Unit 1. Introduction to data communications and networking

1

NETWORKING FUNDAMENTALS

Unit Structure

1.0 OBJECTIVES:

- ✓ Introduce the readers to data communication and its fundamentals
- ✓ Define networks
- ✓ Define protocols
- ✓ Standards in networking

1.1 INTRODUCTION

- This chapter provides an introduction to Computer networks and covers fundamental topics like data, information to the definition of communication and computer networks.
- The main objective of data communication and networking is to enable seamless exchange of data between any two points in the world.
- This exchange of data takes place over a computer network.

1.2 DATA & INFORMATION

- **Data** refers to the raw facts that are collected while **information** refers to processed data that enables us to take decisions.
- Ex. When result of a particular test is declared it contains data of all students, when you find the marks you have scored you have the information that lets you know whether you have passed or failed.
- The word ***data*** refers to any information which is presented in a form that is agreed and accepted upon by is creators and users.

1.3 DATA COMMUNICATION

- Data Communication is a process of exchanging data or information
- In case of computer networks this exchange is done between two devices over a transmission medium.
- This process involves a communication system which is made up of hardware and software. The hardware part involves the sender and receiver devices and the intermediate devices through which the data passes. The software part involves certain rules which specify what is to be communicated, how it is to be communicated and when. It is also called as a Protocol.
- The following sections describes the fundamental characteristics that are important for the effective working of data communication process and is followed by the components that make up a data communications system.

1.3.1 Characteristics of Data Communication

The effectiveness of any data communications system depends upon the following four fundamental characteristics:

1. **Delivery**: The data should be delivered to the correct destination and correct user.
2. **Accuracy**: The communication system should deliver the data accurately, without introducing any errors. The data may get corrupted during transmission affecting the accuracy of the delivered data.
3. **Timeliness**: Audio and Video data has to be delivered in a timely manner without any delay; such a data delivery is called real time transmission of data.
4. **Jitter**: It is the variation in the packet arrival time. Uneven Jitter may affect the timeliness of data being transmitted.

1.3.2 Components of Data Communication

A Data Communication system has five components as shown in the diagram below:

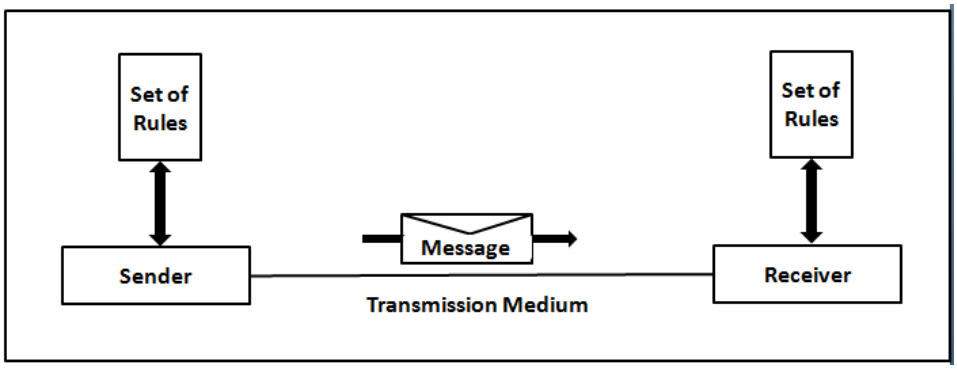


Fig. Components of a Data Communication System

1. Message
Message is the information to be communicated by the sender to the receiver.

2. Sender
The sender is any device that is capable of sending the data (message).

3. Receiver
The receiver is a device that the sender wants to communicate the data (message).

4. Transmission Medium
It is the path by which the message travels from sender to receiver. It can be wired or wireless and many subtypes in both.

5. Protocol

- It is an agreed upon set or rules used by the sender and receiver to communicate data.
- A protocol is a set of rules that governs data communication.
- A Protocol is a necessity in data communications without which the communicating entities are like two persons trying to talk to each other in a different language without know the other language.

1.4 DATA REPRESENTATION

Data is collection of raw facts which is processed to deduce information.

There may be different forms in which data may be represented.
Some of the forms of data used in communications are as follows:

1. Text

- Text includes combination of alphabets in small case as well as upper case.
- It is stored as a pattern of bits. Prevalent encoding system : ASCII, Unicode

2. Numbers

- Numbers include combination of digits from 0 to 9.
- It is stored as a pattern of bits. Prevalent encoding system : ASCII, Unicode

3. Images

- “An image is worth a thousand words” is a very famous saying. In computers images are digitally stored.
- A Pixel is the smallest element of an image. To put it in simple terms, a picture or image is a matrix of pixel elements.
- The pixels are represented in the form of bits. Depending upon the type of image (black n white or color) each pixel would require different number of bits to represent the value of a pixel.
- The size of an image depends upon the number of pixels (also called resolution) and the bit pattern used to indicate the value of each pixel.
- Example: if an image is purely black and white (two color) each pixel can be represented by a value either 0 or 1, so an image made up of 10 x 10 pixel elements would require only 100 bits in memory to be stored.
- On the other hand an image that includes gray may require 2 bits to represent every pixel value (00 - black, 01 – dark gray, 10

– light gray, 11 –white). So the same 10 x 10 pixel image would now require 200 bits of memory to be stored.

- Commonly used Image formats : jpg, png, bmp, etc

4. Audio

- Data can also be in the form of sound which can be recorded and broadcasted. Example: What we hear on the radio is a source of data or information.
- Audio data is continuous, not discrete.

5. Video

- Video refers to broadcasting of data in form of picture or movie

1.5 DATA FLOW

wo devices communicate with each other by sending and receiving data. The data can flow between the two devices in the following ways.

1. Simplex
2. Half Duplex
3. Full Duplex

1.5.1 Simplex

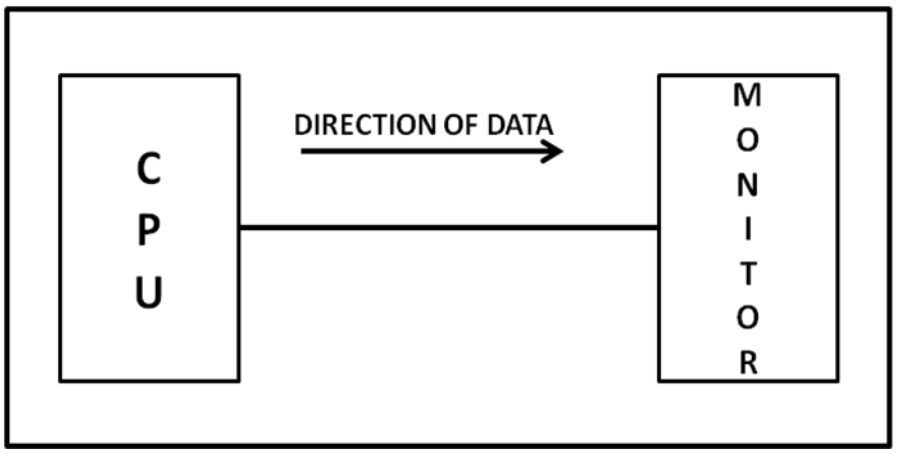


Figure: Simplex mode of communication

- In Simplex, communication is unidirectional
- Only one of the devices sends the data and the other one only receives the data.
- Example: in the above diagram: a cpu send data while a monitor only receives data.

1.5.2 Half Duplex

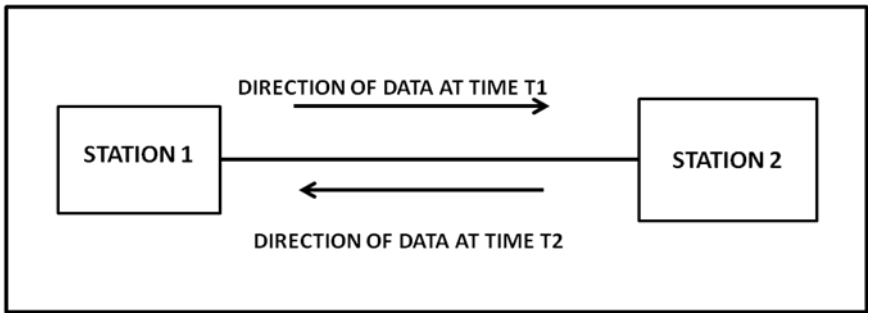


Figure: Half Duplex Mode of Communication

- In half duplex both the stations can transmit as well as receive but not at the same time.
- When one device is sending other can only receive and vice-versa (as shown in figure above.)
- Example: A walkie-talkie.

1.5.3 Full Duplex

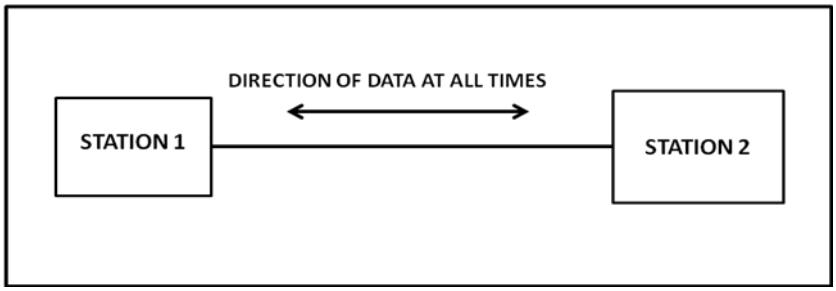


Figure: Full Duplex Mode of Communication

- In Full duplex mode, both stations can transmit and receive at the same time.
- Example: mobile phones

1.6 COMPUTER NETWORK

- Computer Networks are used for data communications

- **Definition:**
  A computer network can be defined as a collection of nodes.
  A node can be any device capable of transmitting or receiving data.
  The communicating nodes have to be connected by communication links.

- A Compute network should ensure
  - **reliability** of the data communication process, should c
  - **security** of the data

- **performance** by achieving higher throughput and smaller delay times

1.6.1 Categories of Network

Networks are categorized on the basis of their size. The three basic categories of computer networks are:

A. **Local Area Networks (LAN)** is usually limited to a few kilometers of area. It may be privately owned and could be a network inside an office on one of the floor of a building or a LAN could be a network consisting of the computers in a entire building.

B. **Wide Area Network (WAN)** is made of all the networks in a (geographically) large area. The network in the entire state of Maharashtra could be a WAN

C. **Metropolitan Area Network (MAN)** is of size between LAN & WAN. It is larger than LAN but smaller than WAN. It may comprise the entire network in a city like Mumbai.

1.7 PROTOCOL

- A Protocol is one of the components of a data communications system. Without protocol communication cannot occur. The sending device cannot just send the data and expect the receiving device to receive and further interpret it correctly.

- When the sender sends a message it may consist of text, number, images, etc. which are converted into bits and grouped into blocks to be transmitted and often certain additional information called control information is also added to help the receiver interpret the data.

- For successful communication to occur, the sender and receiver must agree upon certain rules called protocol.

- **A Protocol is defined as a set of rules that governs data communications.**

- A protocol defines what is to be communicated, how it is to be communicated and when it is to be communicated.

1.7.1 Elements of a Protocol

- There are three key elements of a protocol:

A. Syntax

- It means the structure or format of the data.
- It is the arrangement of data in a particular order.

B. Semantics

- It tells the meaning of each section of bits and indicates the interpretation of each section.
- It also tells what action/decision is to be taken based on the interpretation.

C. Timing

- It tells the sender about the readiness of the receiver to receive the data
- It tells the sender at what rate the data should be sent to the receiver to avoid overwhelming the receiver.

1.7 STANDARDS IN NETWORKING

- Standards are necessary in networking to ensure interconnectivity and interoperability between various networking hardware and software components.

- Without standards we would have proprietary products creating isolated islands of users which cannot interconnect.

1.7.1 Concept of Standard

- Standards provide guidelines to product manufacturers and vendors to ensure national and international interconnectivity.

- Data communications standards are classified into two categories:

1. **De facto Standard**
   - These are the standards that have been traditionally used and mean **by fact** or **by convention**
   - These standards are not approved by any organized body but are adopted by widespread use.

2. **De jure standard**
   - It means by **law** or **by regulation.**
     - These standards are legislated and approved by an body that is officially recognized.

1.7.2 Standard Organizations in field of Networking

- Standards are created by standards creation committees, forums, and government regulatory agencies.

- **Examples of Standard Creation Committees** :
  1. International Organization for Standardization(ISO)
  2. International Telecommunications Union – Telecommunications Standard (ITU-T)
  3. American National Standards Institute (ANSI)
  4. Institute of Electrical & Electronics Engineers (IEEE)
  5. Electronic Industries Associates (EIA)

- **Examples of Forums**
  1. ATM Forum
  2. MPLS Forum
  3. Frame Relay Forum

- **Examples of Regulatory Agencies:**
  1. Federal Communications Committee (FCC)

1.8 REVIEW QUESTIONS

1. Differentiate between data & information. What are the different forms in which data can be represented?
2. What are the characteristics of data communication?
3. What are the components of a data communication system?
4. Define computer network and categorize.
5. Explain protocols in details

1.9 REFERENCES

1. Data Communication & Networking – Behrouz Forouzan

Unit 1 Introduction to data communications and networking

2

Signals

Unit Structure

2.0 OBJECTIVES

- ✓ Introduce the readers to fundamentals of data & signal
- ✓ Types of data & signal
- ✓ Characteristics and nature of analog & digital signal
- ✓ Representation of signal
- ✓ Transmission of digital signals

2.1 INTRODUCTION

Computer networks are designed to transfer data from one point to another. During transit data is in the form of electromagnetic signals. Hence it is important to study data and signals before we move to further concepts in data communication.

2.2 DATA & SIGNALS

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

2.2.1. Data can be Analog or Digital.

1. **Analog data** refers to information that is continuous; ex. sounds made by a human voice
2. **Digital data** refers to information that has discrete states. Digital data take on discrete values.
3. For example, data are stored in computer memory in the form of Os and 1s

2.2.2. Signals can be of two types:

1. **Analog Signal:** They have infinite values in a range.
2. **Digital Signal:** They have limited number of defined values

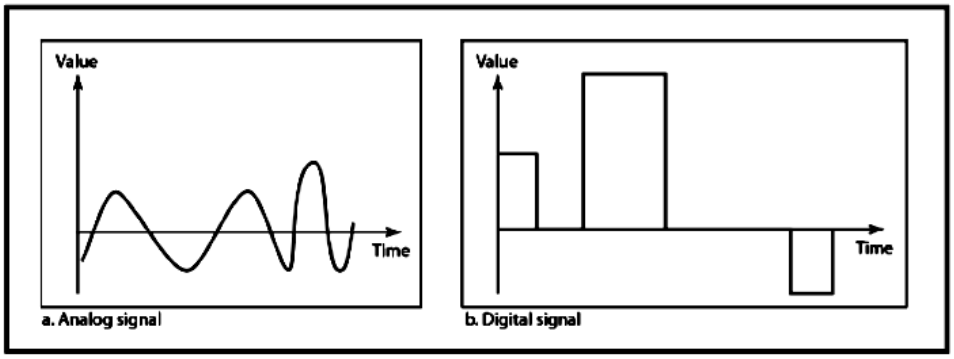


Figure: a. Analog Signal b. Digital Signal*

2.2.3. Periodic & Non Periodic Signals

- Signals which repeat itself after a fixed time period are called Periodic Signals.
- 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 non-periodic digital signals.**

2.3 ANALOG SIGNAL

- An analog signal has infinitely many levels of intensity over a period of time.
- As the wave moves from value *A* to value *B,* it passes through and includes an infinite number of values along its path as it can be seen in the figure below.
- A simple analog signal is a sine wave that cannot be further decomposed into simpler signals.

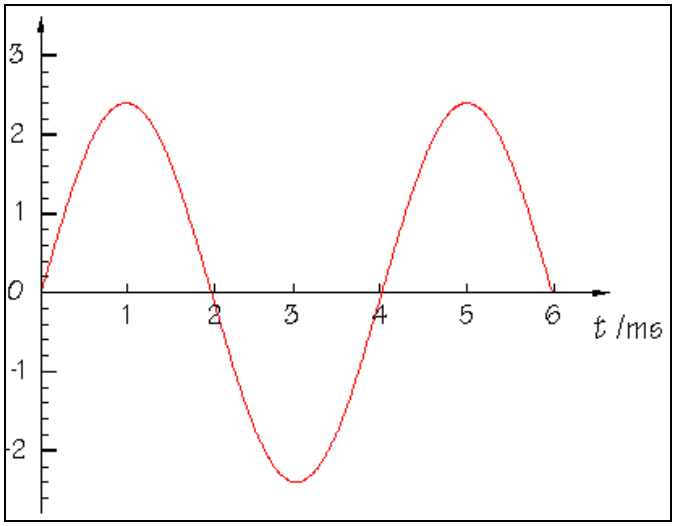


Fig. Sine wave

- A sine wave is characterized by three parameters:
  1. Peak Amplitude
  2. Frequency
  3. Phase

2.3.1 Characteristics of an Analog Signal

2.3.1.1 Peak Amplitude

- The amplitude of a signal is the absolute value of its intensity at time t
- The peak amplitude of a signal is the absolute value of the highest intensity.

- The amplitude of a signal is proportional to the energy carried by the signal

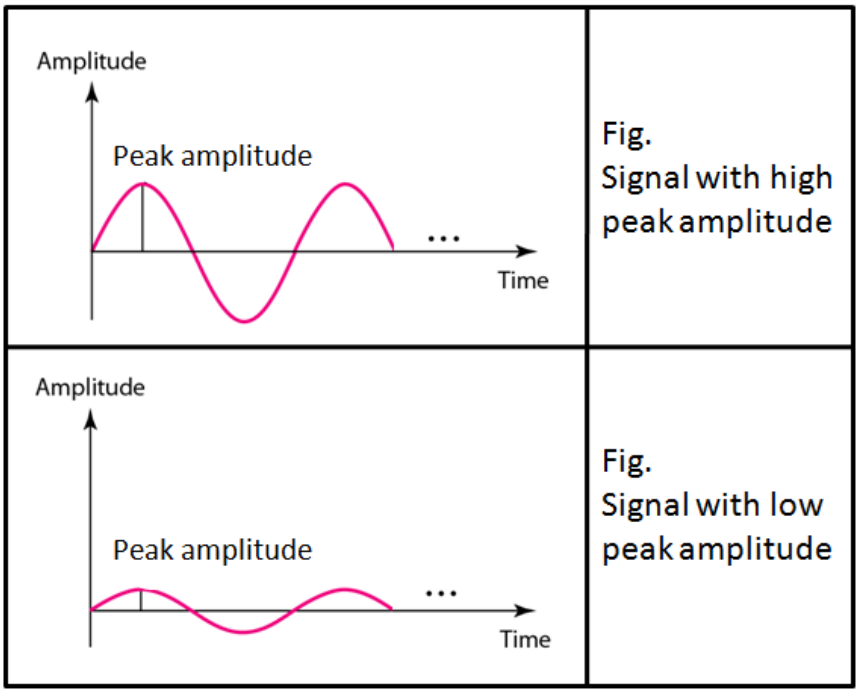


Fig. Amplitude of a sine wave

2.3.1.2. Frequency

- Frequency refers to the number of cycles completed by the wave in one second.
- Period refers to the time taken by the wave to complete one second.

