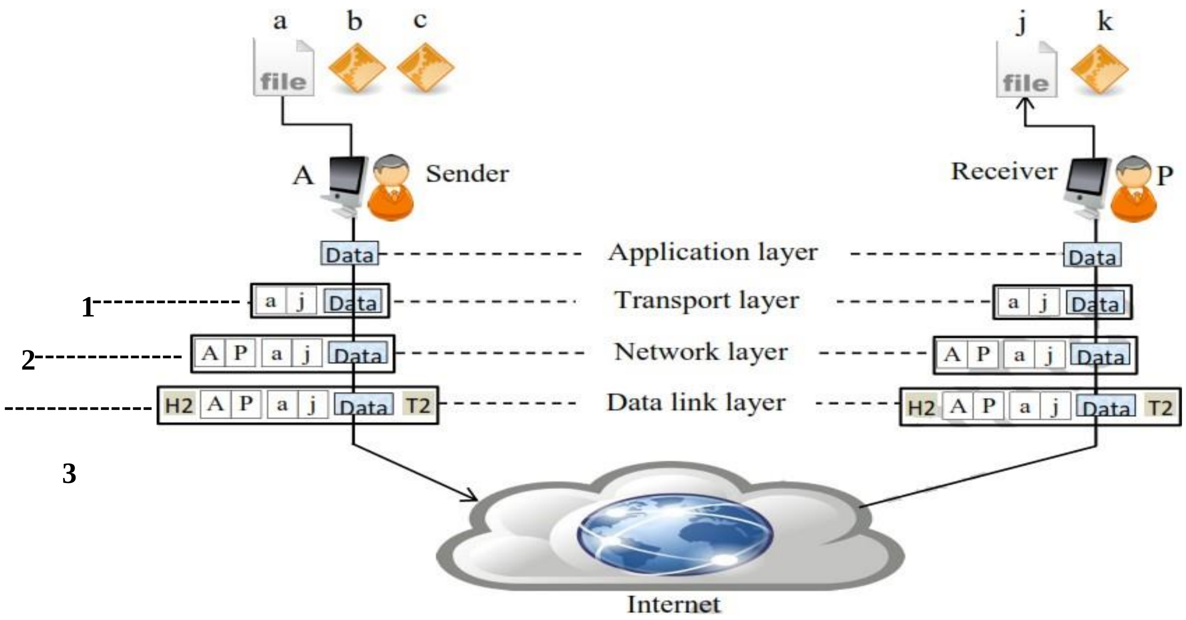


Figure 3.6 Using **ping** instruction to know the corresponding IP address of ww.google.com



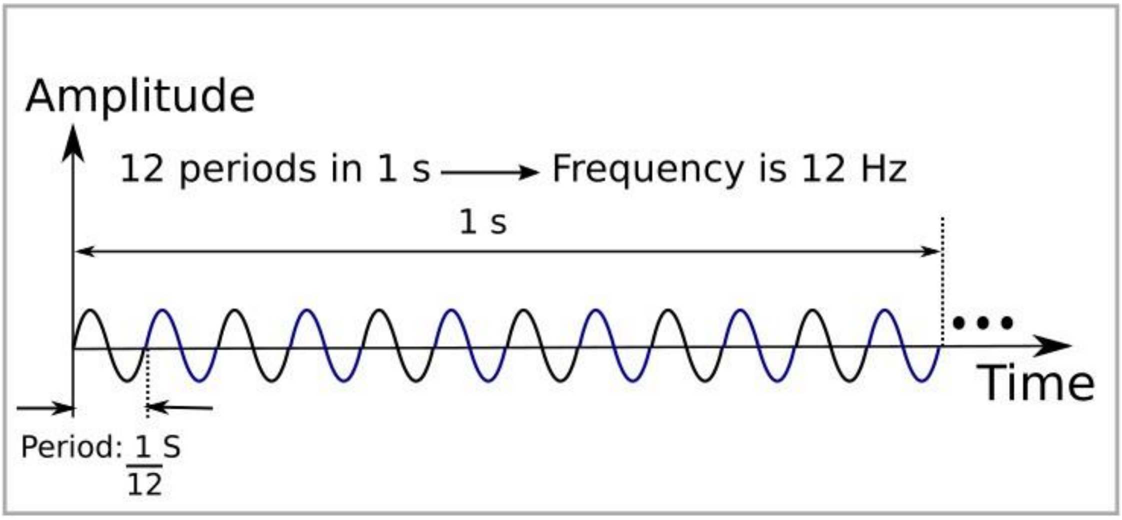
According to the function of each layer, **the encapsulation operation** of the above figure is included the following:

1. The transport layer **encapsulates** data from the application layer in a **packet**
and adds two port addresses (a) and (j), source and destination.

2. The packet from the transport layer is then **encapsulated** in another **packet**
at the network layer with logical source and destination addresses (A and P).

3. Finally, the packet is **encapsulated** in a **frame** with the physical source and destination addresses of the next hop.

Chapter 4: Data and Analog Signals



4. Data and Signals

One of the major functions of the *physical layer* is to move data in the form of electromagnetic signals across a transmission medium. Generally, the data usable to a person or application are not in a form that can be transmitted over a network. For example, a photograph must first be changed to a form that transmission media can accept. Transmission media work by conducting energy along a physical path.

4.1 Analog and Digital Data

Data can be analog or digital. The term analog data refers to information that is continuous; digital data refers to information that has discrete states. **For example: the analog clock and the digital clock.**

Digital data take on discrete values. For example, data are stored in computer memory in the form of 0s and 1s. They can be converted to a digital signal or modulated into an analog signal for transmission across a medium.

Therefore, data can be *analog* or *digital*. **Analog data** are continuous and take continuous values. **Digital data** have discrete states and take discrete values.

4.2 Analog and Digital Signals

Like the data they represent, signals can be either analog or digital. An **analog signal** has infinitely many levels of intensity over a period of time. A **digital signal**, on the other hand, can have only a limited number of defined values. Although each value can be any number, it is often as simple as 1 and 0.

Figure 3.1 illustrates an analog signal and a digital signal. The vertical axis represents the value or strength of a signal. The horizontal axis represents time.

Therefore, Signals can be *analog* or *digital*. Analog signals can have an **infinite number of values** in a range; digital signals can have only a **limited number of values**.