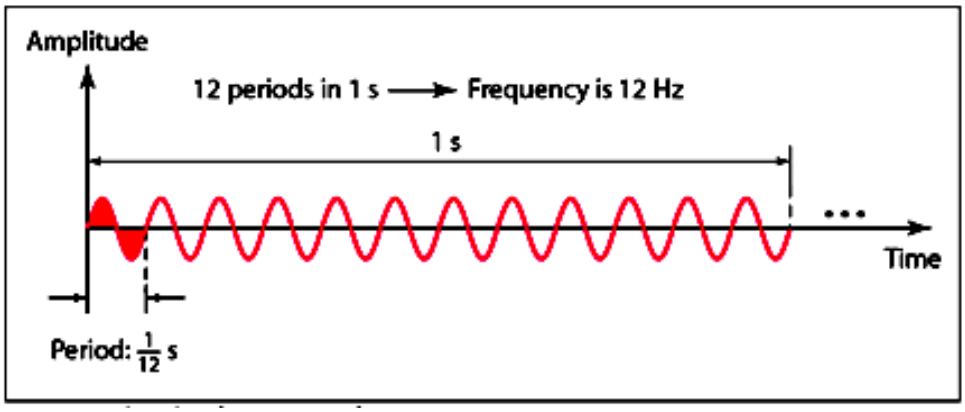


Fig: Frequency & Period of a sine wave

2.3.1.3. Phase

Phase describes the position of the waveform with respect to time (specifically relative to time O).

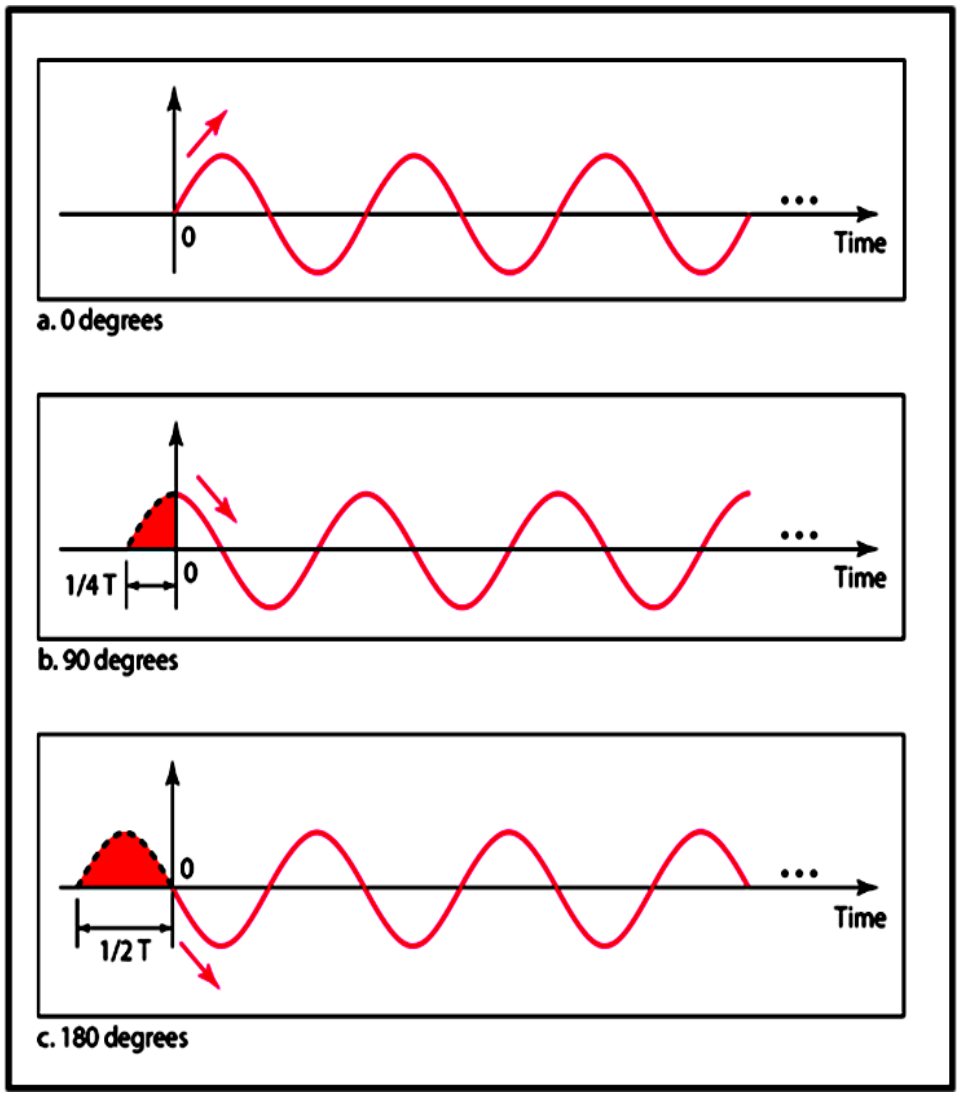


Fig: Phase of a sine wave*

- Phase indicates the forward or backward shift of the waveform from the axis
- It is measured in degrees or radian
- The figure above shows the sine waves with same amplitude and frequency but different phases

2.3.2 Relation between Frequency & Period

- **Frequency & Period are inverse of each other.**
- It is indicated by the following formula:

$$T = 1/f$$
$$Or$$
$$f = 1/T$$

Example1. A wave has a frequency of 100hz. Its period(T) is given by

$$T = 1/F = 1/100 = 0.01 \text{ sec}$$

Example2. A wave completes its one cycle in 0.25 seconds. Its frequency is given by

$$F = 1 / T = 1 / 0.25 = 4 \text{ Hz}$$

2.3.3 Wavelength

- The wavelength of a signal refers to the relationship between frequency (or period) and propagation speed of the wave through a medium.
- The wavelength is the distance a signal travels in one period.
- It is given by

$$\text{Wavelength} = \text{Propagation Speed} \times \text{Period}$$

OR

$$\text{Wavelength} = \text{Propagation Speed} \times \frac{1}{\text{Frequency}}$$

- It is represented by the symbol : λ (pronounced as lamda)
- It is measured in micrometers
- It varies from one medium to another.

2.3.4. Time Domain and Frequency domain representation of signals

- A sine wave can be represented either in the time domain or frequency domain.
- The **time-domain plot** shows changes in signal amplitude with respect to time. It indicates time and amplitude relation of a signal.
- The **frequency-domain plot** shows signal frequency and peak amplitude.
- The figure below show time and frequency domain plots of three sine waves.

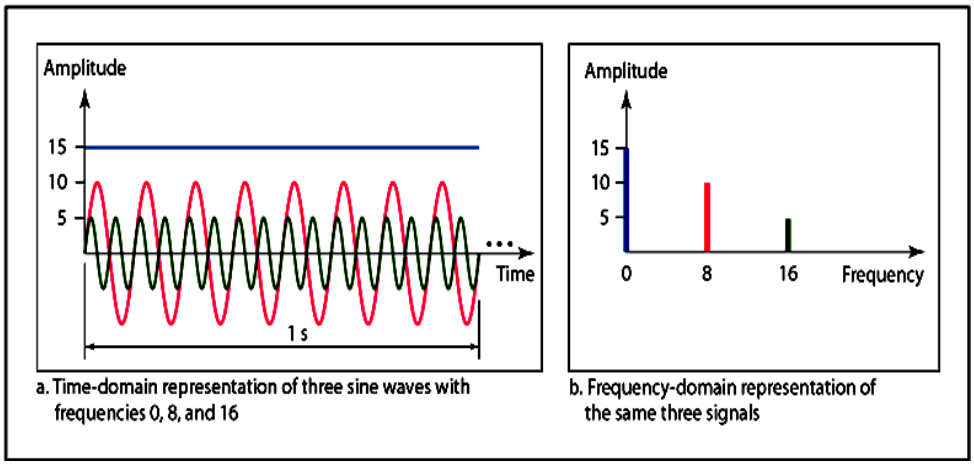


Fig: Time domain and frequency domain plots of three sine waves*

- A complete sine wave in the time domain can be represented by one single spike in the frequency domain

2.3.5. Composite Signal

- A composite signal is a combination of two or more simple sine waves with different frequency, phase and amplitude.
- If the composite signal is periodic, the decomposition gives a series of signals with discrete frequencies; if the composite signal is non-periodic, the decomposition gives a combination of sine waves with continuous frequencies.

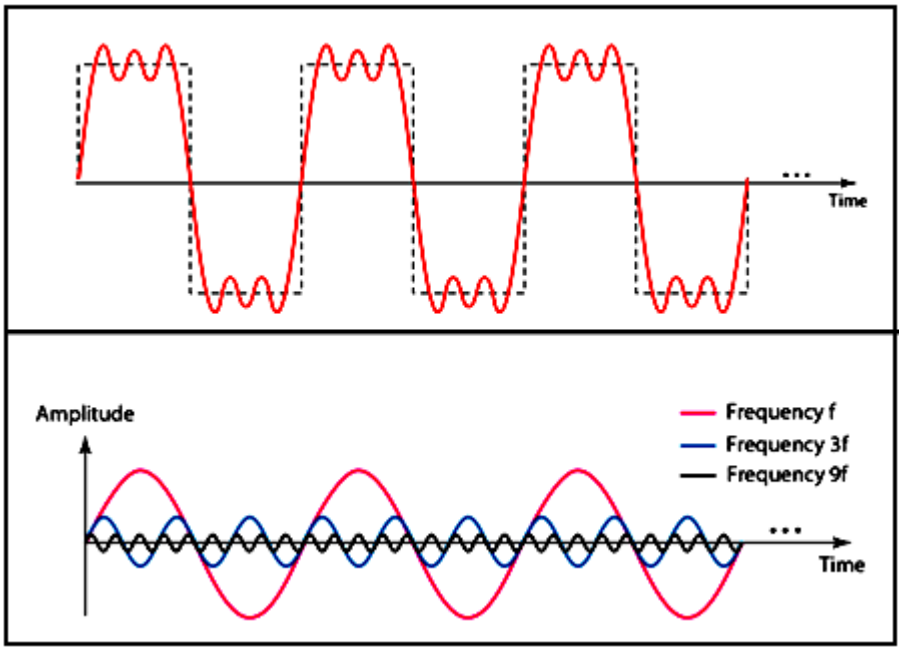


Fig: A Composite signal with three component signals

- For data communication a simple sine wave is not useful, what is used is a composite signal which is a combination of many simple sine waves.

- According to French Mathematician, Jean Baptist, any composite signal is a combination of simple sine waves with different amplitudes and frequencies and phases.

- Composite signals can be periodic or non periodic.

- A periodic composite signal can be decomposed into a series of signals with discrete frequencies.

- A non-periodic signal when decomposed gives a combination of sine waves with continuous frequencies.

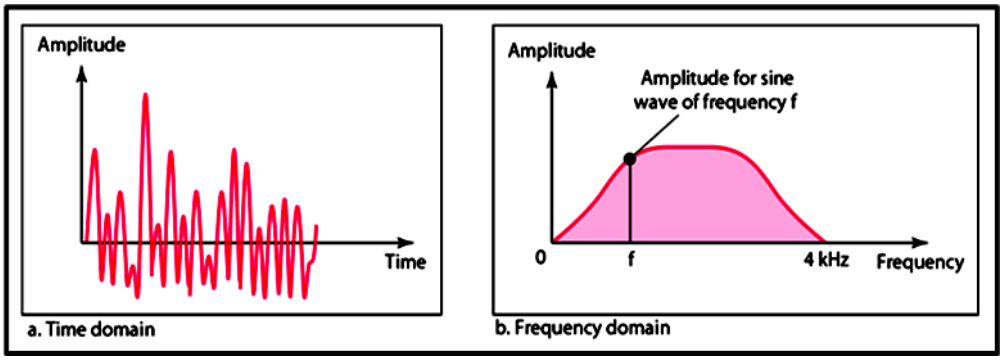


Fig The time and frequency domains of a non-periodic composite analog signal

2.4 Digital Signal

Information can also be explained in the form of a digital signal.

A digital signal can be explained with the help of following points:

2.4.1 Definition:-

- A digital is a signal that has discrete values.
- The signal will have value that is not continuous.

2.4.2 LEVEL

- Information in a digital signal can be represented in the form of voltage levels.
- Ex. In the signal shown below, a ‘1’ is represented by a positive voltage and a ‘0’ is represented by a Zero voltage.

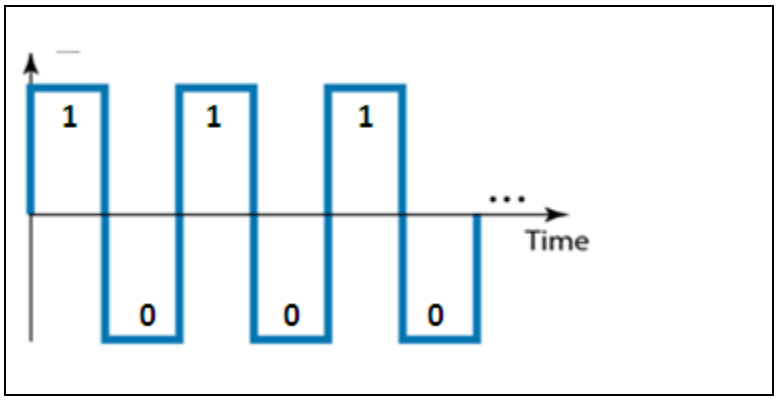


Fig: A digital signal with Two levels. '1' represented by a positive voltage and '0' represented by a negative voltage

- A Signal can have more than two levels

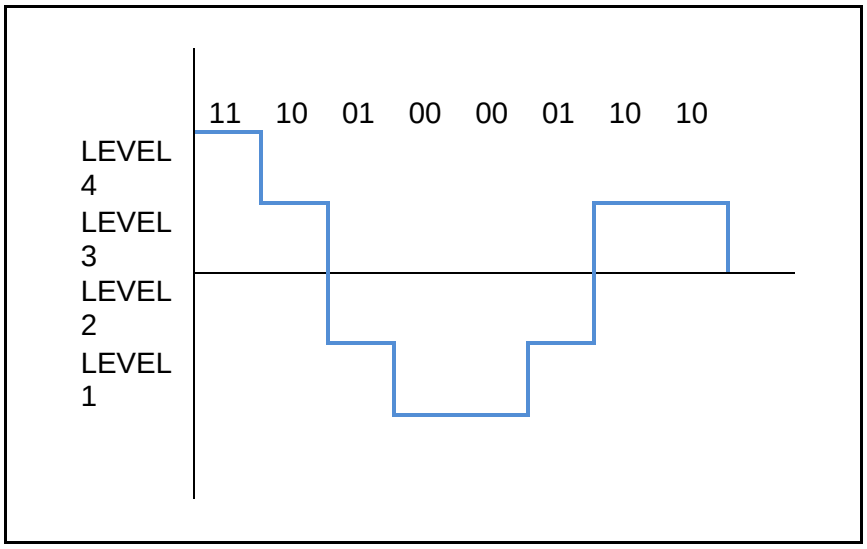


Fig: A digital signal with four levels

- In general, if a signal has L levels then, each level need Log_2L bits
- Example: Consider a digital Signal with four levels, how many bits are required per level?
  Answer: Number of bits per level = Log_2L
  $= Log_24$
  $= 2$
  Hence, 2 bits are required per level for a signal with four levels.

2.4.3 BIT LENGTH or Bit Interval (T_b)

- It is the time required to send one bit.
- It is measured in seconds.

2.4.4 BIT RATE

- It is the number of bits transmitted in one second.
- It is expressed as bits per second (bps).
- Relation between bit rate and bit interval can be as follows

  Bit rate = 1 / Bit interval

2.4.5 Baud Rate

- It is the rate of Signal Speed, i.e the rate at which the signal changes.
- A digital signal with two levels '0' & '1' will have the same baud rate and bit rate & bit rate.
- The diagram below shows three signal of period (T) 1 second
  a) Signal with a bit rate of 8 bits/ sec and baud rate of 8 baud/sec
  b) Signal with a bit rate of 16 bits/ sec and baud rate of 8 baud/sec
  c) Signal with a bit rate of 16 bits/ sec and baud rate of 4 baud/sec

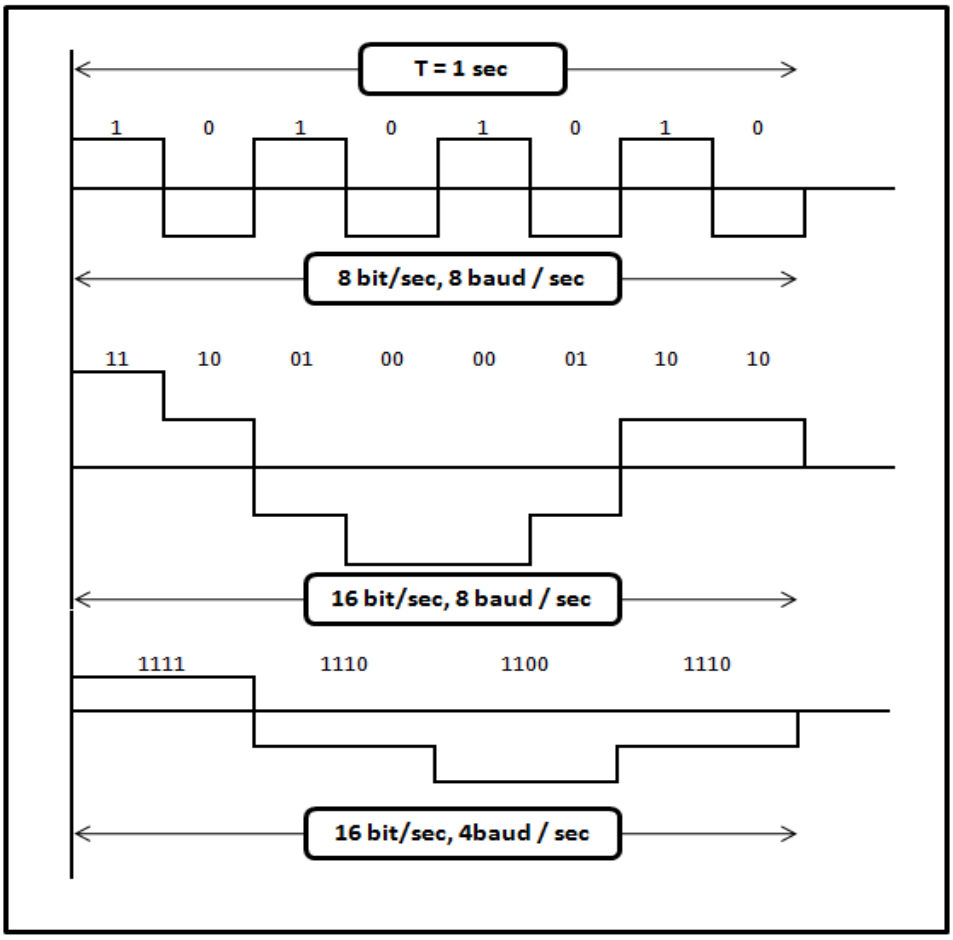


Fig: Three signals with different bit rates and baud rates

2.5 TYPES OF CHANNELS:

Each composite signal has a lowest possible(minimum) frequency and a highest possible (maximum) frequency.

From the point of view of transmission, there are two types of channels:

2.5.1 Low pass Channel

- This channel has the lowest frequency as '0' and highest frequency as some non-zero frequency 'f1'.
- This channel can pass all the frequencies in the range 0 to f1.

2.5.2 Band pass channel

- This channel has the lowest frequency as some non-zero frequency 'f1' and highest frequency as some non-zero frequency 'f2'.
- This channel can pass all the frequencies in the range f1 to f2.

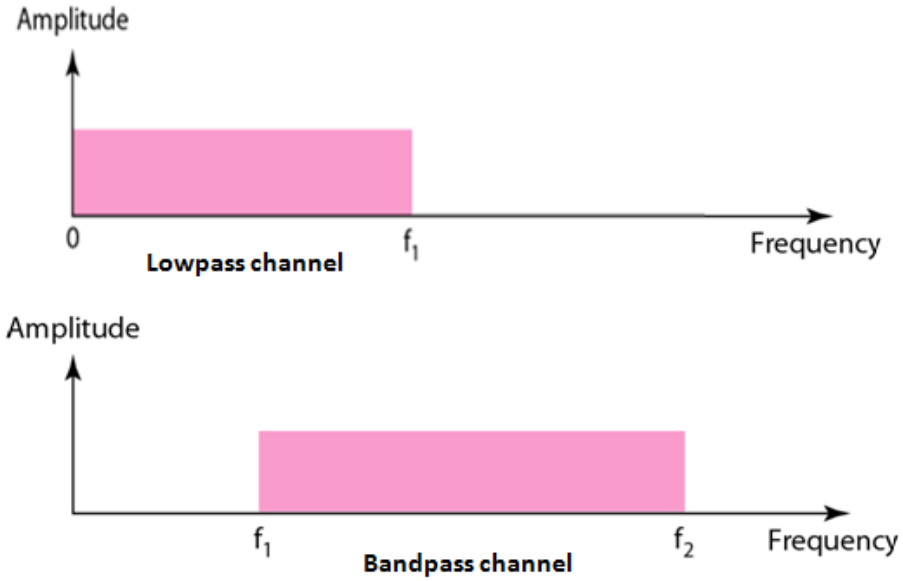


Fig: Lowpass Channel & Bandpass Channel

2.6 Transmission of Digital signal

Digital signal can be transmitted in the following two ways:

2.6.1 Baseband Transmission

- The signal is transmitted without making any change to it (ie. Without modulation)

- In baseband transmission, the bandwidth of the signal to be transmitted has to be less than the bandwidth of the channel.

- Ex. Consider a Baseband channel with lower frequency 0Hz and higher frequency 100Hz, hence its bandwidth is 100 (Bandwidth is calculated by getting the difference between the highest and lowest frequency).

- We can easily transmit a signal with frequency below 100Hz, such a channel whose bandwidth is more than the bandwidth of the signal is called **Wideband** channel

- Logically a signal with frequency say 120Hz will be blocked resulting in loss of information, such a channel whose bandwidth is less than the bandwidth of the signal is called **Narrowband** channel

2.6.2 Broad band Transmission

- Given a bandpass channel, a digital signal cannot be transmitted directly through it

- In broadband transmission we use modulation, i.e we change the signal to analog signal before transmitting it.

- The digital signal is first converted to an analog signal, since we have a bandpass channel we cannot directly send this signal through the available channel. Ex. Consider the bandpass channel with lower frequency 50Hz and higher frequency 80Hz, and the signal to be transmitted has frequency 10Hz.

- To pass the analog signal through the bandpass channel, the signal is modulated using a carrier frequency. Ex. The analog signal (10Hz) is modulated by a carrier frequency of 50Hz resulting in an signal of frequency 60Hz which can pass through our bandpass channel.

- The signal is demodulated and again converted into an digital signal at the other end as shown in the figure below.

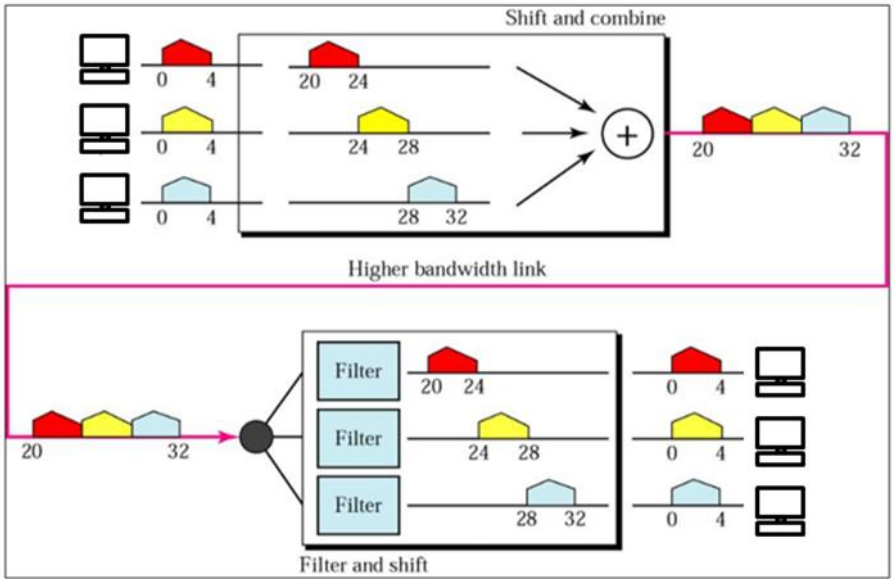


Fig: Broadband Transmission Involving Modulation & Demodulation

2.7 REVIEW QUESTIONS

1. Define analog and digital signals
2. Explain Composite analog signals.
3. Explain Time and Frequency Domain Representation of signals
4. Explain the characteristics of an Analog signal
5. Explain the characteristics of an Digital signal
6. Explain the difference between
    1. Lowpass and Bandpass channel
    2. Narrowband and wideband channel
7. Explain why a digital signal requires to undergo a change before transmitting it through a bandpass channel.

2.8 REFERENCES & FURTHER READING

Data Communication & Networking – Behrouz Forouzan

❖❖❖❖

3

BANDWIDTH

Unit Structure

3.0 OBJECTIVES

To understand

- ✓ Concept of bandwidth
- ✓ Bandwidth of Analog signal
- ✓ Bandwidth of Digital signal
- ✓ Bandwidth of Channel
- ✓ Maximum Data rate of a channel : noisy & noiseless

3.1 INTRODUCTION

This chapter gives insights to the concept of bandwidth. It tells about bandwidth of signal and medium. Also explains how to calculate the bandwidth for a noisy and noiseless channel

3.2 FOURIER ANALYSIS

- In the 19^{th} century, French mathematician Jean-Baptiste Fourier proved that any composite signal is a combination of simple sine waves with different frequencies, amplitudes, and phases.

- A Composite signal can be periodic as well as non periodic.
- A periodic composite signal when decomposed gives a series of simple sine waves with discrete frequencies i.e. frequencies that have integer values (1, 2, 3, etc).
- A non-periodic composite signal when decomposed gives a combination of an infinite number of simple sine waves with continuous frequencies i.e. frequencies that have real values.

3.3 BANDWIDTH OF A SIGNAL

- Bandwidth can be defined as the portion of the electromagnetic spectrum occupied by the signal

- It may also be defined as the frequency range over which a signal is transmitted.

- Different types of signals have different bandwidth. Ex. Voice signal, music signal, etc

- Bandwidth of analog and digital signals are calculated in separate ways; analog signal bandwidth is measured in terms of its frequency (hz) but digital signal bandwidth is measured in terms of bit rate (bits per second, bps)

- Bandwidth of signal is different from bandwidth of the medium/channel

3.3.1 Bandwidth of an analog signal

- Bandwidth of an analog signal is expressed in terms of its frequencies.

- It is defined as the range of frequencies that the composite analog signal carries.

- It is calculated by the difference between the maximum frequency and the minimum frequency.

- Consider the signal shown in the diagram below:

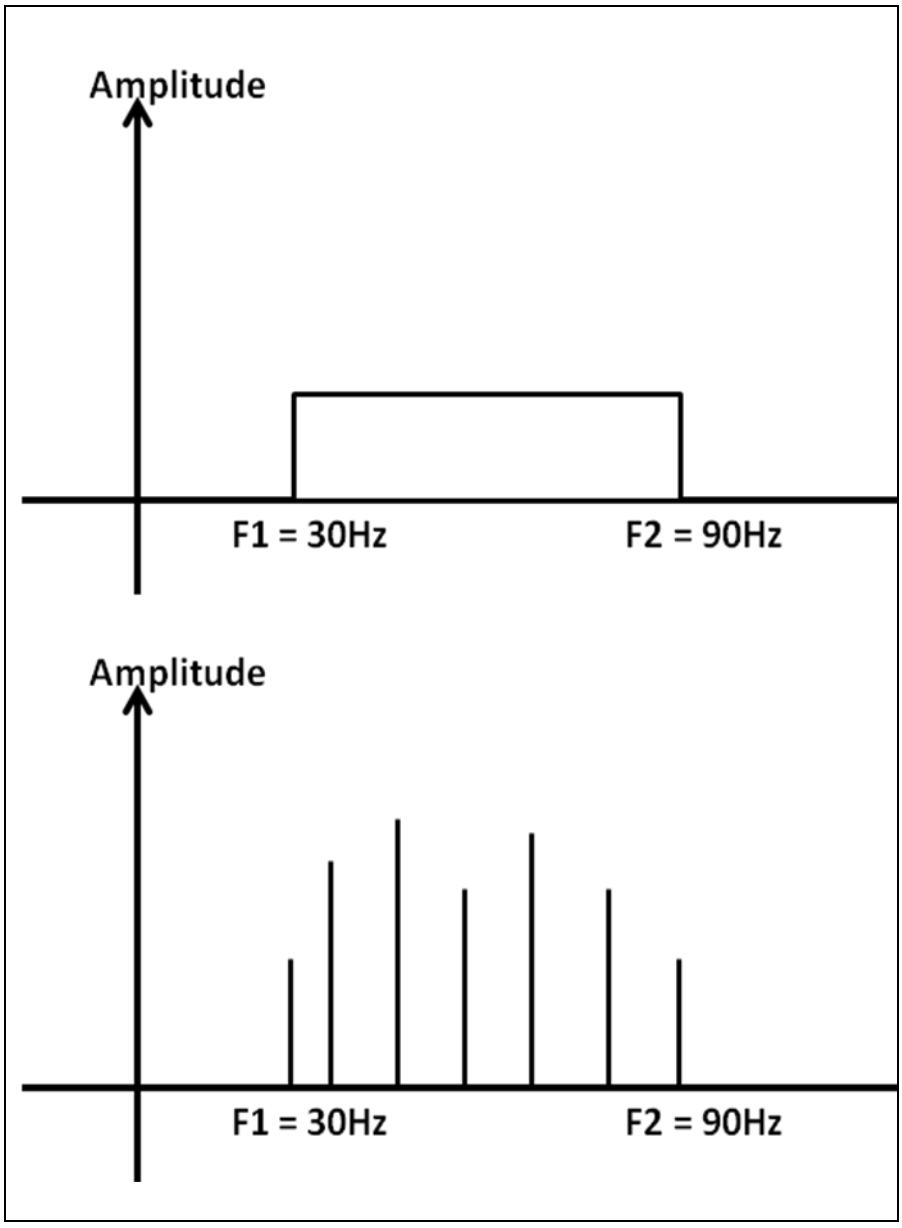


Fig: Bandwidth of a signal in time domain and frequency domain

- The signal shown in the diagram is an composite analog signal with many component signals.
- It has a minimum frequency of F1 = 30Hz and maximum frequency of F2 = 90Hz.
- Hence the bandwidth is given by F2 – F1 = 90 – 30 = 60 Hz

3.3.2 Bandwidth of a digital signal

- It is defined as the maximum bit rate of the signal to be transmitted.
- It is measured in bits per second.

3.4 BANDWIDTH OF A CHANNEL

- A channel is the medium through which the signal carrying information will be passed.

- In terms of analog signal, bandwidth of the channel is the range of frequencies that the channel can carry.

- In terms of digital signal, bandwidth of the channel is the maximum bit rate supported by the channel. i.e. the maximum amount of data that the channel can carry per second.

- The bandwidth of the medium should always be greater than the bandwidth of the signal to be transmitted else the transmitted signal will be either attenuated or distorted or both leading in loss of information.

- The channel bandwidth determines the type of signal to be transmitted i.e. analog or digital.

3.5 THE MAXIMUM DATA RATE OF A CHANNEL

Data rate depends on three factors:

1. The bandwidth available
2. The level of the signals we use
3. The quality of the channel (the level of noise)

The quality of the channel indicates two types:

a) **A Noiseless or Perfect Channel**
   - An ideal channel with no noise.
   - The Nyquist Bit rate derived by Henry Nyquist gives the bit rate for a Noiseless Channel.

b) A Noisy Channel
   - A realistic channel that has some noise.
   - The Shannon Capacity formulated by Claude Shannon gives the bit rate for a Noisy Channel

3.5.1 Nyquist Bit Rate

The Nyquist bit rate formula defines the theoretical maximum bit rate for a noiseless channel

$$\text{Bitrate} = 2 \times \text{Bandwidth} \times \text{Log}_2 L$$

Where,

- Bitrate is the bitrate of the channel in bits per second
- Bandwidth is the bandwidth of the channel
- L is the number of signal levels.

Example
What is the maximum bit rate of a noiseless channel with a bandwidth of 5000 Hz transmitting a signal with two signal levels.

Solution:
The bit rate for a noiseless channel according to Nyquist Bit rate can be calculated as follows:

$$\text{BitRate} = 2 \times \text{Bandwidth} \times \text{Log}_2 L$$
$$= 2 \times 5000 \times \log_2 2 = \mathbf{10000\ bps}$$

3.5.2 Shannon Capacity

The Shannon Capacity defines the theoretical maximum bit rate for a noisy channel

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

Where,

- Capacity is the capacity of the channel in bits per second
- Bandwidth is the bandwidth of the channel
- SNR is the Signal to Noise Ratio

Shannon Capacity for calculating the maximum bit rate for a noisy channel does not consider the number of levels of the signals being transmitted as done in the Nyquist bit rate.
Example:
Calculate the bit rate for a noisy channel with SNR 300 and bandwidth of 3000Hz
Solution:

The bit rate for a noisy channel according to Shannon Capacity can be calculated as follows:

$$\text{Capacity} = \text{bandwidth} \times \log_2 (1 + \text{SNR})$$
$$= 3000 \times \log_2 (1 + 300)$$
$$= 3000 \times \log_2 (301)$$
$$= 3000 \times 8.23$$
$$= \mathbf{24{,}690bps}$$

3.6 REVIEW QUESTIONS

1. Explain the term bandwidth of a signal
2. Explain the term bandwidth of a channel.
3. Write short note on maximum data rate of a channel.

3.7 REFERENCES & FURTHER READING

Data Communication & Networking – Behrouz Forouzan

4

NETWORK MODELS

Unit Structure

4.0 OBJECTIVES

- ✓ Understand concept of dividing a job into layered tasks
- ✓ Get introduced to the OSIRM
- ✓ Understand the functions of the various layers of the OSI Mode.

4.1 INTRODUCTION

In the study of computer networks it is essential to study the way our networks work. Computer networks are operated by network models; most prominently the OSIRM and the TCP/ IP Model. This chapter gives the understanding of the OSI reference model.