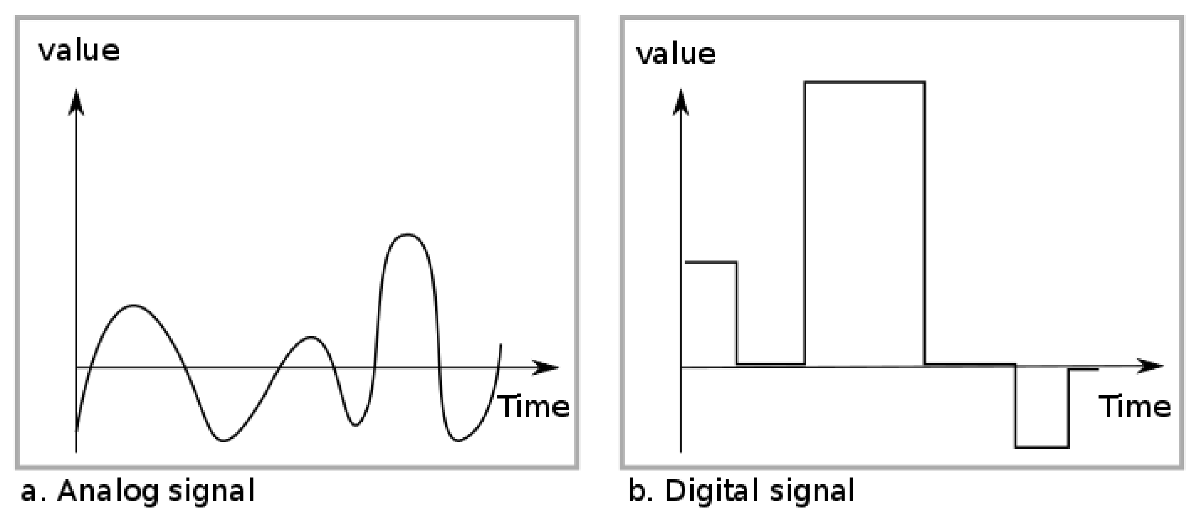


Figure 4.1 Comparison of analog and digital signals

Periodic and Non-periodic Signals

A **periodic signal** completes a pattern within a measurable time frame, called a **period**, and repeats that pattern over identical periods. The completion of one full pattern is called a **cycle**.

A **nonperiodic signal** changes without exhibiting a pattern or cycle that repeats over time. Both analog and digital signals can be periodic or nonperiodic.

In data communications, we commonly use ***periodic analog signals*** (because they need less bandwidth) and ***nonperiodic digital signals*** (because they can represent variation in data).

4.3 Periodic analog signals

Periodic analog signals can be classified as simple or composite.

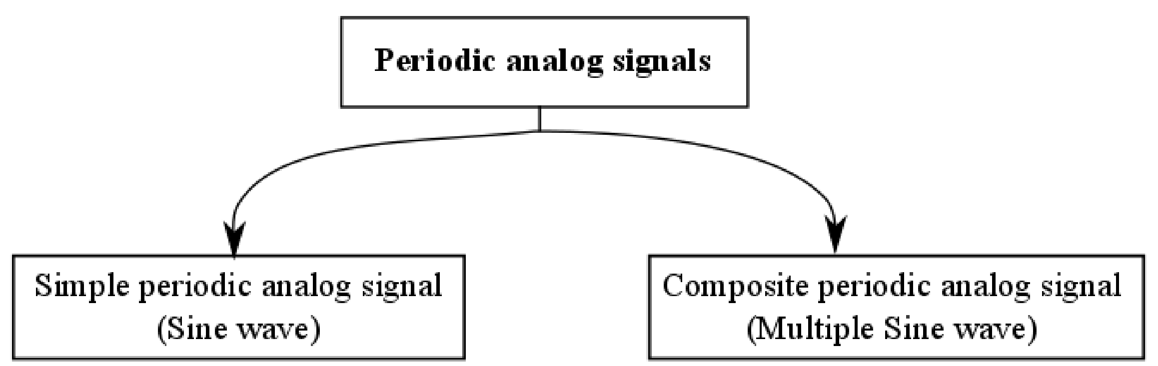


Figure 4.2 Classification of periodic analog signals

4.3.1 Sine Wave

Figure 4.3 shows a sine wave. Each cycle consists of a single arc above the time axis followed by a single arc below it. A sine wave can be represented by **three parameters**: the 1) *peak amplitude,* 2) the *frequency,* and 3) the *phase.*

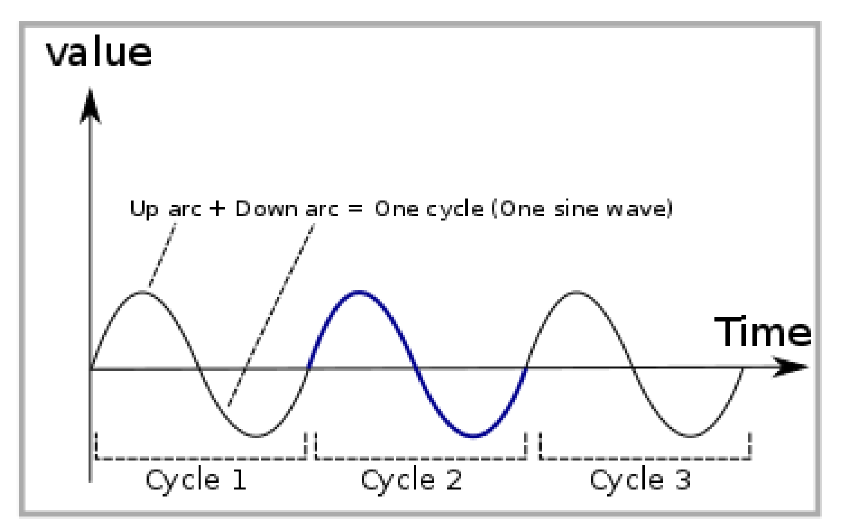


Figure 4.3 Sine wave

1) <u>Peak amplitude</u>

The peak amplitude of a signal is the absolute **value** of its highest intensity (the energy). Figure 4.4 shows two signals and their peak amplitudes. For electric signals, peak amplitude is normally measured in ***volts***. The power in your house can be represented by a sine wave with a peak amplitude of 220 to 240 V, whereas the peak value of an AA battery is normally 1.5 V.

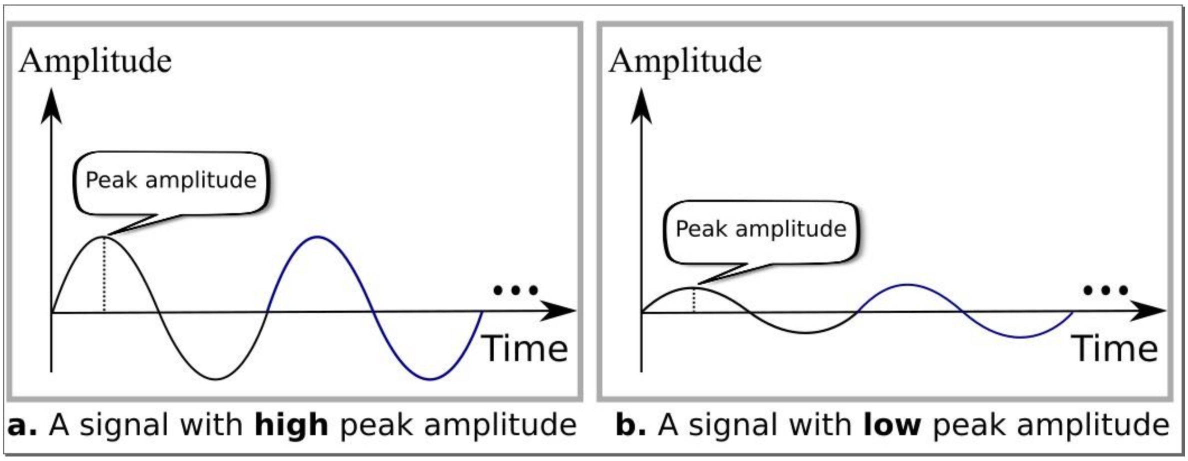


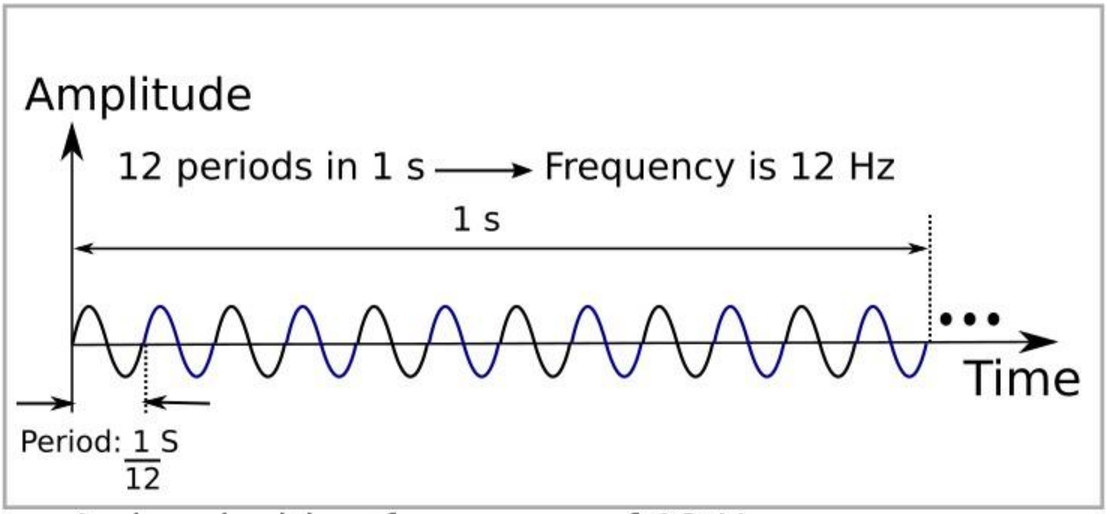
Figure 4.4 Two signals with the same phase and frequency, but different amplitudes

2) Period and Frequency

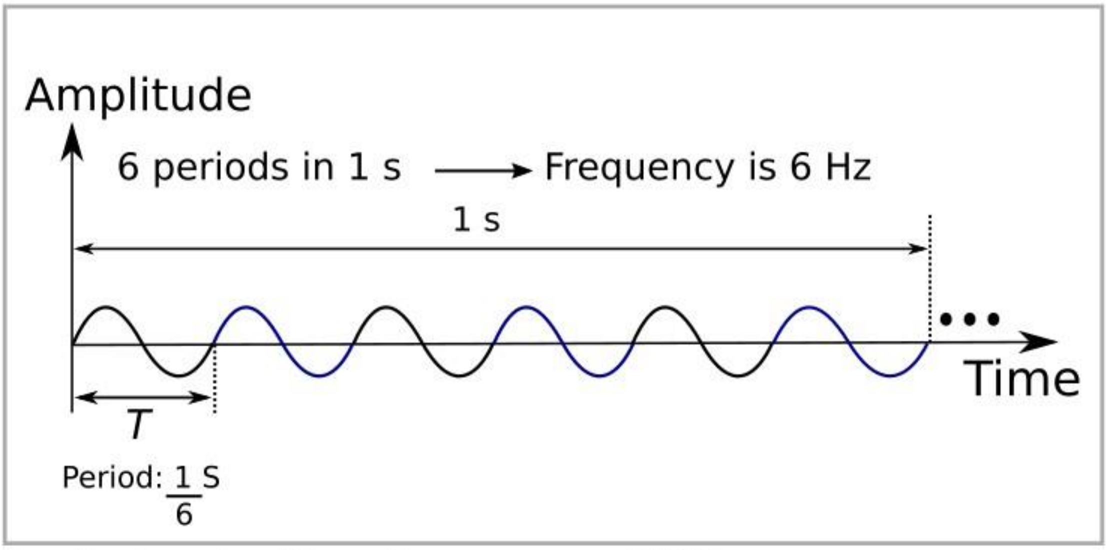
Period refers to the amount of time, in seconds, a signal needs to complete 1 cycle. **Frequency refers** to the number of periods in 1 s. Note that period and frequency are just one characteristic defined in two ways. Period is the inverse of frequency, and frequency is the inverse of period, as the following formulas show.

$$f = \frac{1}{T} \text{ and } T = \frac{1}{f} \quad (f \text{ :frequency } T \text{ : time })$$

Figure 4.5 shows two signals and their frequencies.



a. A signal with a frequency of 12 Hz



b. A signal with a frequency of 6 Hz

Figure 4.5 Two signals with the same amplitude and phase, but different frequencies

Period is formally expressed in seconds. Frequency is formally expressed in Hertz (Hz), which is cycle per second. Units of period and frequency are shown in Table 4.1.

Table 4.1 Units of period and frequency

Unit	*Equivalent*
Seconds (s)	1s
Milliseconds (ms)	10^{-3} s
Microseconds (μs)	10^{-6} s
Nanoseconds (ns)	10^{-9} s
Picoseconds (ps)	10^{-12} s

Units of period

Unit	*Equivalent*
Hertz (Hz)	1 Hz
Kilohertz (kHz)	10^{3} Hz
Megahertz (MHz)	10^{6} Hz
Gigahertz (GHz)	10^{9} Hz
Terahertz (THz)	10^{12} Hz

Units of frequency

Example 1

The power we use at home has a frequency of 60 Hz. The period of this sine wave can be determined as follows:

$$T \frac{1}{f} = \frac{1}{60} = 0.0166 \text{ s} = 0.0166 \text{ x } 10^{3} \text{ ms} = 16.6 \text{ ms}$$

This means that the period of the power for our lights at home is 0.0116 s, or 16.6 ms. Our eyes are not sensitive enough to distinguish these rapid changes in amplitude.

Example 2

Express a period of 100 ms in microseconds.

Solution

From Table 4.1 we find the equivalents of 1 ms (l ms is 10^{-3} s) and (1 s is 10^{6} μs). We make the following substitutions:

$$100 \text{ ms} = 100 \text{ x } 10^{-3} \text{ x } 10^{6} \text{ μs} = 10^{2} \text{ x } 10^{-3} \text{ x } 10^{6} \text{ μs} = 10^{5} \text{ μs}$$

لتحويل الوقت من وحدة الملي ثانية (milliseconds) الى وحدة الثانية (second)

لتحويل الوقت من وحدة الثانية (second) الى وحدة المايكرو ثانية (Microseconds)

عند الضرب تُجمع الأسس

Example 3

The period of a signal is 100 ms. What is its frequency in kilohertz?

Solution

First we change 100 ms to seconds, and then we calculate the frequency from the period (1 Hz = 10^{-3} kHz).

$$100 \text{ ms} = 100 \times 10^{-3} \text{ s} = 10^{-1} \text{ s}$$

$$f = \frac{1}{T} = \frac{1}{10^{-1}} \text{ Hz} = 10 \text{ Hz} = 10 \times 10^{-3} \text{ kHz} = 10^{-2} \text{ kHz}$$

3) Phase

The term phase describes the position of the waveform relative to time 0. It indicates the status of the first cycle (how much the wave is shifted from 0 on the time axis).