4.2 CONCEPT OF LAYERED TASK

i. The main objective of a computer network is to be able to transfer the data from sender to receiver. This task can be done by breaking it into small sub tasks, each of which are well defined.

ii. Each subtask will have its own process or processes to do and will take specific inputs and give specific outputs to the subtask before or after it. In more technical terms we can call these sub tasks as layers.

iii. In general, every task or job can be done by dividing it into sub task or layers. Consider the example of sending a letter where the sender is in City A and receiver is in city B.

iv. The process of sending letter is shown below:

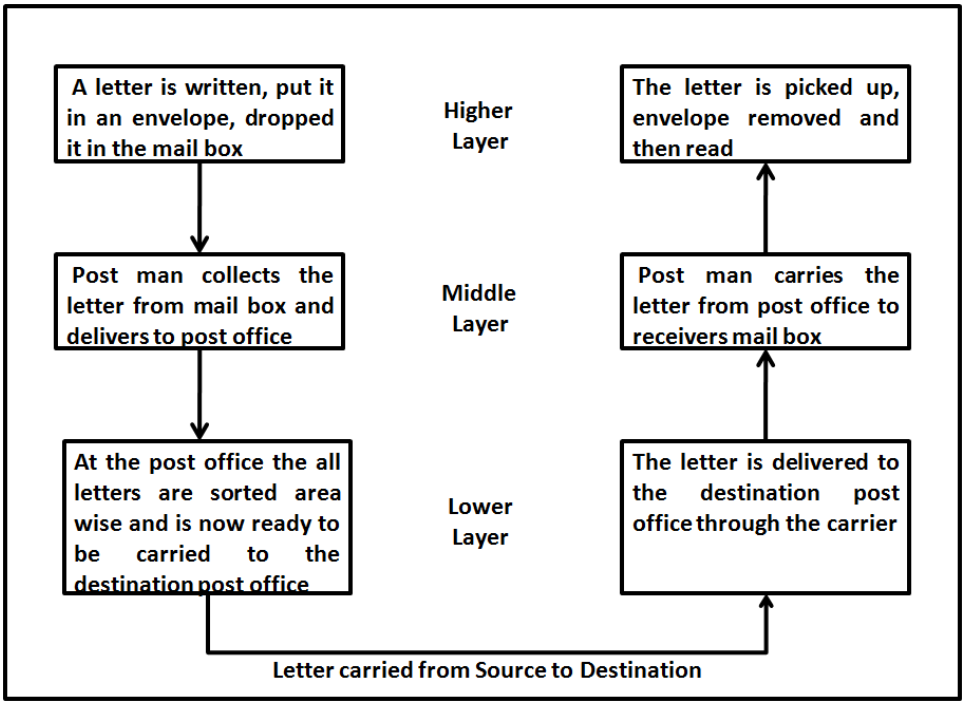


Fig: Concept of layer task: sending a letter

v. The above figure shows
   a. Sender, Receiver & Carrier
   b. Hierarchy of layers

vi. At the sender site, the activities take place in the following descending order:
   a. Higher Layer: The sender writes the letter along with the sender and receivers address and put it in an envelope and drop it in the mailbox.
   b. Middle Layer: The letter is picked up by the post man and delivered to the post office
   c. Lower Layer: The letters at the post office are sorted and are ready to be transported through a carrier.

vii. During transition the letter may be carried by truck, plane or ship or a combination of transport modes before it reaches the destination post office.

viii. At the Receiver site, the activities take place in the following ascending order:
   a. Lower Layer: The carrier delivers the letter to the destination post office
   b. Middle Layer: After sorting, the letter is delivered to the receivers mail box
   c. Higher Layer: The receiver picks up the letter, opens the envelope and reads it.

ix. Hierarchy of layers: The activities in the entire task are organized into three layers. Each activity at the sender or receiver side occurs in a particular order at the hierarchy.

x. The important and complex activities are organized into the Higher Layer and the simpler ones into middle and lower layer.

4.3 OPEN SYSTEMS INTER CONNECTION REFERENCE MODEL (OSIRM)

4.3.1 Introduction to OSI Model & its layers

- The Open Systems Interconnection (OSI) Model was developed by International Organization for Standardization (ISO).
- ISO is the organization, OSI is the model
- It was developed to allow systems with different platforms to communicate with each other. Platform could mean hardware, software or operating system.
- It is a network model that defines the protocols for network communications.
- It is a hierarchical model that groups its processes into layers. It has 7 layers as follows: (Top to Bottom)
    1. Application Layer
    2. Presentation Layer
    3. Session Layer
    4. Transport Layer
    5. Network Layer
    6. Data Link Layer
    7. Physical Layer
- Each layer has specific duties to perform and has to co-operate with the layers above and below it.

4.3.2 Layered Architecture of OSI Model

- The OSI model has 7 layers each with its own dedicated task.

- A message sent from Device A to Device B passes has to pass through all layers at A from top to bottom then all layers at B from bottom to top as shown in the figure below.

- At Device A, the message is sent from the top layer i.e Application Layer A then all the layers till it reaches its physical layer and then it is transmitted through the transmission medium.

- At Device B, the message received by the physical layer passes through all its other layers and moves upwards till it reaches its Application Layer.

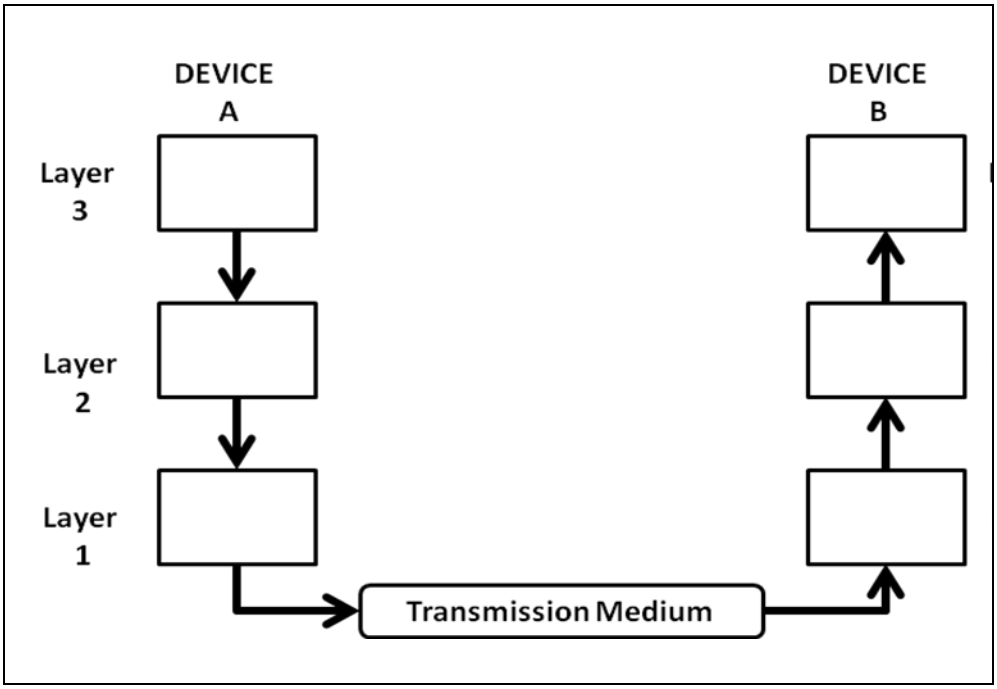


Fig: Flow of Data from Device A to Device B through various layers

- As the message travels from device A to device B, it may pass through many intermediate nodes. These intermediate nodes usually involve only the first three layers of the OSI model as shown below.

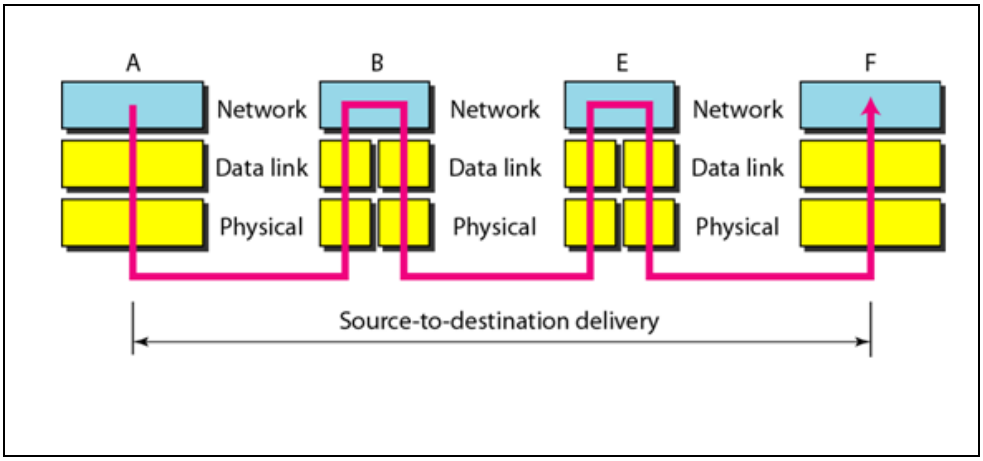


Fig: Data Transfer through Intermediate nodes

- The Data Link layer determines the next node where the message is supposed to be forwarded and the network layer determines the final recipient.

4.3.3 Communication & Interfaces

- For communication to occur, each layer in the sending device adds its own information to the message it receives from the layer just above it and passes the whole package to the layer just below it. Each layer in the receiving device removes the information added at the corresponding layer and sends the obtained data to the layer above it.

- Every Layer has its own dedicated function or services and is different from the function of the other layers.

- On every sending device, each layer calls upon the service offered by the layer below it.

- On every receiving device, each layer calls upon the service offered by the layer above it.

- Between two devices, the layers at corresponding levels communicate with each other .i.e layer 2 at receiving end can communicate and understand data from layer 2 of sending end. This is called peer –to – peer communication.

- For this communication to be possible between every two adjacent layers there is an interface. An interface defines the service that a layer must provide. Every layer has an interface to the layer above and below it as shown in the figure below

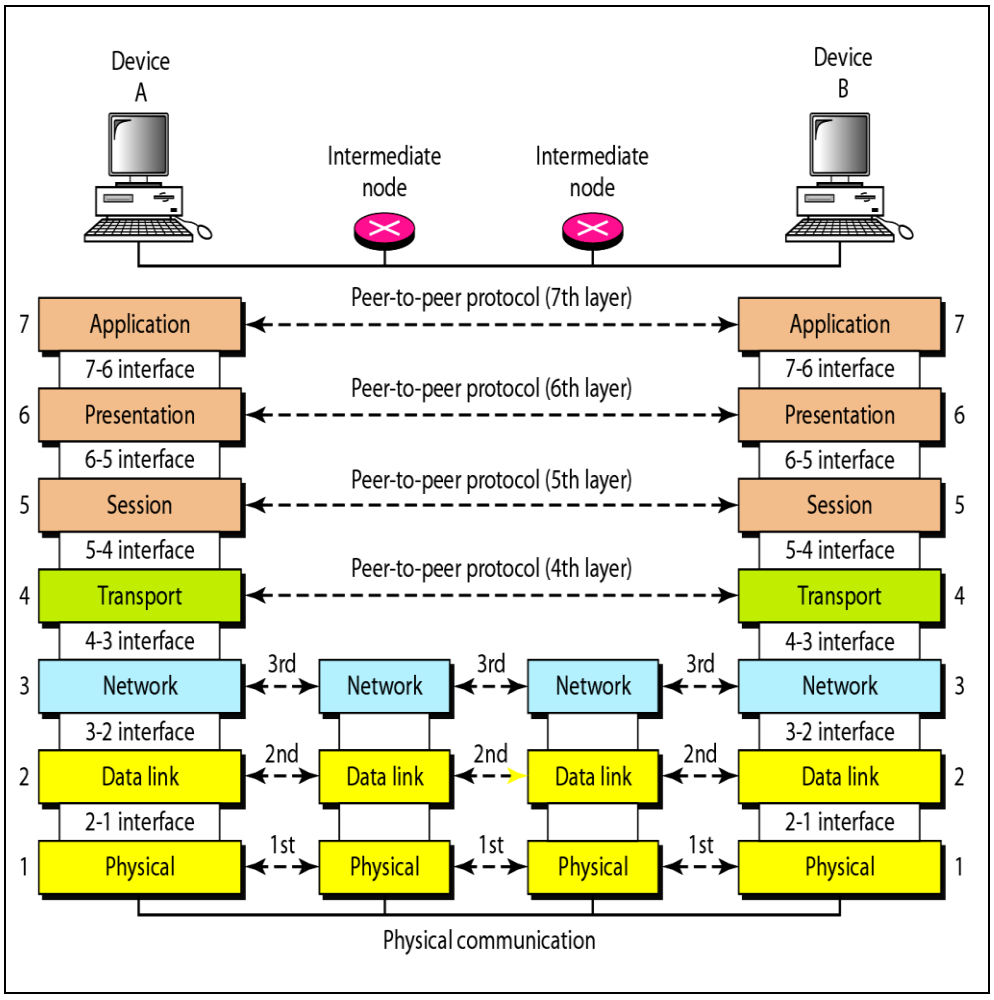


Fig: Communication & Interfaces in the OSI model

4.3.4 Encapsulation of Data

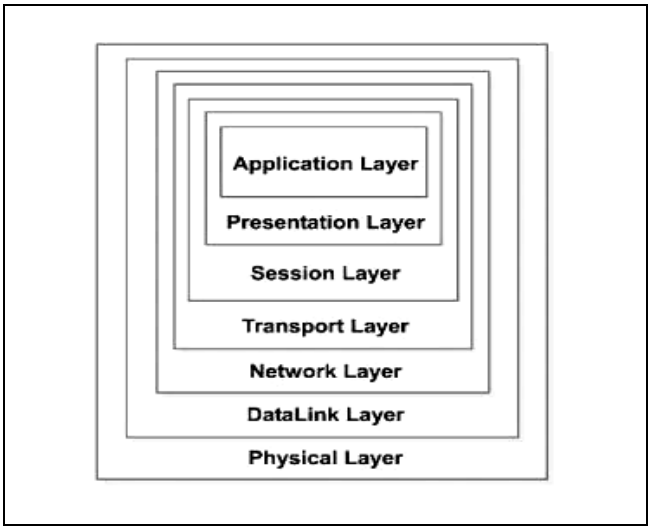


Fig: Encapsulation

- As shown in the figure above the data at layer 7 i.e the Application layer along with the header added at layer 7 is given to layer 6, the Presentation layer. This layer adds Its header and passed the whole package to the layer below.

- The corresponding layers at the receiving side removes the corresponding header added at that layer and sends the remaining data to the above layer.

- The above process is called encapsulation

4.3.5 Description of Layers in the OSI Model

4.3.5.1 Physical Layer

I. The Physical Layer provides a standardized interface to physical transmission media, including :
   a. Mechanical specification of electrical connectors and cables, for example maximum cable length
   b. Electrical specification of transmission line
   c. Bit-by-bit or symbol-by-symbol delivery

II. On the sender side, the physical layer receives the data from Data Link Layer and encodes it into signals to be transmitted onto the medium. On the receiver side, the physical layer receives the signals from the transmission medium decodes it back into data and sends it to the Data Link Layer as shown in the figure below:

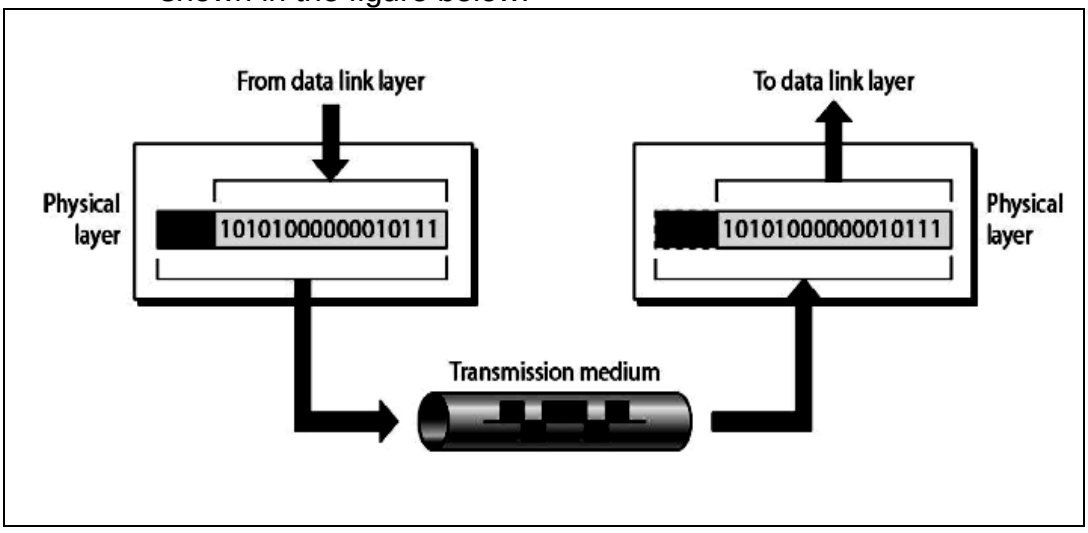


Fig: Transmission of data to and from Physical Layer

III. **Interface**

The Physical Layer defines the characteristics of interfaces between the devices & transmission medium.

IV. **Representation of bits**
The physical layer is concerned with transmission of signals from one device to another which involves converting data (1's & 0's) into signals and vice versa. It is not concerned with the meaning or interpretation of bits.

V. **Data rate**
The physical layer defines the data transmission rate i.e. number of bits sent per second. It is the responsibility of the physical layer to maintain the defined data rate.

VI. **Synchronization of bits**
To interpret correct and accurate data the sender and receiver have to maintain the same bit rate and also have synchronized clocks.

VII. **Line configuration**
The physical layer defines the nature of the connection .i.e. a point to point link, or a multi point link.

VIII. **Physical Topology**
The physical layer defines the type of topology in which the device is connected to the network. In a mesh topology it uses a multipoint connection and other topologies it uses a point to point connection to send data.

IX. **Transmission mode**
The physical layer defines the direction of data transfer between the sender and receiver. Two devices can transfer the data in simplex, half duplex or full duplex mode

X. **Main responsibility of the physical layer**
Transmission of bits from one hop to the next.

4.3.5.2 Data Link Layer

I. The Data Link layer adds reliability to the physical layer by providing error detection and correction mechanisms.

II. On the sender side, the Data Link layer receives the data from Network Layer and divides the stream of bits into fixed size manageable units called as **Frames** and sends it to the physical layer. On the receiver side, the data link layer receives the stream of bits from the physical layer and regroups them into frames and sends them to the Network layer. This process is called **Framing.** It is shown in the figure below:

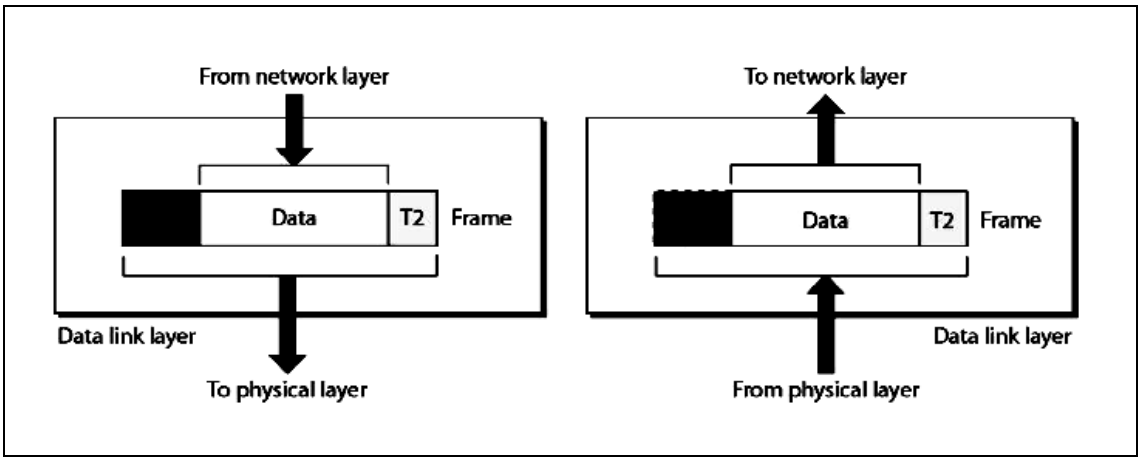


Fig: Data Link Layer: The process of Framing

III. Physical Addressing (inside / outside senders network)

a. The Data link layer appends the physical address in the header of the frame before sending it to physical layer.

b. The physical address contains the address of the sender and receiver.

c. In case the receiver happens to be on the same physical network as the sender; the receiver is at only one hop from the sender and the receiver address contains the receiver's physical address.

d. In case the receiver is not directly connected to the sender, the physical address is the address of the next node where the data is supposed to be delivered.

IV. Flow control

a. The data link layer makes sure that the sender sends the data at a speed at which the receiver can receive it else if there is an overflow at the receiver side the data will be lost.

b. The data link layer imposes flow control mechanism over the sender and receiver to avoid overwhelming of the receiver.

V. Error control

a. The data link layer imposes error control mechanism to identify lost or damaged frames, duplicate frames and then retransmit them.

b. Error control information is present in the trailer of a frame.

VI. Access Control

a. The data link layer imposes access control mechanism to determine which device has right to send data in an multipoint connection scenario.

VII. Main Responsibility

i. The main responsibility of the data link layer is hop to hop transmission of frames.

4.3.5.3 Network Layer

I. The network layer makes sure that the data is delivered to the receiver despite multiple intermediate devices.

II. The network layer at the sending side accepts data from the transport layer, divides it into packets, adds addressing information in the header and passes it to the data link layer. At the receiving end the network layer receives the frames sent by data link layer, converts them back into packets, verifies the physical address (verifies if the receiver address matches with its own address) and the send the packets to the transport layer.

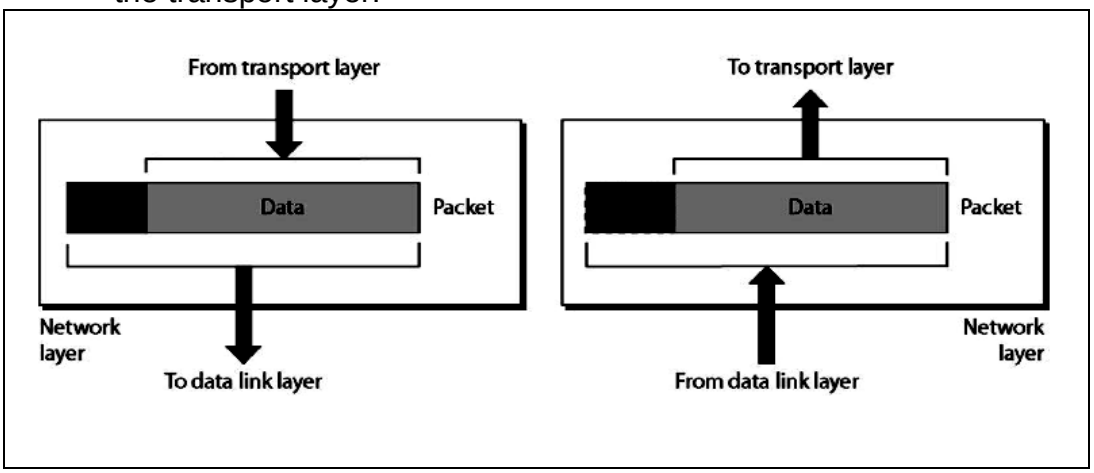


Fig: Network Layer

III. The network layer is responsible for source to destination of delivery of data. Hence it may have to route the data through multiple networks via multiple intermediate devices. In order to achieve this the network layer relies on two things:

a. Logical Addressing
b. Routing

IV. **Logical Addressing**

- The network layer uses logical address commonly known as IP address to recognize devices on the network.

- An IP address is a universally unique address which enables the network layer to identify devices outside the sender's network.
- The header appended by the network layer contains the actual sender and receiver IP address.

- At every hop the network layer of the intermediate node check the IP address in the header, if its own IP address does not match with the IP address of the receiver found in the header, the intermediate node concludes that it is not the final node but an intermediate node and passes the packet to the data link layer where the data is forwarded to the next node.

V. **Routing**

- **VI.** The network layer divides data into units called packets of equal size and bears a sequence number for rearranging on the receiving end.

- Each packet is independent of the other and may travel using different routes to reach the receiver hence may arrive out of turn at the receiver.

- Hence every intermediate node which encounters a packet tries to compute the best possible path for the packet. The best possible path may depend on several factors such as congestion, number of hops, etc

- This process of finding the best path is called as Routing. It is done using routing algorithms.

VI. The Network layer does not perform any flow control or error control

VII. Main Responsibility

- The main responsibility of Network Layer is transmission of packets from source to destination

4.3.5.4 Transport Layer

I. A logical address at network layer facilitates the transmission of data from source to destination device. But the source and the destination both may be having multiple processes communicating with each other. Hence it is important to deliver the data not only from the sender to the receiver but from the correct process on the sender to the correct process on the receiver. The transport layer takes care of process to process delivery of data and makes sure that it is intact and in order.

II. At the sending side, the transport layer receives data from the session layer, divides it into units called segments and sends it to the network layer. At the receiving side, the transport layer receives packets from the network layer, converts and arranges into proper sequence of segments and sends it to the session layer.

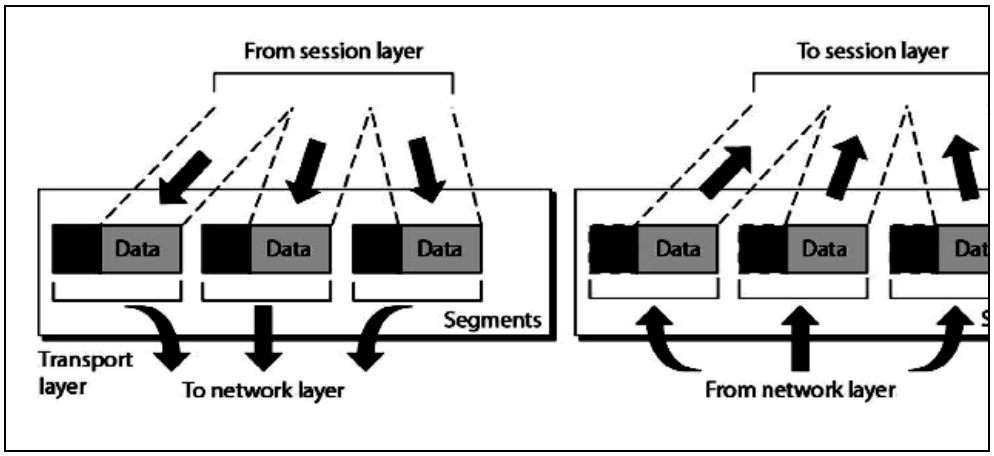


Fig: Transport Layer

III. To ensure process to process delivery the transport layer makes use of **port address** to identify the data from the sending and receiving process. A Port Address is the name or label given to a process. It is a 16 bit address. Ex. TELNET uses port address 23, HTTP uses port address 80. Port address is also called as Service Point Address

IV. The data can be transported in a connection oriented or connectionless manner. If the connection is connection oriented then all segments are received in order else they are independent of each other and are received out of order and have to be rearranged.

V. The Transport layer is responsible for segmentation and reassembly of the message into segments which bear sequence numbers. This numbering enables the receiving transport layer to rearrange the segments in proper order.

VI. Flow Control & Error control: the transport layer also carries out flow control and error control functions; but unlike data link layer these are end to end rather than node to node.

VII. Main Responsibility

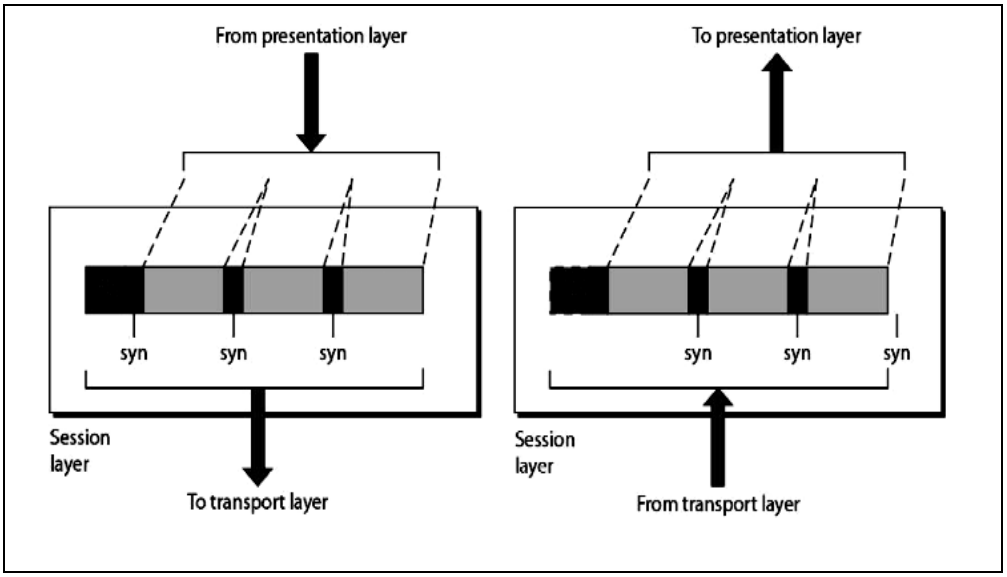
- The main responsibility of the transport layer is process to process delivery of the entire message.

4.3.5.5 Session Layer

I. The session layer establishes a session between the communicating devices called dialog and synchronizes their interaction. It is the responsibility of the session layer to establish and synchronize the dialogs. It is also called the network dialog controller.

II. The session layer at the sending side accepts data from the presentation layer adds checkpoints to it called syn bits and passes the data to the transport layer. At the receiving end the session layer receives data from the transport layer removes the checkpoints inserted previously and passes the data to the presentation layer.

III. The checkpoints or synchronization points is a way of informing the status of the data transfer. Ex. A checkpoint after first 500 bits of data will ensure that those 500 bits are not sent again in case of retransmission at 650^{th} bit.



IV. **Main responsibility of session layer is** dialog control and synchronizatoin

4.3.5.6 Presentation Layer

I. The communicating devices may be having different platforms. The presentation layer performs translation, encryption and compression of data.

II. The presentation layer at sending side receives the data from the application layer adds header which contains information related to encryption and compression and sends it to the session layer. At the receiving side, the presentation layer receives data from the session layer decompresses and decrypts the data as required and translates it back as per the encoding scheme used at the receiver.

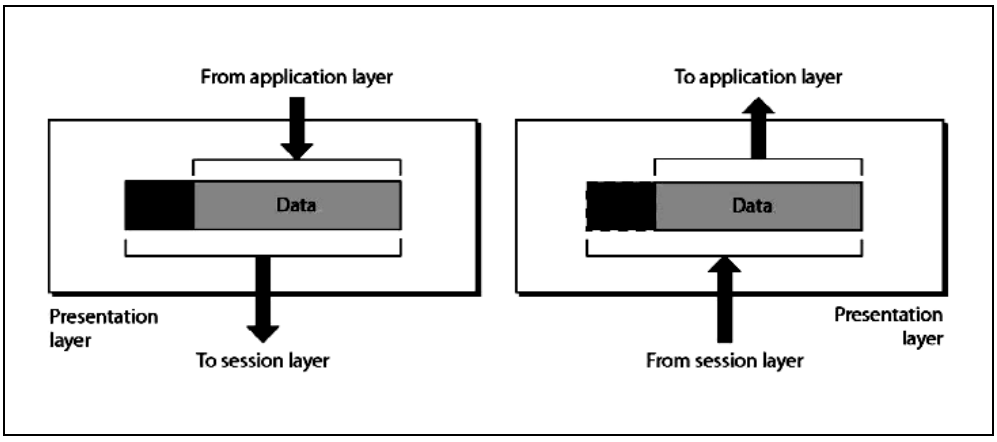


Fig : Presentation Layer

III. Translation

The sending and receiving devices may run on different platforms (hardware, software and operating system). Hence it is important that they understand the messages that are used for communicating. Hence a translation service may be required which is provided by the Presentation layers

IV. Compression

Compression ensures faster data transfer. The data compressed at sender has to be decompressed at the receiving end, both performed by the Presentation layer.

V. Encryption

It is the process of transforming the original message to change its meaning before sending it. The reverse process called decryption has to be performed at the receiving end to recover the original message from the encrypted message.

VI. **Main responsibility**

The main responsibility of the Presentation layer is translation, compression and encryption.

4.3.5.7 Application Layer

I. The application layer enables the user to communicate its data to the receiver by providing

certain services. For ex. Email is sent using X.400 service.

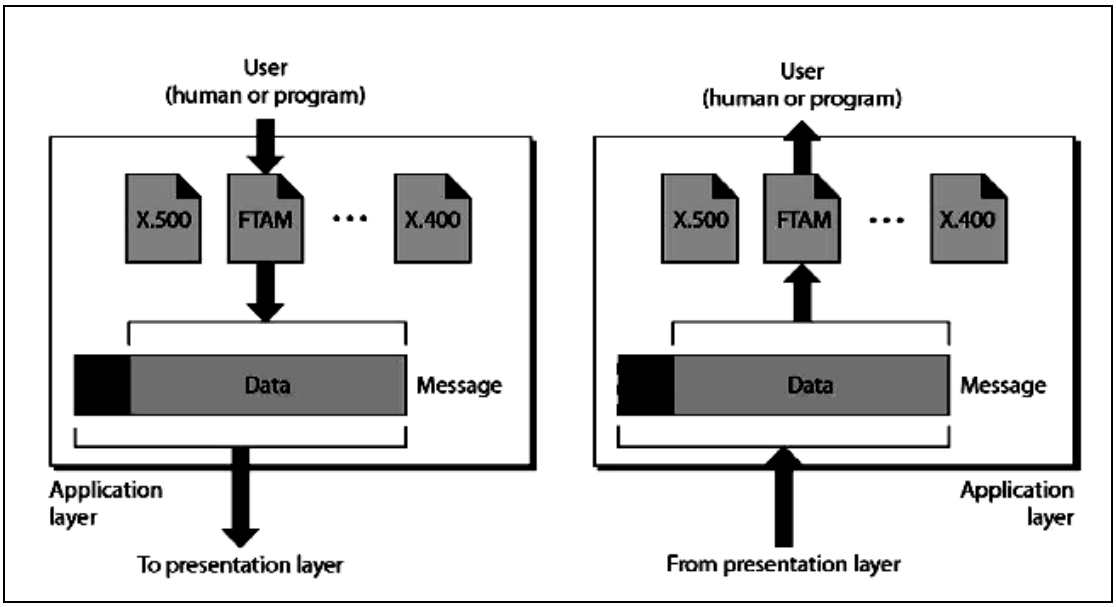


Fig : Application Layer

II. **X500** is a directory service used to provide information and access to distributed objects

III. **X400** is services that provides basis for mail storage and forwarding

IV. **FTAM (File transfer, access and management)** provides access to files stored on remote computers and mechanism for transfer and manage them locally.

V. Main Responsibility
Main Responsibility of Application layer is to provide access to network resources.

4.4 SUMMARY

The responsibilities of the 7 layers of OSI model can be summarized as follows:

1. Application Layer : To provide the users access to network resources
2. Presentation Layer: To provide the functions of translation, encryption and compression.
3. Session Layer: To establish, manage and terminate sessions

4. Transport Layer: To provide process to process delivery of message
5. Network Layer: To provide source to destination delivery of packets.
6. Datalink Layer: To provide hop to hop delivery of frames
7. Physical Layer: To transmit data over a bit stream from one hop to the next and provide electrical and mechanical specification.

4.5 REVIEW QUESTIONS

1. Explain the concept of layered task.
2. What is the OSI model? List its layers and explain their responsibility in exactly one line.
3. Explain how the communication takes place between layers of OSI model.
4. Write a short note on encapsulation of data in OSI model.
5. Differentiate between the working of Data link layer, Network layer and Transport layer.

4.6 REFERENCE & FURTHER READING

Data Communication & Networking – Behrouz Forouzan

5

TCP/IP MODEL, ADDRESSING IN TCP/IP – IPV4

Unit Structure

5.0 OBJECTIVES

- ✓ Understand the basics of TCP/IP model
- ✓ Understand the functions of the different layers and protocols involved
- ✓ Understand the Addressing mechanisms used under the TCP/IP
- ✓ Understand IPv4 and importantly IP address and IP header format

5.1 INTRODUCTION

After an understand of the concept of layered task and then understanding the OSI model we introduce the TCP/IP model. This model is currently being used on our systems. TCP/IP model is a collection of protocols often called a protocol suite. It offers a rich variety of protocols from which we can choose from.