Looking at Figure 4.6, we can say that

1. A sine wave with a phase of 0° starts at time 0 with a zero amplitude. ***The amplitude is increasing.***
2. A sine wave with a phase of 90° starts at time 0 with a peak amplitude. ***The amplitude is decreasing.***
3. A sine wave with a phase of 180° starts at time 0 with a zero amplitude. ***The amplitude is decreasing.***

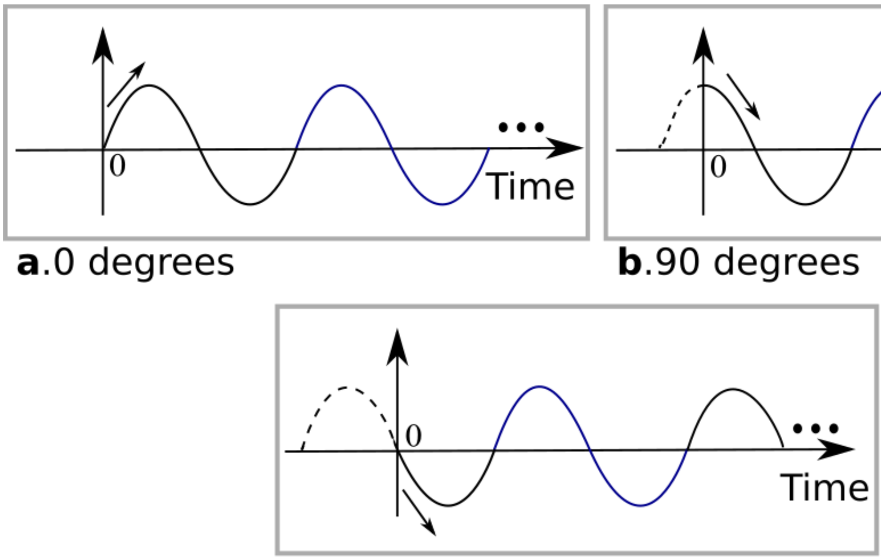


Figure 4.6 Three sine waves with the same amplitude and frequency, but different phases

4.3.2 Composite Signals

The previous section focused on simple sine waves which have many applications in daily life. We can send a single sine wave to carry electric energy from one place to another. For example, the power company sends a single sine wave with a frequency of 50 Hz to distribute electric energy to houses and businesses.

If we had only one single sine wave to convey a conversation over the phone, it would make no sense and carry no information. We would just hear a buzz. **Therefore,** we need to send a composite signal to communicate data; a single-frequency sine wave is not useful in data communications.

The **composite signal** is a combination of simple sine waves with different frequencies, amplitudes, and phases.

A composite signal can be periodic or nonperiodic as shown in the following:

- **A periodic composite signal** can be decomposed into a series of simple sine waves with *discrete frequencies* (the frequencies that have integer values: 1, 2, 3, and so on).
- **A nonperiodic composite signal** can be decomposed into a combination of an infinite number of simple sine waves with continuous frequencies (the frequencies that have real values: 0.1, 0.2, 0.3, and so on).

4.3.3 Bandwidth

Bandwidth is the range of frequencies contained in a composite signal. The bandwidth is normally a difference between two numbers. For example, if a composite signal contains frequencies between 1000 and 5000, its bandwidth is 5000 - 1000, or 4000.

Figure 4.8 shows the concept of bandwidth. The bandwidth of the periodic signal contains all integer frequencies between 1000 and 5000 (1000, 1001, 1002, ...). The bandwidth of the nonperiodic signals has the same range, but the frequencies are **continuous**.

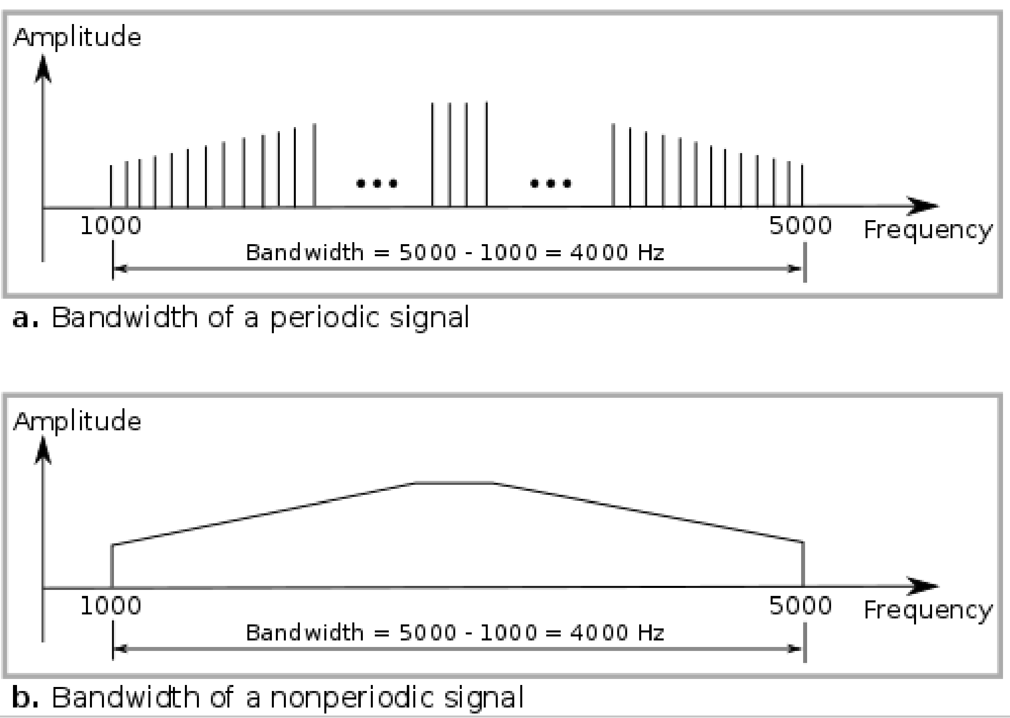


Figure 4.8 The bandwidth of periodic and nonperiodic composite signals

Example 4

If a periodic signal is decomposed into five sine waves with frequencies of 100, 300, 500, 700, and 900 Hz, what is its bandwidth? Draw the spectrum, assuming all components have **a maximum amplitude of 10 V**.

Solution

Let f_h be the highest frequency, f_l the lowest frequency, and B the bandwidth.
Then: $B = f_h - f_l = 900 - 100 = 800$ Hz

The spectrum has only five spikes, at 100, 300, 500, 700, and 900 Hz (see Figure 4.9).

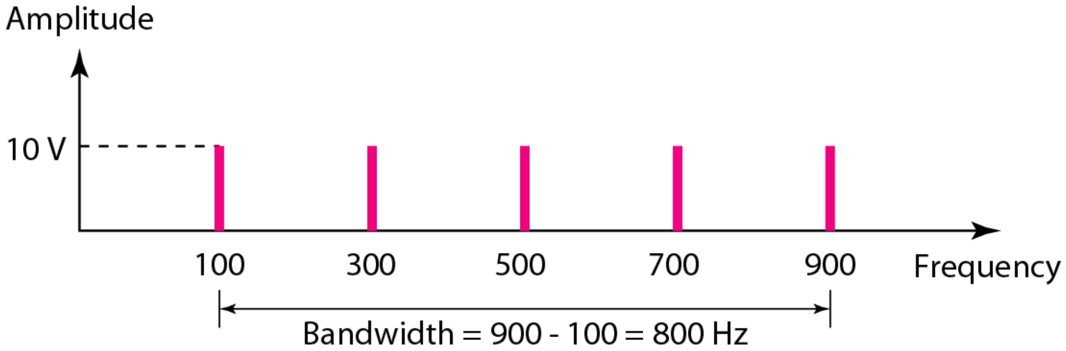


Figure 4.9 The bandwidth for Example 4

Example 5

A periodic signal has a bandwidth of 20 Hz. The highest frequency is 60 Hz. What is the lowest frequency? Draw the spectrum if the signal contains all frequencies of the **same amplitude**.