5.2 TCP/IP MODEL

- It is also called as the TCP/IP protocol suite. It is a collection of protocols.
- IT is a hierarchical model, ie. There are multiple layers and higher layer protocols are supported by lower layer protocols.
- It existed even before the OSI model was developed.
- Originally had four layers (bottom to top):
    1. Host to Network Layer
    2. Internet Layer
    3. Transport Layer
    4. Application Layer
- The figure for TCP/IP model is as follows:

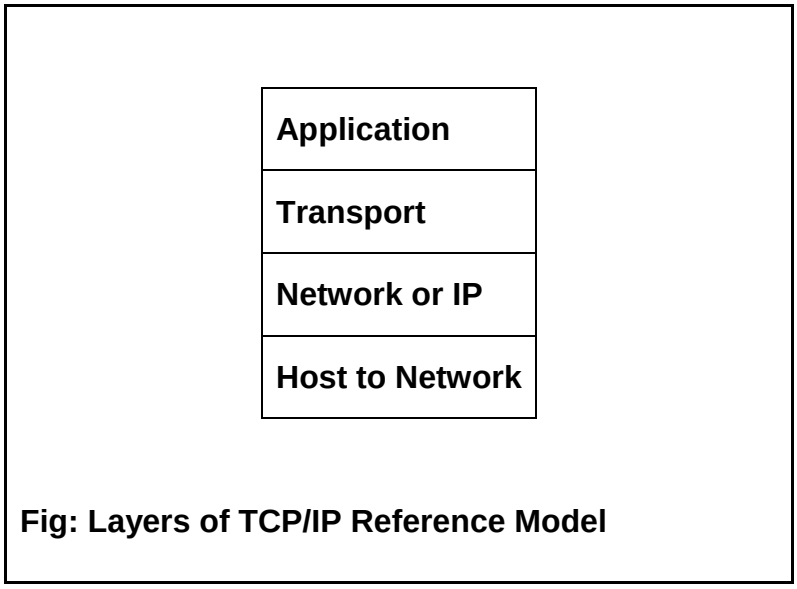


Fig: Layers of TCP/IP Reference Model

- The structure TCP/IP model is very similar to the structure of the OSI reference model. The OSI model has seven layers where the TCP/IP model has four layers.
- The Application layer of TCP/IP model corresponds to the Application Layer of Session, Presentation & Application Layer of OSI model.
- The Transport layer of TCP/IP model corresponds to the Transport Layer of OSI model
- The Network layer of TCP/IP model corresponds to the Network Layer of OSI model
- The Host to network layer of TCP/IP model corresponds to the Physical and Datalink Layer of OSI model.
- The diagram showing the comparison of OSI model and TCP/IP model along with the protocols is as shown below:

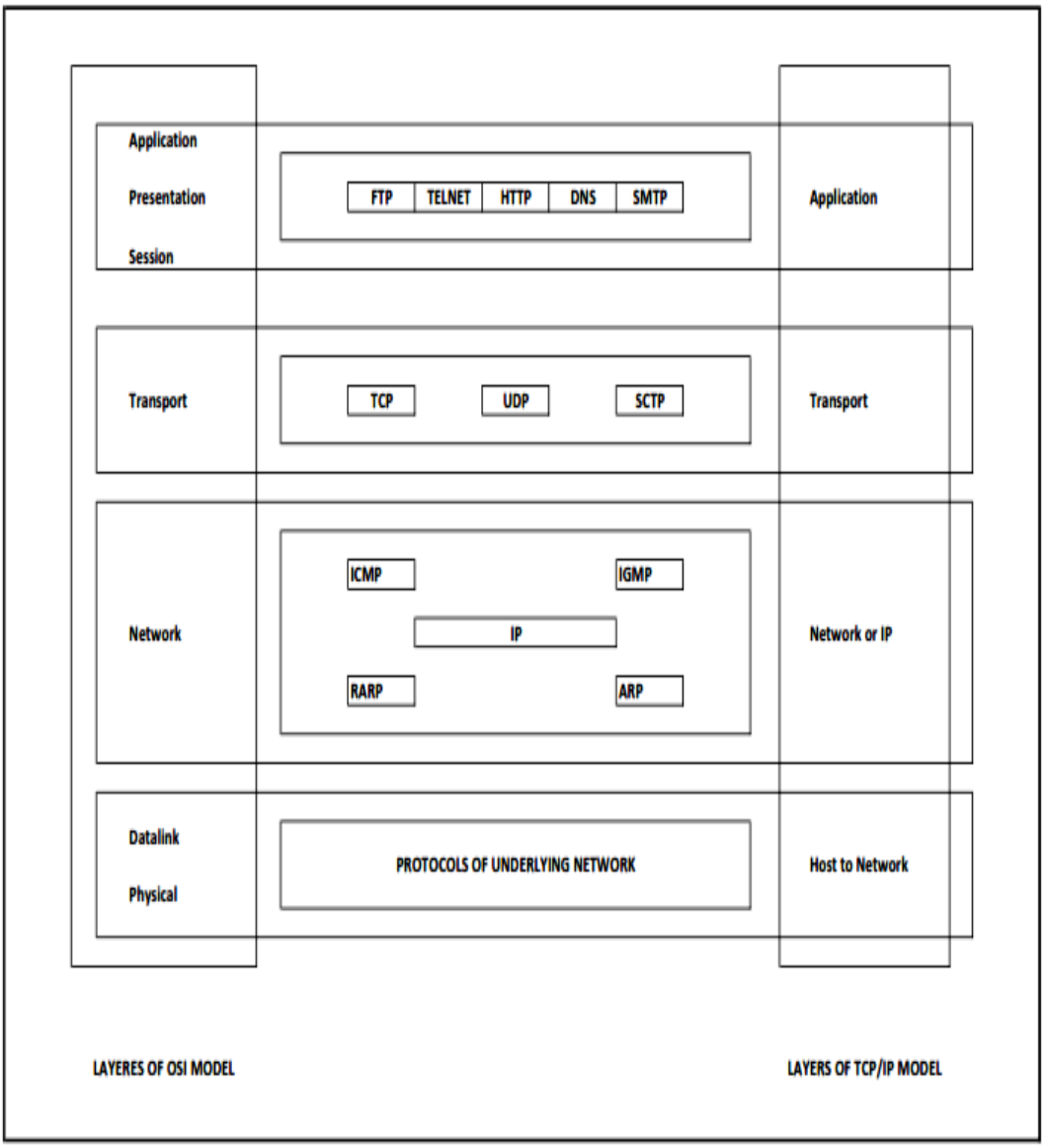


Fig: Comparison of OSI model and TCP/IP model

Functions of the Layers of TCP/IP model:

A. Host to Network Layer

This layer is a combination of protocols at the physical and data link layers.

It supports all standard protocols used at these layers.

B. Network Layer or IP

- Also called as the Internetwork Layer (IP). It holds the IP protocol which is a network layer protocol and is responsible for source to destination transmission of data.

- The Internetworking Protocol (IP) is an **connection-less** & **unreliable protocol.**

- It is a best effort delivery service. i.e. there is no error checking in IP, it simply sends the data and relies on its underlying layers to get the data transmitted to the destination.

- IP transports data by dividing it into **packets or datagrams** of same size. Each packet is independent of the other and can be transported across different routes and can arrive out of order at the receiver.

- In other words, since there is no connection set up between the sender and the receiver the packets find the best possible path and reach the destination. Hence, the word **connection-less**.

- The packets may get dropped during transmission along various routes. Since IP does not make any guarantee about the delivery of the data its call an **unreliable** protocol.

- Even if it is unreliable IP cannot be considered weak and useless; since it provides only the functionality that is required for transmitting data thereby giving maximum efficiency. Since there is no mechanism of error detection or correction in IP, there will be no delay introduced on a medium where there is no error at all.

- IP is a combination of four protocols:
    1. ARP
    2. RARP
    3. ICMP
    4. IGMP

1. ARP – Address Resolution Protocol

I. It is used to resolve the physical address of a device on a network, where its logical address is known.

II. Physical address is the 48 bit address that is imprinted on the NIC or LAN card, Logical address is the Internet Address or commonly known as IP address that is used to uniquely & universally identify a device.

2. RARP– Reverse Address Resolution Protocol

I. It is used by a device on the network to find its Internet address when it knows its physical address.

3. ICMP- Internet Control Message Protocol

I. It is a signaling mechanism used to inform the sender about datagram problems that occur during transit.

II. It is used by intermediate devices.
III. In case and intermediate device like a gateway encounters any problem like a corrupt datagram it may use ICMP to send a message to the sender of the datagram.

4. IGMP- Internet Group Message Protocol

I. It is a mechanism that allows to send the same message to a group of recipients.

C. Transport Layer

- Transport layer protocols are responsible for transmission of data running on a process of one machine to the correct process running on another machine.
- The transport layer contains three protocols:
    1. TCP
    2. UDP
    3. SCTP

1. TCP – Transmission Control Protocol

I. TCP is a reliable connection-oriented, reliable protocol. i.e. a connection is established between the sender and receiver before the data can be transmitted.

II. It divides the data it receives from the upper layer into segments and tags a sequence number to each segment which is used at the receiving end for reordering of data.

2. UDP – User Datagram Protocol

I. UDP is a simple protocol used for process to process transmission.

II. It is an unreliable, connectionless protocol for applications that do not require flow control or error control.

III. It simply adds port address, checksum and length information to the data it receives from the upper layer.

3. SCTP – Stream Control Transmission Protocol

I. SCTP is a relatively new protocol added to the transport layer of TCP/IP protocol suite.
II. It combines the features of TCP and UDP.
III. It is used in applications like voice over Internet and has a much broader range of applications

D. Application Layer

I. The Application Layer is a combination of Session, Presentation & Application Layers of OSI models and define high level protocols like File Transfer (FTP), Electronic Mail (SMTP), Virtual Terminal (TELNET), Domain Name Service (DNS), etc.

5.3 ADDRESSING IN TCP/IP

The TCP/IP protocol suited involves 4 different types of addressing:

1. Physical Address
2. Logical Address
3. Port Address
4. Specific Address

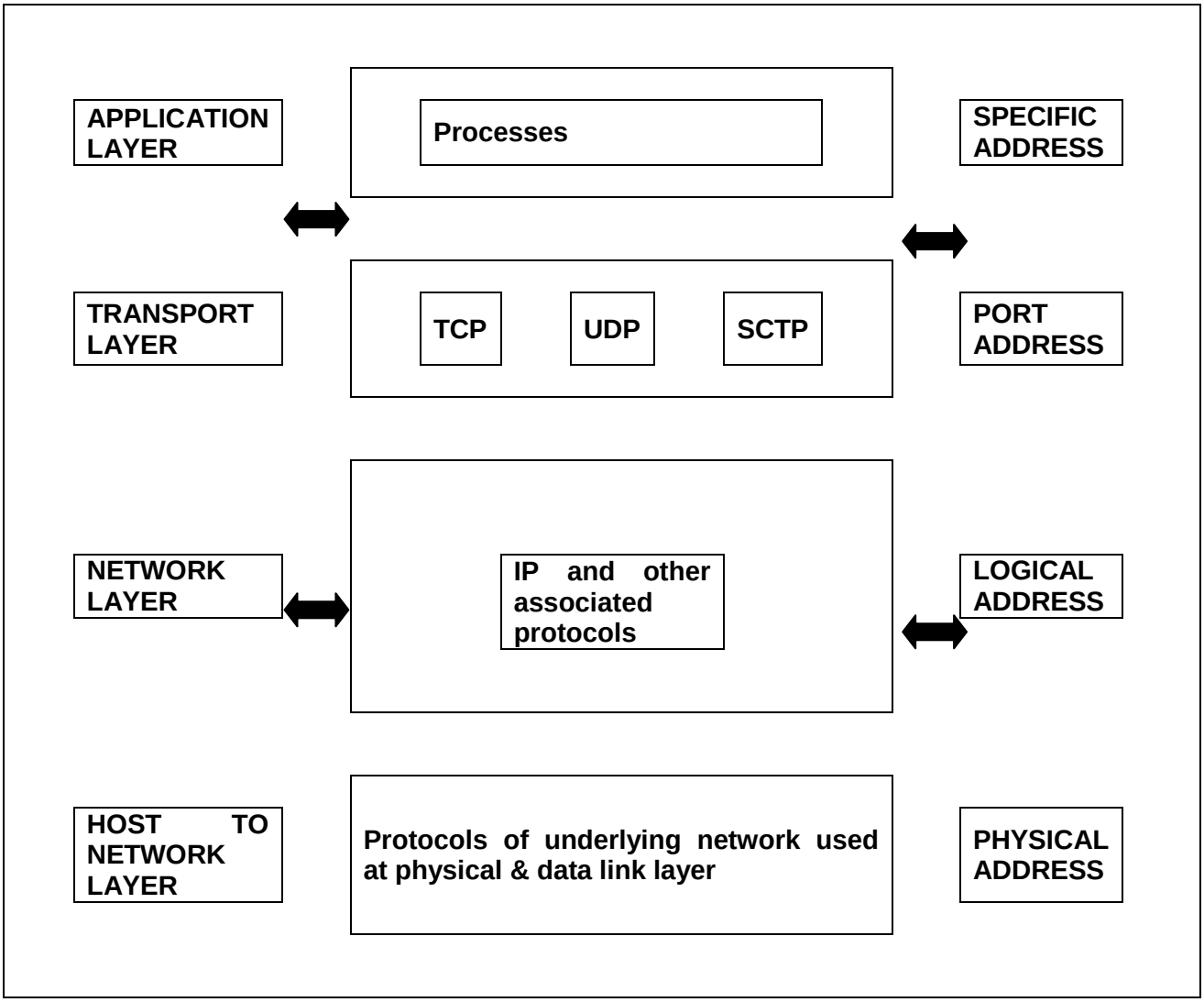


Fig: Addressing in TCP/IP model

Each of these addresses are described below:

1. **Physical Address**
   - **i.** Physical Address is the lowest level of addressing, also known as link address.
   - **ii.** It is local to the network to which the device is connected and unique inside it.
   - **iii.** The physical address is usually included in the frame and is used at the data link layer.
   - **iv.** MAC is a type of physical address that is 6 byte (48 bit) in size and is imprinted on the Network Interface Card (NIC) of the device.
   - **v.** The size of physical address may change depending on the type of network. Ex. An Ethernet network uses a 6 byte MAC address.

2. **Logical Address**
   - **i.** Logical Addresses are used for universal communication.
   - **ii.** Most of the times the data has to pass through different networks; since physical addresses are local to the network there is a possibility that they may be duplicated across multiples networks also the type of physical address being used may change with the type of network encountered. For ex: Ethernet to wireless to fiber optic. Hence physical addresses are inadequate for source to destination delivery of data in an internetwork environment.
   - **iii.** Logical Address is also called as IP Address (Internet Protocol address).
   - **iv.** At the network layer, device i.e. computers and routers are identified universally by their IP Address.
   - **v.** IP addresses are universally unique.
   - **vi.** Currently there are two versions of IP addresses being used:
     a. **IPv4**: 32 bit address, capable of supporting 2^{32} nodes
     b. **IPv6:** 128 bit address, capable of supporting 2^{128} nodes

3. **Port Address**
   - **VIII.** A logical address facilitates the transmission of data from source to destination device. But the source and the destination both may be having multiple processes communicating with each other.

Ex. Users A & B are chatting with each other using Google Talk, Users B & C are exchanging emails using Hotmail. The IP address will enable transmitting data from A to B, but still the data needs to be delivered to the correct process. The data from A cannot be given to B on yahoo messenger since A & B are communicating using Google Talk.

IX. Since the responsibility of the IP address is over here there is a need of addressing that helps identify the source and destination processes. In other words, data needs to be delivered not only on the correct device but also on the correct process on the correct device.

X. A Port Address is the name or label given to a process. It is a 16 bit address.

XI. Ex. TELNET uses port address 23, HTTP uses port address 80

4. Specific Address

i. Port addresses address facilitates the transmission of data from process to process but still there may be a problem with data delivery.

For Ex: Consider users A, B & C chatting with each other using Google Talk. Every user has two windows open, user A has two chat windows for B & C, user B has two chat windows for A & C and so on for user C

Now a port address will enable delivery of data from user A to the correct process (in this case Google Talk) on user B but now there are two windows of Google Talk for user A & C available on B where the data can be delivered.

ii. Again the responsibility of the port address is over here and there is a need of addressing that helps identify the different instances of the same process.

iii. Such address are user friendly addresses and are called specific addresses.

iv. Other Examples: Multiple Tabs or windows of a web browser work under the same process that is HTTP but are identified using **Uniform Resource Locators (URL**), Email addresses.

5.4 IP PROTOCOL – IPV4

Packets in the IPv4 format are called datagram. An IP datagram consists of a header part and a text part (payload). The header has a 20-byte fixed part and a variable length optional part. It is transmitted in big-endian order: from left to right, with the high-order bit of the Version field going first.
IPv4 can be explained with the help of following points:

1. IP addresses
2. Address Space
3. Notations used to express IP address
4. Classfull Addressing
5. Subnetting
6. CIDR
7. NAT
8. IPv4 Header Format

5.4.1 IP addresses

- Every host and router on the Internet has an IP address, which encodes its network number and host number.
- The combination is unique: in principle, no two machines on the Internet have the same IP address.
- An IPv4 address is 32 bits long
- They are used in the Source address and Destination address fields of IP packets.
- An IP address does not refer to a host but it refers to a network interface.

5.4.2 Address Space

- An address space is the total number of addresses used by the protocol. If a protocol uses *N* bits to define an address, the address space is 2^N because each bit can have two different values (0 or 1) and *N* bits can have 2^N values.
- IPv4 uses 32-bit addresses, which means that the address space is 2^{32} or 4,294,967,296 (more than 4 billion).

5.4.3 Notations

- There are two notations to show an IPv4 address:
    1. **Binary notation**
       The IPv4 address is displayed as 32 bits.
       ex. 11000001 10000011 00011011 11111111

2. **Dotted decimal notation**
   To make the IPv4 address easier to read, Internet addresses are usually written in decimal form with a decimal point (dot) separating the bytes.

   Each byte (octet) is 8 bits hence each number in dotted-decimal notation is a value ranging from 0 to 255.
   Ex. 129.11.11.239

5.4.4 Classful addressing

In classful addressing, the address space is divided into five classes: A, B, C, D, and E.

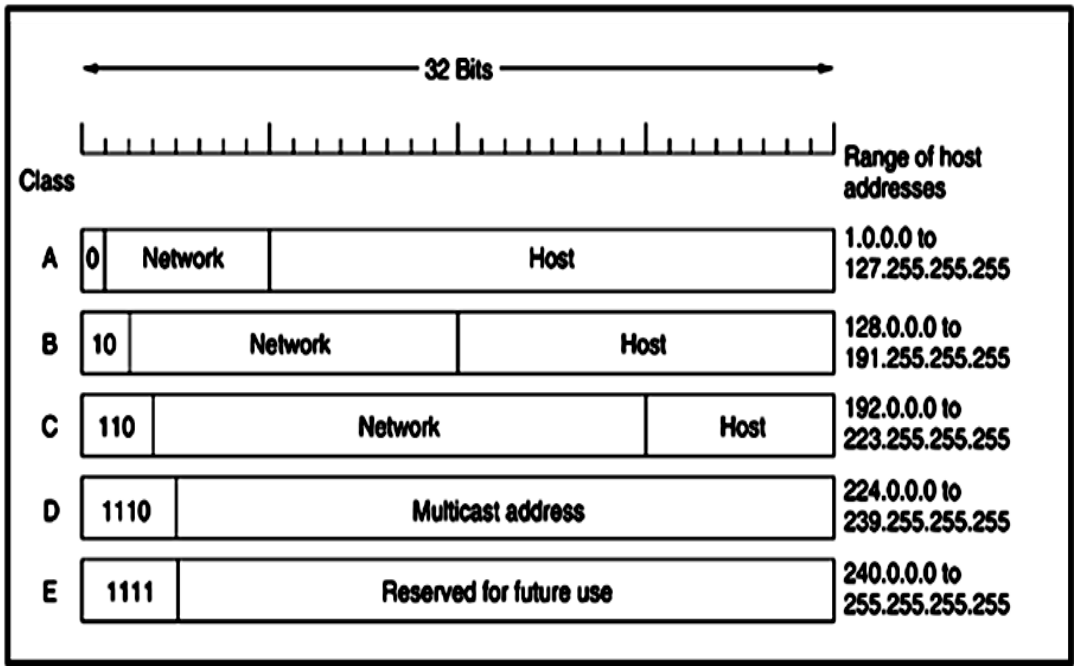


Figure: Classful addressing : IPv4

Netid and Hostid

- In classful addressing, an IP address in class A, B, or C is divided into netid and hostid.