Solution

Let fh be the highest frequency, fl the lowest frequency, and B the bandwidth.

Then: $B = fh - fl \quad \rightarrow 20 = 60 - fl \quad \rightarrow fl = 60 - 20 = 40$ Hz

The spectrum contains all integer frequencies. We show this by a series of spikes (see Figure 4.10).

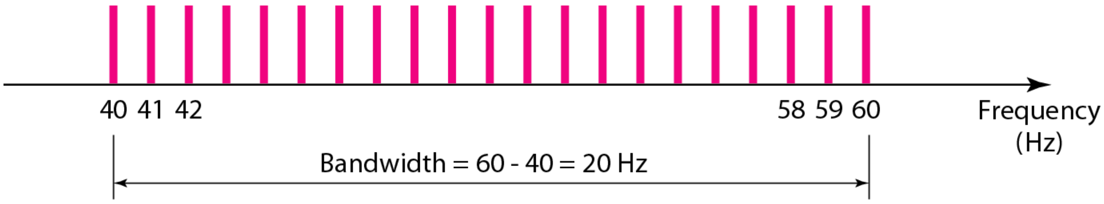
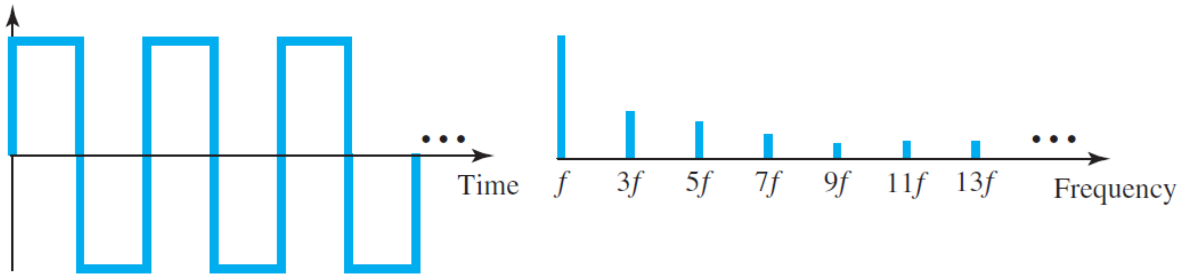


Figure 4.10 The bandwidth for Example 5

Chapter 5: Digital Signals & Transmission Impairment



5. Digital Signals

In addition to being represented by an analog signal, information can also be represented by a digital signal. For example, a 1 can be encoded as a positive voltage and a 0 as zero voltage. A digital signal can have more than two levels. In this case, we can send more than 1 bit for each level. Figure 5.1 shows two signals, one with two levels and the other with four. We send 1 bit per level in part a of the figure and 2 bits per level in part b of the figure.

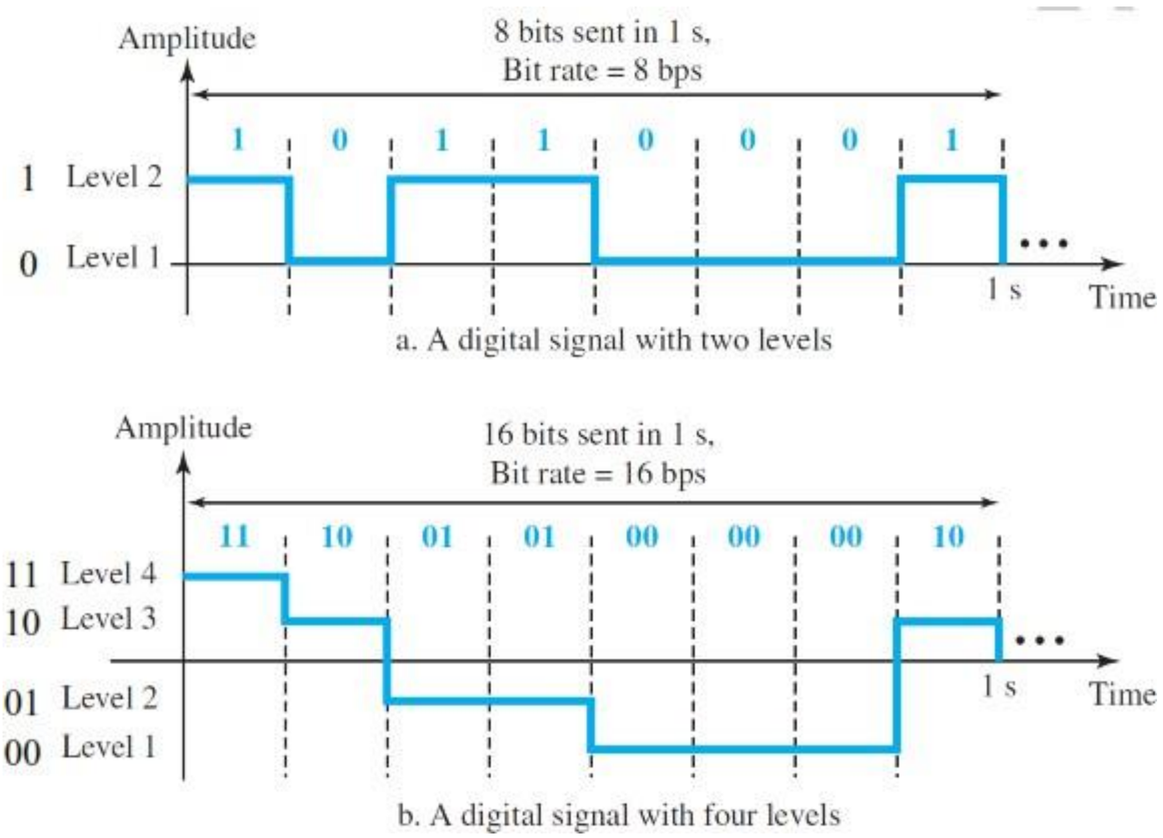


Figure 5.1 Two digital signals: one with two signal levels and the other with four signal levels

5.1 Bit Rate

Most digital signals are nonperiodic, and thus period and frequency are not appropriate characteristics. Another term—***bit rate*** (instead of *frequency*)—is **used to describe digital signals**. The **bit rate** is the number of bits sent in 1s, expressed in bits per second (bps). Figure 5.1 shows the bit rate for two signals.

Example 5.1

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

Digital data take on discrete values. For example, data are stored in computer memory in the form of 0s and 1s. They can be converted to a digital signal or modulated into an analog signal for transmission across a medium.

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 x 24 x 80 x 8 = 1,536,000 bps = 1.536 Mbps ///////

Example 5.2

A digitized voice channel 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 x 4000 x 8 = 64,000 bps = 64 kbps ///////

Example 5.3

What is the bit rate for high-definition TV (HDTV)?