- These parts are of varying lengths, depending on the class of the address as shown above.

Information on the Number of networks and host in each class is given below:

Class	Number of Networks	Number of Hosts	Application
A	128	16,777,216	Unicast
B	16,384	65,536	Unicast
C	2,097,152	256	Unicast
D	1	268,435,456	Multicast
E	1	268,435,456	Reserved

- The IP address 0.0.0.0 is used by hosts when they are being booted.

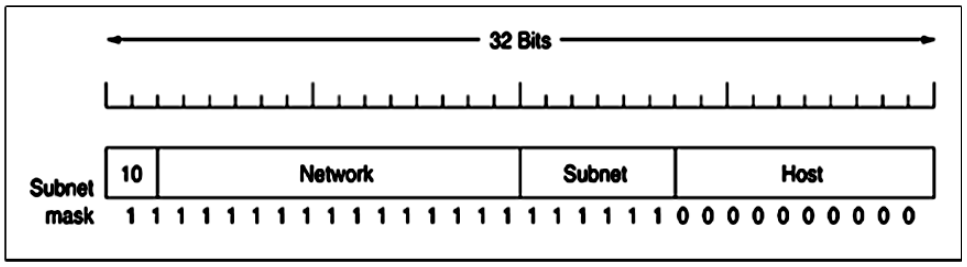
- All addresses of the form 127.xx.yy.zz are reserved for loopback testing, they are processed locally and treated as incoming packets.

5.4.5 Subnetting

- It allows a network to be split into several parts for internal use but still act like a single network to the outside world.

- To implement subnetting, the router needs a subnet mask that indicates the split between network + subnet number and host. Ex. 255.255.252.0/22. A"/22" to indicate that the subnet mask is 22 bits long.

- Consider a class B address with 14 bits for the network number and 16 bits for the host number where some bits are taken away from the host number to create a subnet number.



4Fig: A Class B network subnetted into 64 subnets.

- If 6 bits from the host Id are taken for subnet then available bits are :
  14 bits for network + 6 bits for subnet + 10 bits for host
- With 6 bits for subnet the number of possible subnets is 2^6 which is 64.
- With 10 bits for host the number of possible host are 2^{10} which is 1022 (0 & 1 are not available)

5.4.6 CIDR

A class B address is far too large for most organizations and a class C network, with 256 addresses is too small. This leads to granting Class B address to organizations who do not require all the address in the address space wasting most of it.

This is resulting in depletion of Address space.

A solution is CIDR (Classless InterDomain Routing) The basic idea behind CIDR, is to allocate the remaining IP addresses in variable-sized blocks, without regard to the classes.

5.4.7 NAT (Network Address Translation)

- The scarcity of network addresses in IPv4 led to the development of IPv6.

- IPv6 uses a 128 bit address, hence it has 2^{128} addresses in its address space which is larger than 2^{32} addresses provided by IPv4.
- Transition from IPv4 to IPv6 is slowly occurring, but will take years to complete, because of legacy hardware and its incompatibility to process IPv6 address.

- NAT (Network Address Translation) was used to speed up the transition process

- The only rule is that no packets containing these addresses may appear on the Internet itself. The three reserved ranges are:
  10.0.0.0 – 10.255.255.255/8 (16,777,216 hosts)
  172.16.0.0 – 172.31.255.255/12 (1,048,576 hosts)
  192.168.0.0 – 192.168.255.255/16 (65,536 hosts)

- **Operation:**
  Within the Organization, every computer has a unique address of the form 10.x.y.z. However, when a packet leaves the organization, it passes through a NAT box that converts the internal IP source address, 10.x.y.z, to the organizations true IP address, 198.60.42.12 for example.

5.4.8 IP Header

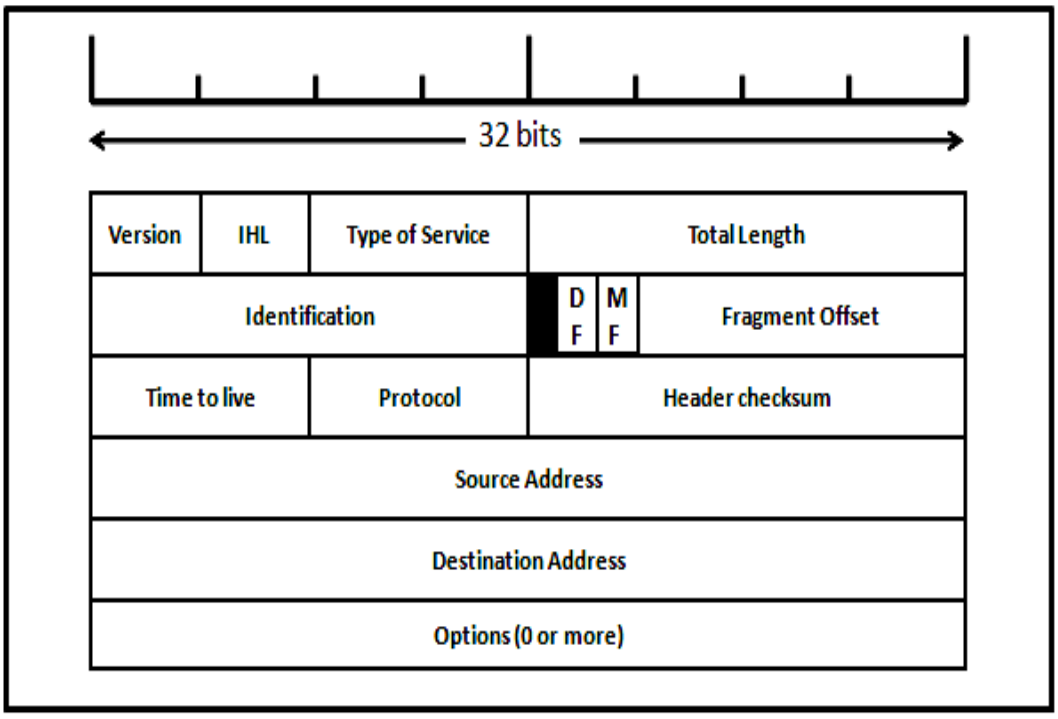


Figure: The IPv4 (Internet Protocol) header

The description of the fields shown in the diagram is as follows:

No	Field Name	Description
1	Version	Keeps track of the version of the protocol the datagram belongs to (IPV4 or IPv6)
2	IHL	Used to indicate the length of the Header. Minimum value is 5 Maximum value 15
3	Type of service	Used to distinguish between different classes of service
4	Total length	It includes everything in the datagram—both header and data. The maximum length is 65,535 bytes
5	Identification	Used to allow the destination host to identify which datagram a newly arrived fragment belongs to. All the fragments of a datagram contain the same Identification value

6	DF	1 bit field. It stands for Don't Fragment. Signals the routers not to fragment the datagram because the destination is incapable of putting the pieces back together again
7	MF	MF stands for More Fragments. All fragments except the last one have this bit set. It is needed to know when all fragments of a datagram have arrived.
8	Fragment offset	Used to determine the position of the fragment in the current datagram.
9	Time to live	It is a counter used to limit packet lifetimes. It must be decremented on each hop. When it hits zero, the packet is discarded and a warning packet is sent back to the source host.
10	Header checksum	It verifies Header for errors.
11	Source address	IP address of the source
12	Destination address	IP address of the destination
13	Options	The options are variable length. Originally, five options were defined: 1. Security : specifies how secret the datagram is 2. Strict source routing : Gives complete path to be followed 3. Loose source routing : Gives a list of routers not to be missed 4. Record route: Makes each router append its IP address 5. Timestamp: Makes each router append its IP address and timestamp

5.5 SUMMARY

1. TCP/IP has 4 layers: host to network, IP, Transport & Application
2. It uses 4 levels of address: physical, logical, port & specific

3. IP address uniquely identifies a device on the Internet.

5.6 REVIEW QUESTIONS

6. Explain the structure of TCP/IP protocol
7. Explain in short the functions of every layer of TCP/IP
8. Explain the function of every protocol of the IP layer
9. Explain the concept of IP addresses in detail
10. Why do we use subnetting?
11. What Is NAT? why is it used for?
12. Explain the header of and IPv4 Packet.

5.7 REFERENCE & FURTHER READING

Data Communication & Networking – Behrouz Forouzan

6

INFORMATION ENCODING

Unit Structure

1. Data Communication & Networking – BehrouzForouzan

6.0 OBJECTIVES

- ✓ **Introduce representing different symbols**
- ✓ **Introduction to error**
- ✓ **Multimedia**
- ✓ **Sampling**
- ✓ **Quantization**

6.1 INTRODUCTION

In communications and information processing, **encoding** is the process by which information from a source is converted into symbols to be communicated.

In computer system we have to use encoding to represent the Information in the format that computer understand i.e. “binary” language.

Encoding enables us to improve the communication in places where written languages is difficult to use or impossible.

6.2 REPRESENTING DIFFERENT SYMBOLS

In information theory and computer science, a code is usually considered as an algorithm which uniquely represents symbols from some source alphabet, by *encoded* strings, which may be in some other target alphabet.

The purpose of the symbols is to communicate the idea or meaning. We use the different symbols to represent the Information in computer understandable format.

There are different symbols available e.g. in English language we have 26 capital letters from “A to Z” and same way we have small letters from “a to z”, we have numeric symbols like (0,1,…9) and special symbols like (!,@,#,$,%,^,&,*, etc.)

6.3 MINIMIZING ERRORS

One way to represent the information is to use sound beep. The different sound intensities can be utilized to represent the around 162 different symbols.

Practically we won’t be able to distinguish between all 162 sound levels. So it will lead to errors in identifying a symbol correctly. We should follow the representation that minimizes the errors.

So another way of doing work is just use two states OFF or ON. With sound beep we get some sound or none at all. When it is no sound then the value must be 0. When we hear some sound then value must be 1.

With this beep and no-beep we can represent only two symbols correctly. We call this system as binary system.

We can have the following table to represent our symbols:

Sound	**Represented symbols**	**Code**
No-beep	A	0
Beep	B	1

We must note that the system which is explained above, even though the error in representation of symbols is minimized, is restricted with two symbols. So let us see how to represent more symbols using the binary system.

6.3.1 Representing more symbols:

Now we see what happens when we use two sound devices. As shown in the following table we get the different combinations like 00, 01, 10 and 11. And now we can represent four different symbols.

Sound device 1	Sound device 2	Represented symbols	Code
No-beep	No-beep	A	0 0
No-beep	Beep	B	0 1
Beep	No-beep	C	1 0
Beep	Beep	D	1 1

Thus with the pair of sound devices, we get four ON/OFF combinations and hence we can represent four symbols.

If we extend the same technique with three sound devices then we can get eight different codes to represent our symbols like (000,001,010,011,100,101,110 and 111)

The generalized we can say that with N number of sound devices we can have 2^N symbols to be represented.

Number of states : 2^N

Where N is number of sound devices.

6.4 Multimedia:

Now a day the computers that we use have additional facilities such as:

1. Drawing, capturing, storing and viewing pictures of different formats.

2. Recording, storing the sound/songs and playing them back.

3. Capturing, editing, storing the video information and playing them back.

Since video, pictures are not made of alphabets; we cannot represent them with the help of the character codes like ASCII (American Standard Code for Information Interchange) or EBCDIC (Extended Binary Coded Decimal Interchange Code) etc.

So we use the concept of multimedia in order to code the picture or videos.

With the help of multimedia technique it’s possible to use the computer system to store, play and process the audio, video and picture information along with the textual data.

6.4.1 Pictures/Images:

We can represent the data in the form of Images / Pictures. Images are represented by the pixels i.e. the smallest element in the picture.

The basic idea is shown in the following figure:

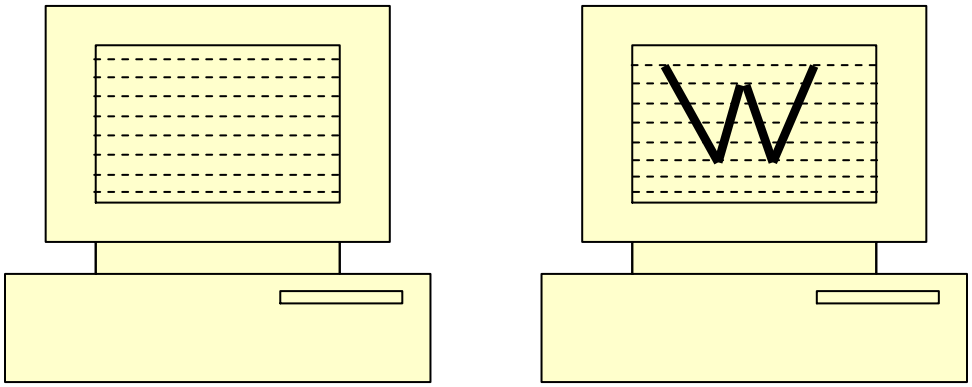


Figure (a) shows the computer screen made up of number of dots (pixels). Figure (b) shows the letter W by illuminating specific dots.

We can use a large number of pixels for better resolution. Higher resolution gives better quality to the picture.

When we divide a picture into pixels, each pixel is represented by a unique pattern.

If Image is black and white then we can use only one bit per pattern. Bit 1 is used to represent white and bit 0 is used to represent black.

If the image is having gray shade then we can use 2 bit pattern to represent each pixel. So 00 is black, 01 is dark gray, 10 is light gray and 11 will represent white.

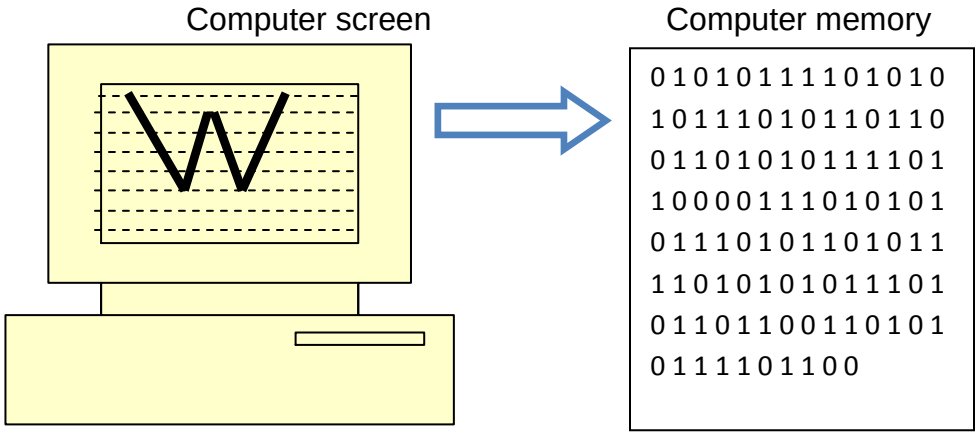
If the image is colored then it is to be represented by the pixels that are constituted of the three primary colorsnamely red, green and blue (RGB).

It means that any picture can be drawn on screen by illuminating or darkening specific pixels.

We know that computer can understand on binary values (0 and 1). So it is important that the way in which we make computer understand the concept of pixels.

We can consider the coding scheme where illuminated pixel is considered to be binary 1 and darkened pixel is considered to be binary 0. We can imagine that any picture that we draw on screen can be mapped first to a series of pixels. Which in turn, get mapped to a series of zeros and ones.

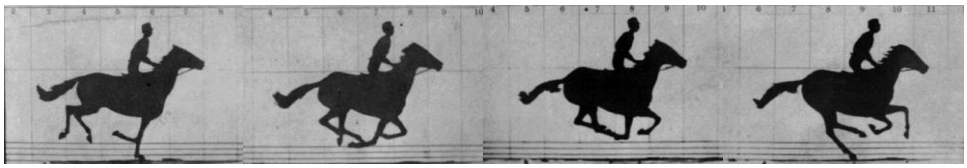
We can show this idea in following figure:



6.4.2 Video:

Animation is used as a basic technique for creating videos.

We get the animation if we show set of pictures rapidly, the human eye gets an illusion that the picture is in motion.



The above pictures show the idea behind the video. It shows the movements of the horse and his rider. If you were shown these four pictures one after the other very fast, you would believe that the horse is actually running.

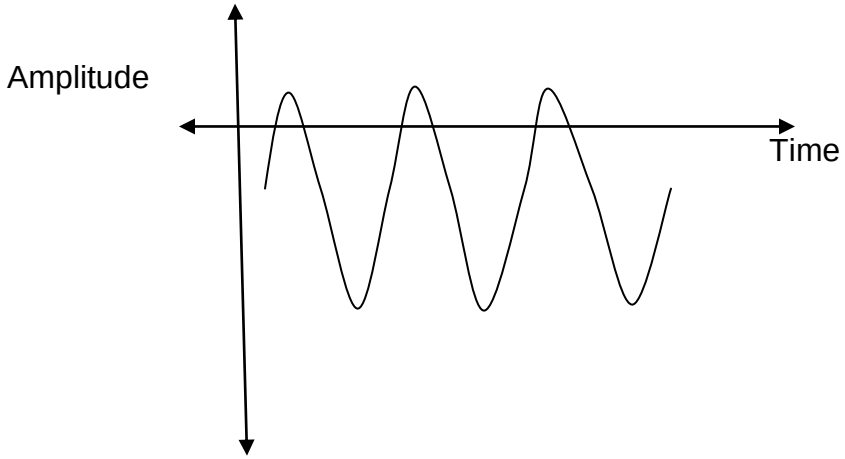
We use the same principle to store pictures in the disk / memory of the computer in their binary form and show them rapidly at the rate of 24 such pictures or images per second on the screen.