Solution

HDTV uses digital signals to broadcast high quality video signals. The HDTV screen is normally a ratio of 16:9 (in contrast to 4:3 for regular TV), which means the screen is wider. There are 1920 by 1080 pixels per screen, and the screen is renewed 30 times per second. Twenty-four bits represents one colour pixel. We can calculate the bit rate as

1920 x 1080 x 30 x 24 = 1,492,992,000 ≈ 1.5 Gbps

The TV stations reduce this rate to 20 to 40 Mbps through compression.

5.2 Digital Signal as a Composite Analog Signal

Based on Fourier analysis, a **digital signal** is a ***composite analog signal*** with an infinite bandwidth. A digital signal, in the time domain, comprises connected vertical and horizontal line segments as shown in Figure 5.2.

- A vertical line in the time domain means a frequency of infinity (sudden change in time);
- A horizontal line in the time domain means a frequency of zero (no change in time).

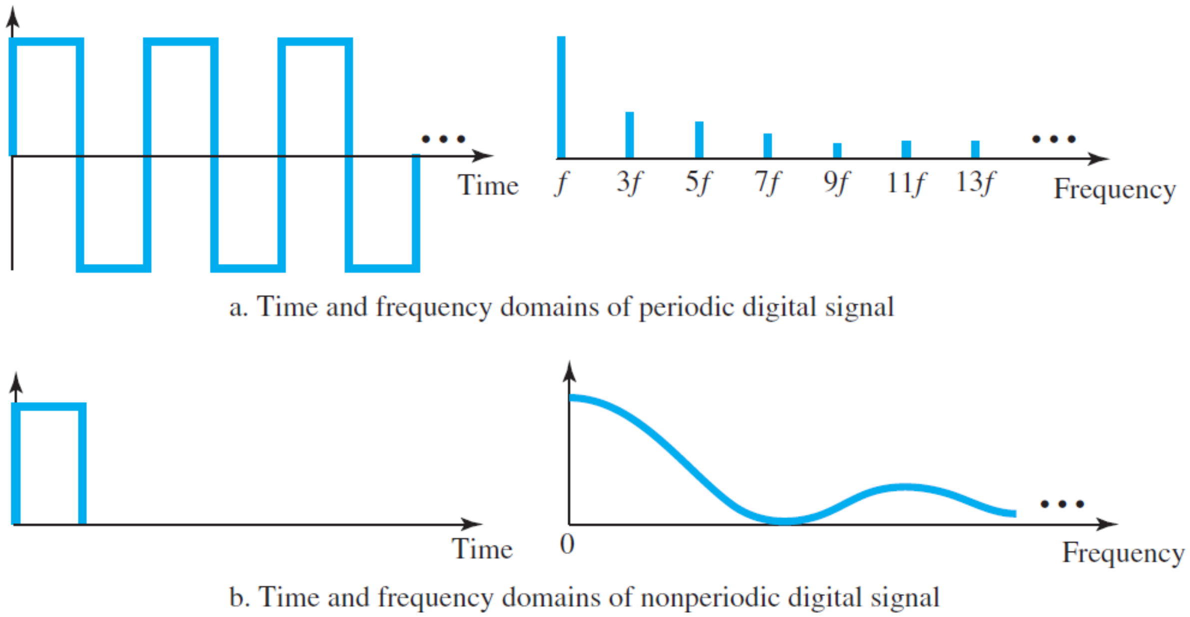
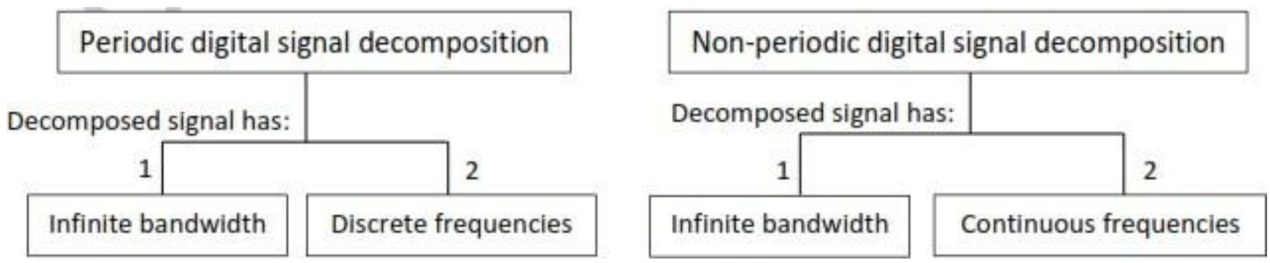


Figure 5.2 The time and frequency domains of periodic and nonperiodic digital signals

Fourier analysis can be used to decompose a digital signal as shown in the following chart:



Note that both bandwidths are infinite, **but** the periodic signal has **discrete frequencies** while the nonperiodic signal has **continuous frequencies**.

5.3 Transmission of Digital Signals

We can transmit a digital signal from point *A* to point *B* by using one of two different approaches: ***baseband transmission*** or ***broadband transmission*** (using modulation).

For the remainder of this lecture, let us consider the case of a nonperiodic digital signal because it is used often in data communications.

5.3.1 Baseband Transmission

Baseband transmission means sending a digital signal over a channel without changing the digital signal to an analog signal. Figure 5.3 shows baseband transmission.

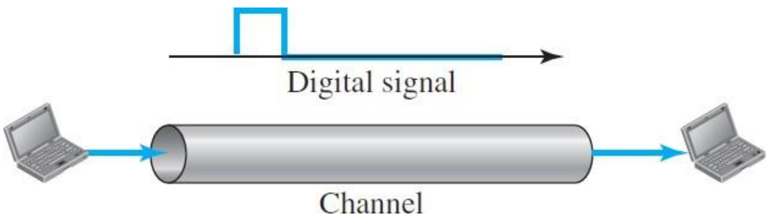


Figure 5.3 Baseband transmission

Baseband transmission requires a **low-pass channel**, a channel with a bandwidth that starts from zero. This is the case if we have a **dedicated medium** with a bandwidth constituting only one channel. **For example**, the entire bandwidth of a cable connecting two computers is one single channel. As **another example**, we may connect several computers to a bus, but not allow more than two stations to communicate at a time.

5.3.2 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**—a channel with a bandwidth that does not start from zero. This type of channel is **more available** than a low-pass channel. Figure 5.4 shows a bandpass channel.

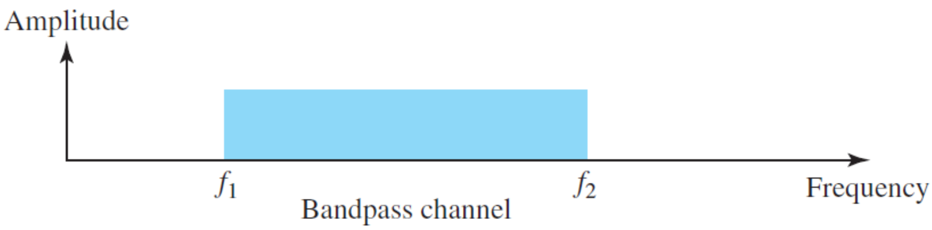


Figure 5.4 Bandwidth of a bandpass channel

Figure 5.5 **shows the modulation of a digital signal**. In the figure, a digital signal is converted to a composite analog signal. We have used a single-frequency analog signal (called a ***carrier***); the amplitude of the carrier has been changed to look like the digital signal. The result, however, is not a single-frequency signal; it is a composite signal. At the receiver, the received analog signal is converted to digital, and the result is a replica of what has been sent.

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.

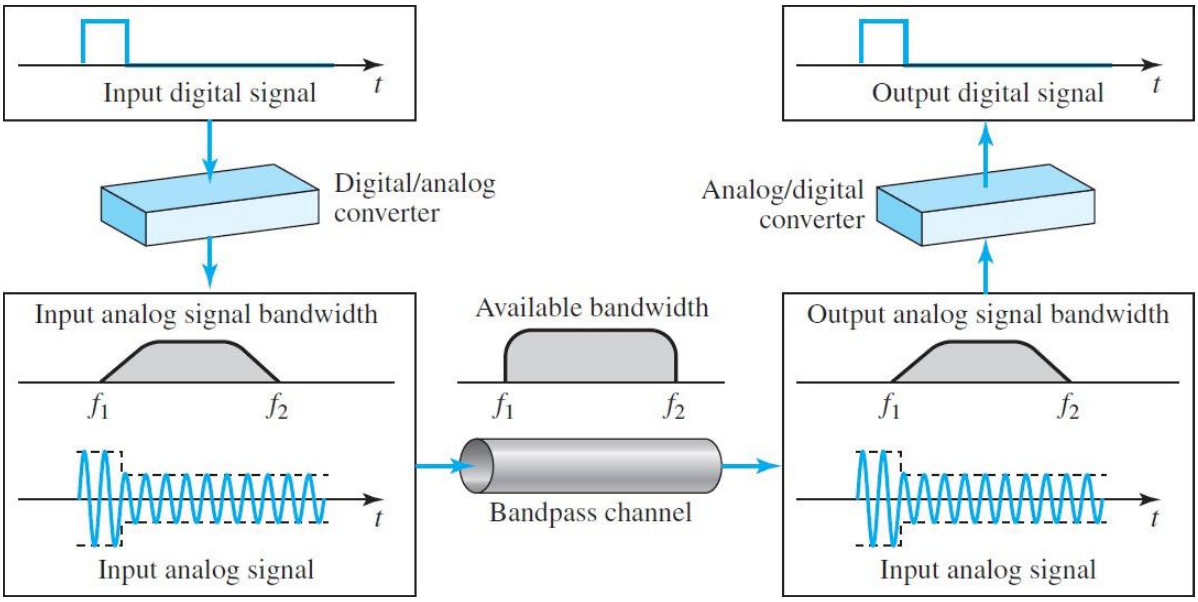


Figure 5.5 Modulation of a digital signal for transmission on a bandpass

<u>Example 5.4</u>

An **example** of broadband transmission using modulation is the sending of computer data through a **telephone lines**. The digital signal in the computer is converted to an analog signal, and then sending the analog signal. At the sending and receiving ends we can install two converters to change the digital signal to analog and vice versa. The converter, in this case, is called a ***modem*** (***mo***dulator/***dem***odulator).

Example 5.5

For better reception, the **digital cellular phones** convert the analog audio signal to digital and then convert it again to analog for transmission over a bandpass channel.

Analog audio signal → digital → analog --Transmission-->

5.4 Transmission Impairment

Signals travel through transmission media, which are not perfect. The imperfection causes signal ***impairment***. This means that the signal at the beginning of the medium is not the same as the signal at the end of the medium (what is sent is not what is received). The **causes** of impairment are ***attenuation, distortion,*** and ***noise*** (see Figure 5.6).

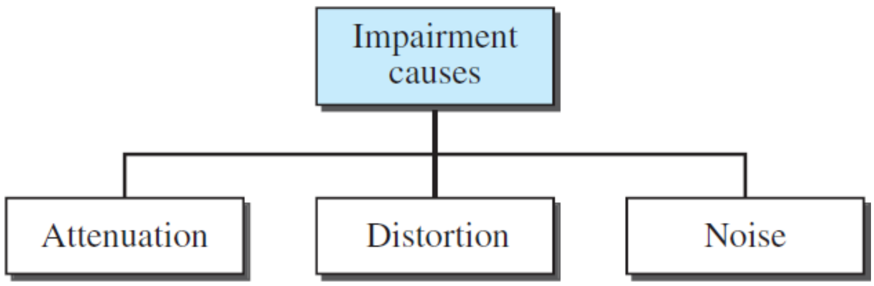


Figure 5.6 Causes of impairment

5.4.1 Attenuation

Attenuation means a loss of energy. When a signal (simple or composite) travels through a medium, it loses some of its energy in overcoming the resistance of the medium. **That is why a wire carrying electric signals gets warm**, if not hot, after a while. Some of the electrical energy in the signal is converted to heat. To recover this loss, **amplifiers** are used to amplify the signal. Figure 5.7 shows the effect of attenuation and amplification.

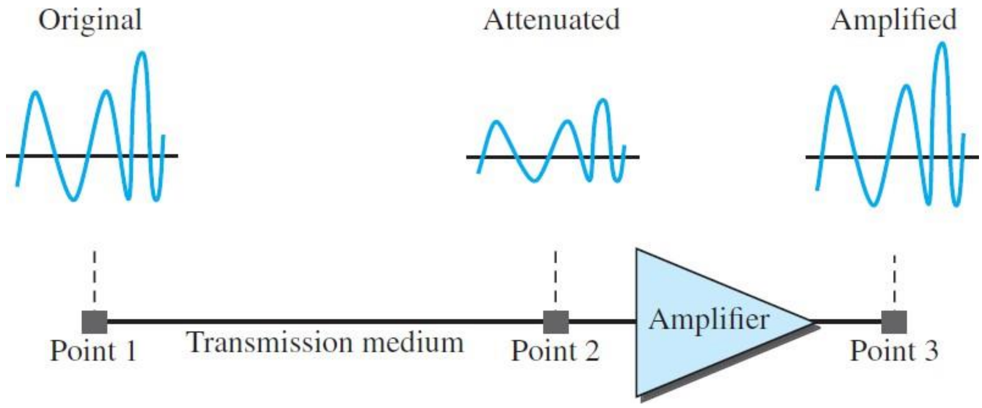


Figure 5.7 Attenuation and amplification

Decibel

To show that a signal has lost or gained strength, engineers use the unit of the decibel. The decibel (**dB**) measures the relative strengths of two signals or one signal at two different points. Note that the decibel is negative if a signal is attenuated and positive if a signal is amplified.

$$dB = 10 \log_{10} \frac{p_2}{p_1}$$

Variables P_1 and P_2 are the powers of a signal at points 1 and 2.

The $\log_{10}$ means the logarithm to base 10, which is called the *common logarithm* or the *decimal logarithm* (اللوغاريتن العام او العشري*).

في الرياضيات، اللوغاريتم هي العملية العكسية للدوال الأسية ويُعرّف لوغاريتم عدد ما بالنسبة لأساس ما، بأنه الأس المرفوع على الأساس والذي سينتج ذلك العدد. فعلى سبيل المثال فلوغاريتم ١٠٠٠ بالنسبة للأساس ١٠ هو ٣ لأن:

$$10^3 = ١٠ \times ١٠ \times ١٠ = ١٠٠٠ \quad (\log_{10} 1000 = 3)$$

* يُعرف اللوغاريتم العام او العشري بأنه لوغاريتم عدد ما بالنسبة للأساس ١٠ والذي يستخدم بشكل كبير في حساب التطبيقات العلمية والهندسية. **في هذه المحاضرة سيتم فقط استخدام اللوغارتم العشري.**

يوجد ايضا انواع اخرى (ليست مستخدمة في هذه المحاضرة) مثلا اللوغاريتم الطبيعي والذي له تطبيقات كثيرة في الحسابات الهندسية والعلمية و في الرياضيات البحتة وخاصة في التفاضل والتكامل. في حين يعرف اللوغاريتم الثنائي لعدد ما بأنه لوغاريتمه بالنسبة للأساس ٢ ويستخدم بشكل كبير في علم الحاسوب والدوال المنطقية مثلا:

The binary logarithm of 4 is 2 ($\log_2 4 = 2$), and the binary logarithm of 32 is 5 ($\log_2 32 = 5$)

Example 5.6

Suppose a signal travels through a transmission medium and its power is reduced to one-half. This means that P2 = 0.5 P1. In this case, the attenuation (loss of power) can be calculated as:

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

Example 5.7

A signal travels through an amplifier, and its power is increased 10 times. This means that P2 = 10P1. In this case, the amplification (gain of power) can be calculated as:

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

5.4.2Distortion

Distortion means that the signal **changes** its form or shape. Distortion can occur in a composite signal made of different frequencies. Each signal component has its own propagation speed through a medium and, therefore, its own delay in arriving at the final destination. Differences in delay may create a difference in phase if the delay is not exactly the same as the period duration (see Figure 5.8).

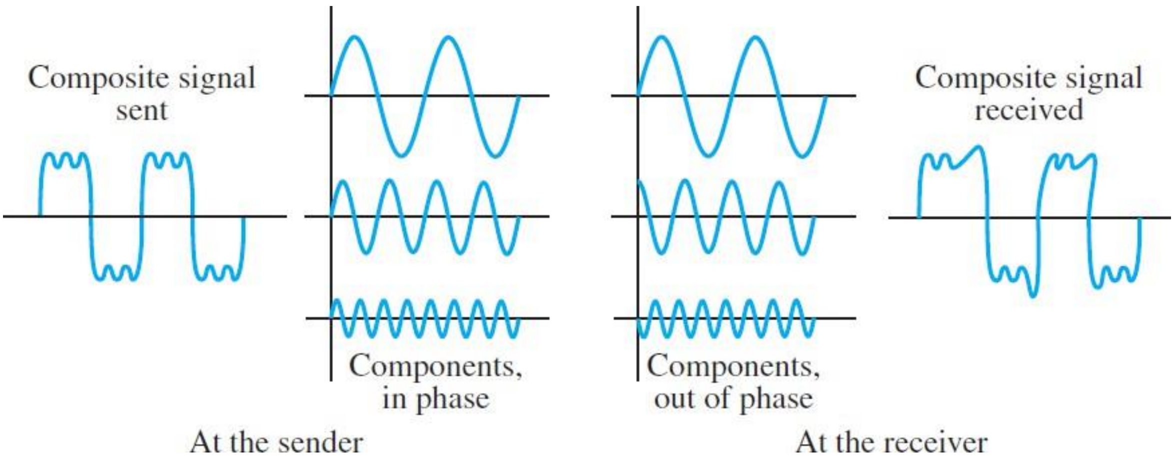


Figure 5.8 Distortion

5.4.3Noise

Noise is another cause of impairment. 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 5.9 shows the effect of noise on a signal.

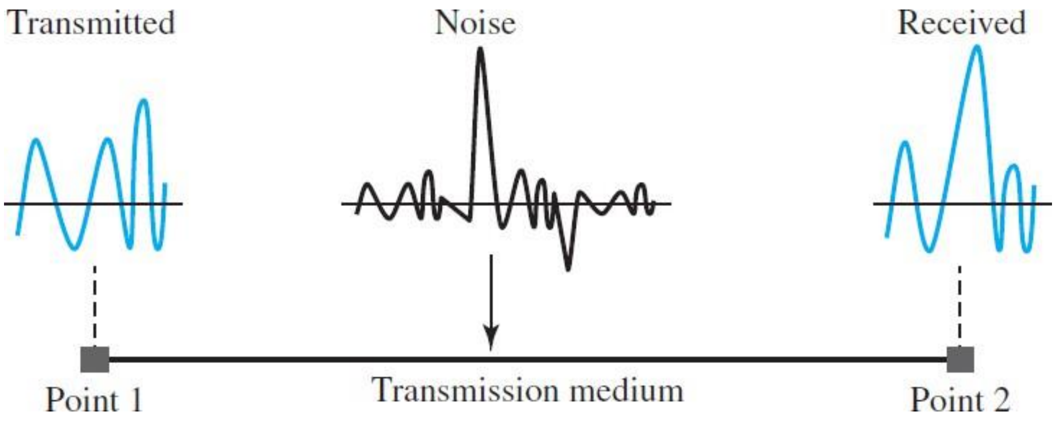


Figure 5.9 Noise

<u>Signal-to-Noise Ratio (SNR)</u>

The signal-to-noise ratio is defined as:

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

We need to consider the average signal power and the average noise power because these may change with time. Figure 5.10 shows the idea of SNR.

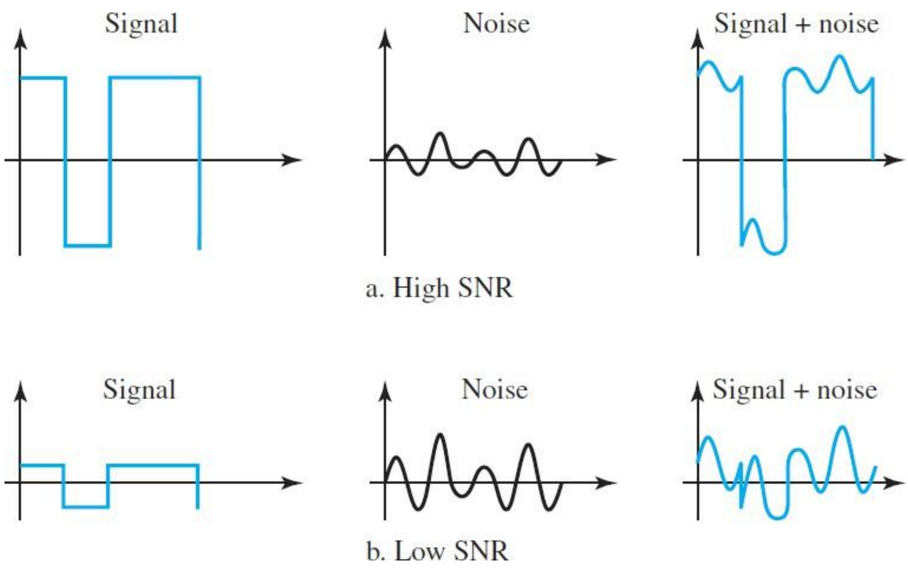


Figure 5.10 Two cases of SNR: a high SNR and a low SNR

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 two powers, it is often described in ***decibel*** units, SNR_{dB}, defined as

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

Example 5.8

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 5.9

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.