6.4.3 Sound:

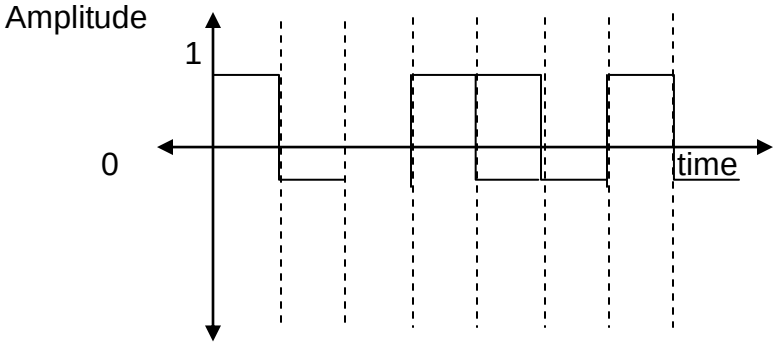
A sound wave in its most basic form continuous in nature. It is continuous in two aspects:

First, the strength (the amplitude) and time.

A typical sound signal takes the form of sine wave as shown in following figure:



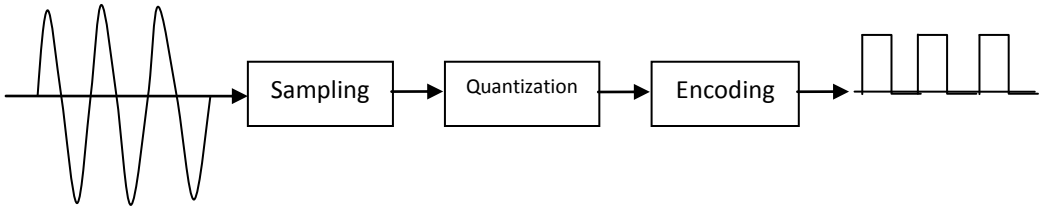
The sound needs to be converted into the digital form in order to store it in the computer system. Thus if we have to show the zeros and ones inside a computer's memory graphically, then we get the following figure:



In order to convert the sound into digital form, we have carry out the following processes on the analog sound signal:

1. Sampling
2. Quantization
3. Encoding

These processes are collectively called as Pulse Code Modulation (PCM). The following figure shows all the processes:



Analog sound Digital sound

6.4.3.1 Sampling:

When we transfer the signal using pulse code modulation and digital modulation the signal must be in the discrete time form.

If the message is generated from the computer system or any other digital source then it is in the proper form for processing by the digital communication system.

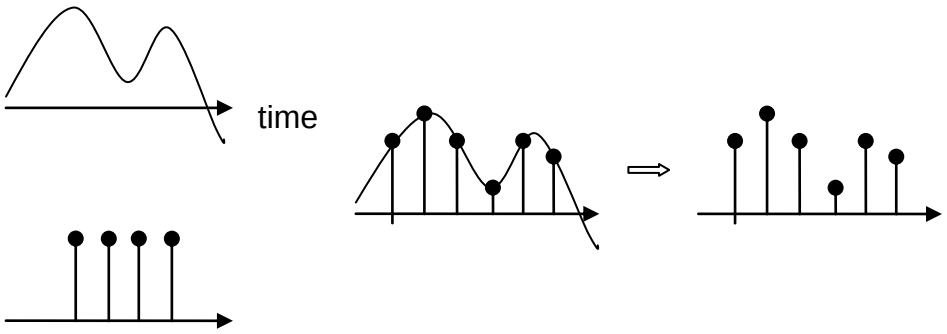
But in real life the signal can be of analog type (e.g. voice). In such a case it has to be first converted into discrete time signal.

For this we use “Sampling” method. Thus using the sampling process we convert the continuous time signal (analog) into the discrete time signal (digital).

The sampling process should satisfy the following requirements:

1. Sampled signal should represent the original signal.
2. It should be possible to reconstruct the original signal from its sampled form.

The following figure explains the sampling process:



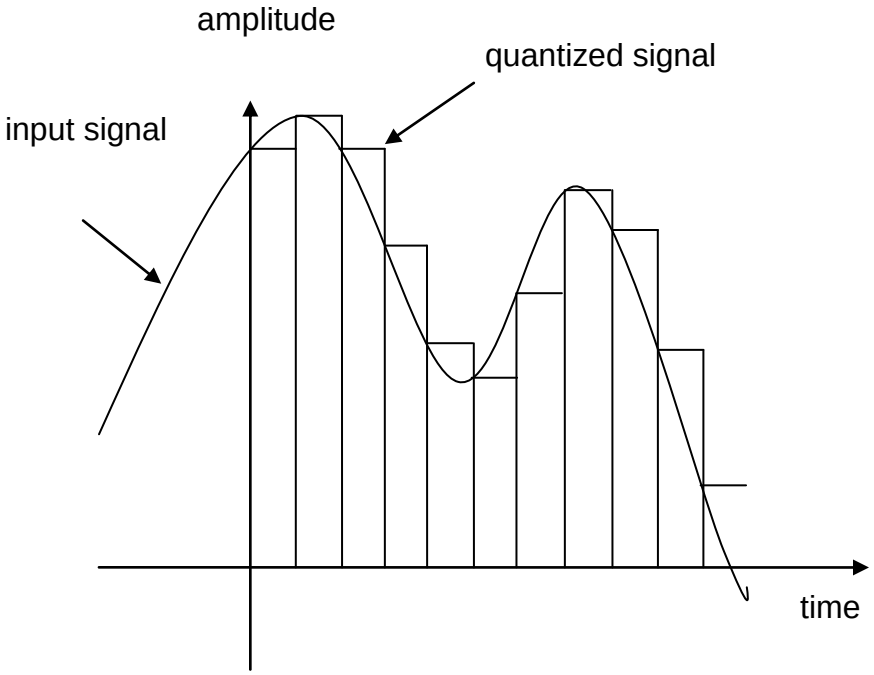
6.4.3.2 Quantization:

Quantization is the process in which we assign the numbers to the discrete values depending upon their amplitude values.

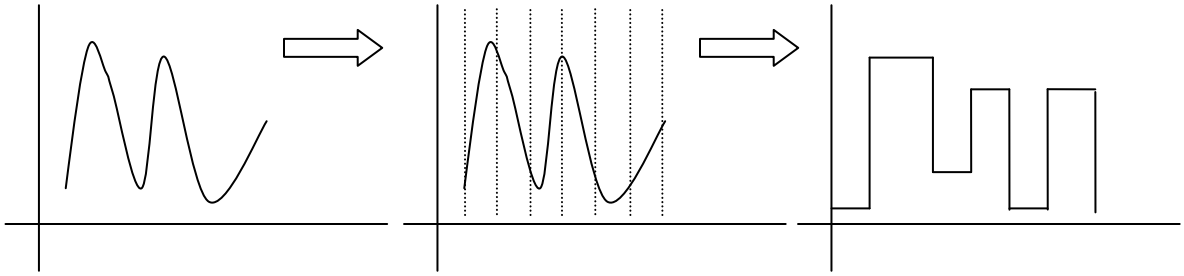
Quantizer converts the sampled signal into an approximate quantized signal which consists of only finite number of predecided voltage levels.

Each sampled value at the input of the quantizer is approximated or rounded to the nearest standard predecided voltage level (Quantization levels).

The following figure shows the overview of the process:



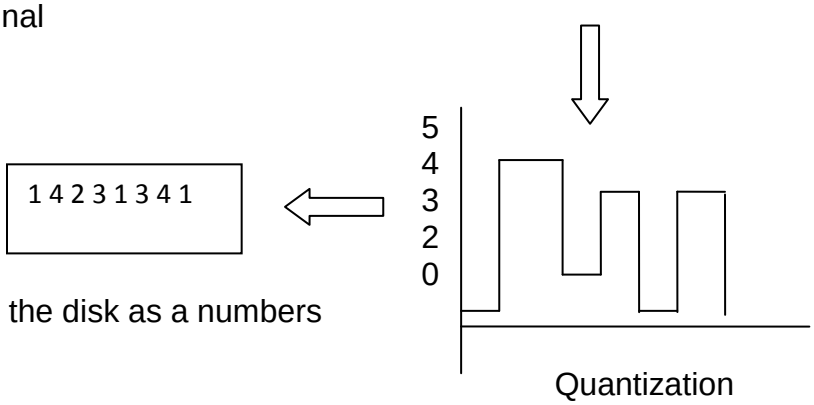
We can show the sampling and quantization process as follows:



Original signal
for the signal

Sampling

Measuring



Stored on the disk as a numbers

Quantization

6.5 MULTIMEDIA AND DATA COMPRESSION

When we store the information in the form of pictures, videos and sounds there would be lot of repetition that can be observed.

Storing such repeated information would cause the wastage of computer memory.

We must find the better schemes that eliminate the duplication of the data or redundant information.

There are various multimedia file formats available which allow storage of multimedia files in more efficient manner by getting rid of duplication/ redundancy by means of the process called as **Data Compression.**

6.6 REVIEW QUESTIONS

8. Define multimedia
9. What is quantization
10. What is sampling and explain it's importance

6.7 REFERENCES

1. Data Communication & Networking – BehrouzForouzan

7

ERRORS, DETECTION & CORRECTION

Unit Structure

7.0 OBJECTIVE

- ✓ Understand error classification
- ✓ Types of error
- ✓ Understand concept redundancy
- ✓ Hamming code concept
- ✓ CRC concept
- ✓ Checksum technic

7.1 INTRODUCTION:

Errors in the data are basically caused due to the various impairments that occur during the process of transmission.

When there is an imperfect medium or environment exists in the transmission it prone to errors in the original data.

Errors can be classified as follows:

Attenuation
Noise
Distortion

7.2 ERRORS CLASSIFICATION

Following are the categories of the errors:

1. Attenuation:

As signal travels through the medium, its strength decreases as distance increases, as shown in the figure 7.1, the example is voice, it becomes weak over the distance and loses its contents beyond a certain distance. As the distance increases attenuation also increases.

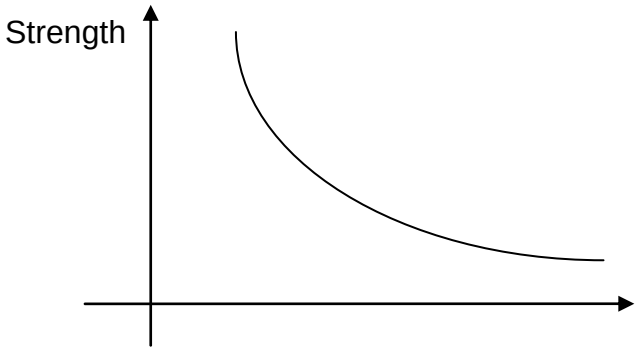


Figure 7.1

2. Noise:

Noise is defined as an unwanted data. When some electromagnetic signal gets inserted during the transmission, it is generally called as a Noise. Due to Noise it is difficult to retrieve the original data or information.

3. Distortion:

When there is an interference of the different frequencies who travel across the medium with the different speed, Distortion occurs. So it is important to have a space (guard space) between the different frequencies.

7.3 TYPES OF ERRORS:

If the signal comprises of binary data there can be two types of errors which are possible during the transmission:

1. Single bit errors
2. Burst Errors

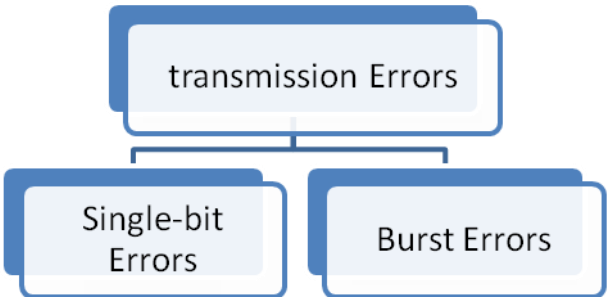


Figure 7.2

1. Single-bit errors:

In single-bit error, a bit value of 0 changes to bit value 1 or vice versa. Single bit errors are more likely to occur in parallel transmission. Figure 7.3 (a)

2. Burst errors:

In Burst error, multiple bits of the binary value changes. Burst error can change any two or more bits in a transmission. These bits need not be adjacent bits. Burst errors are more likely to occur in serial transmission. Figure 7.3 (b)

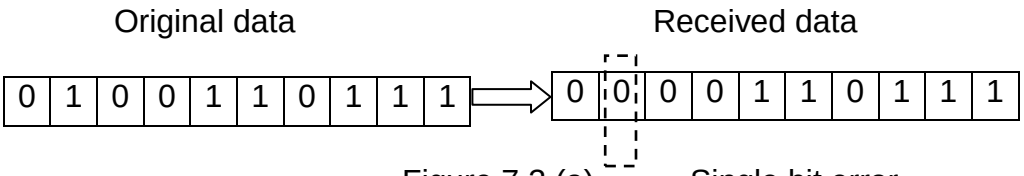


Figure 7.3 (a) Single bit error

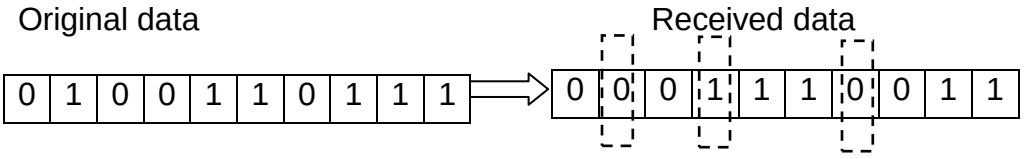


Figure 7.3 (b) Burst error

7.4 REDUNDANCY

In order to detect and correct the errors in the data communication we add some extra bits to the original data. These extra bits are nothing but the redundant bits which will be removed by the receiver after receiving the data.

Their presence allows the receiver to detect or correct corrupted bits. Instead of repeating the entire data stream, a short group of bits may be attached to the entire data stream.This technique is called redundancy because the extra bits are redundant to the information: they are discarded as soon as the accuracy of the transmission has been determined.

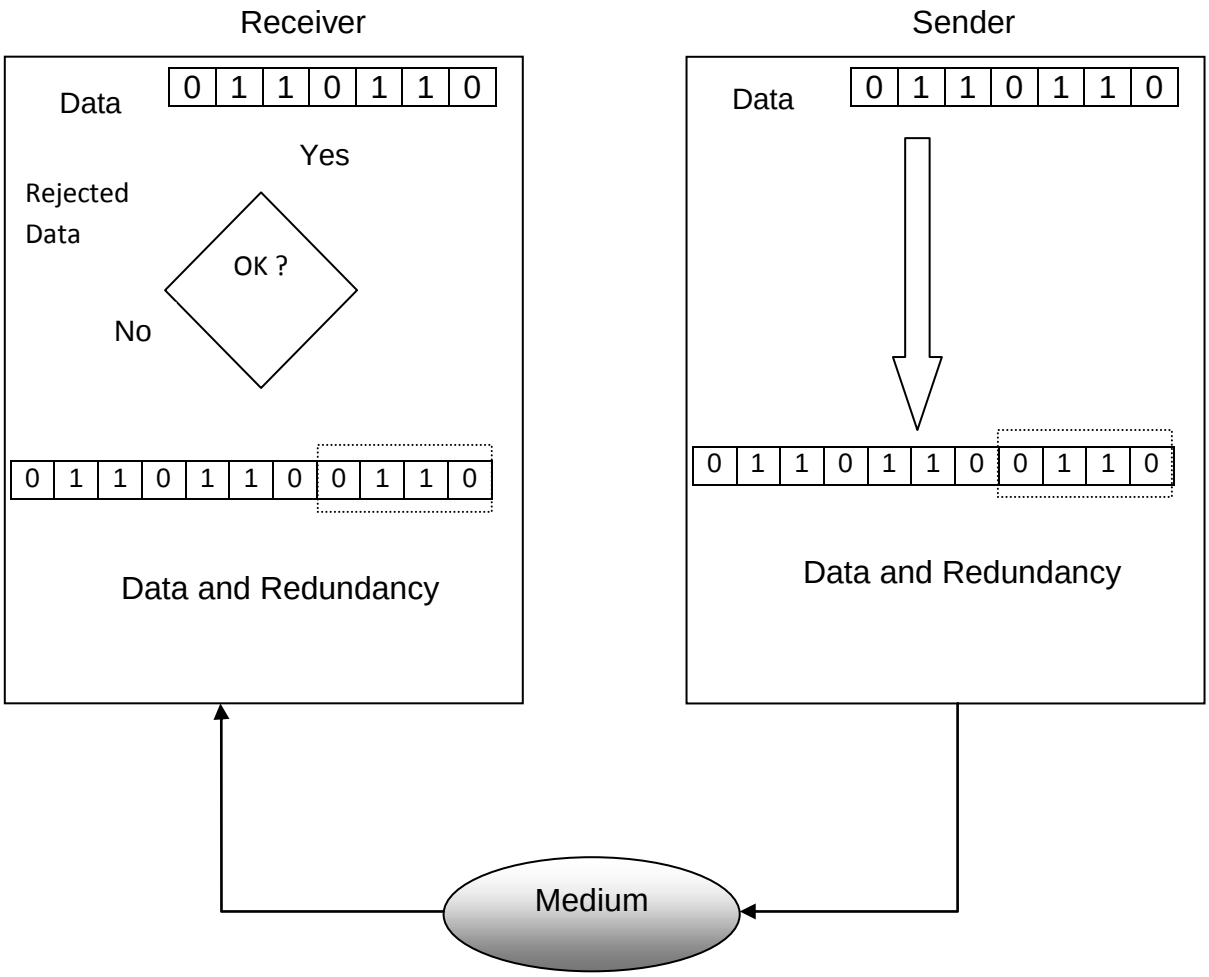


Figure 7.4

There are different techniques used for transmission error detection and correction.

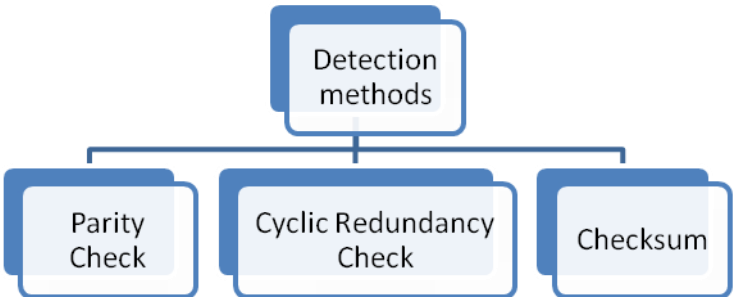
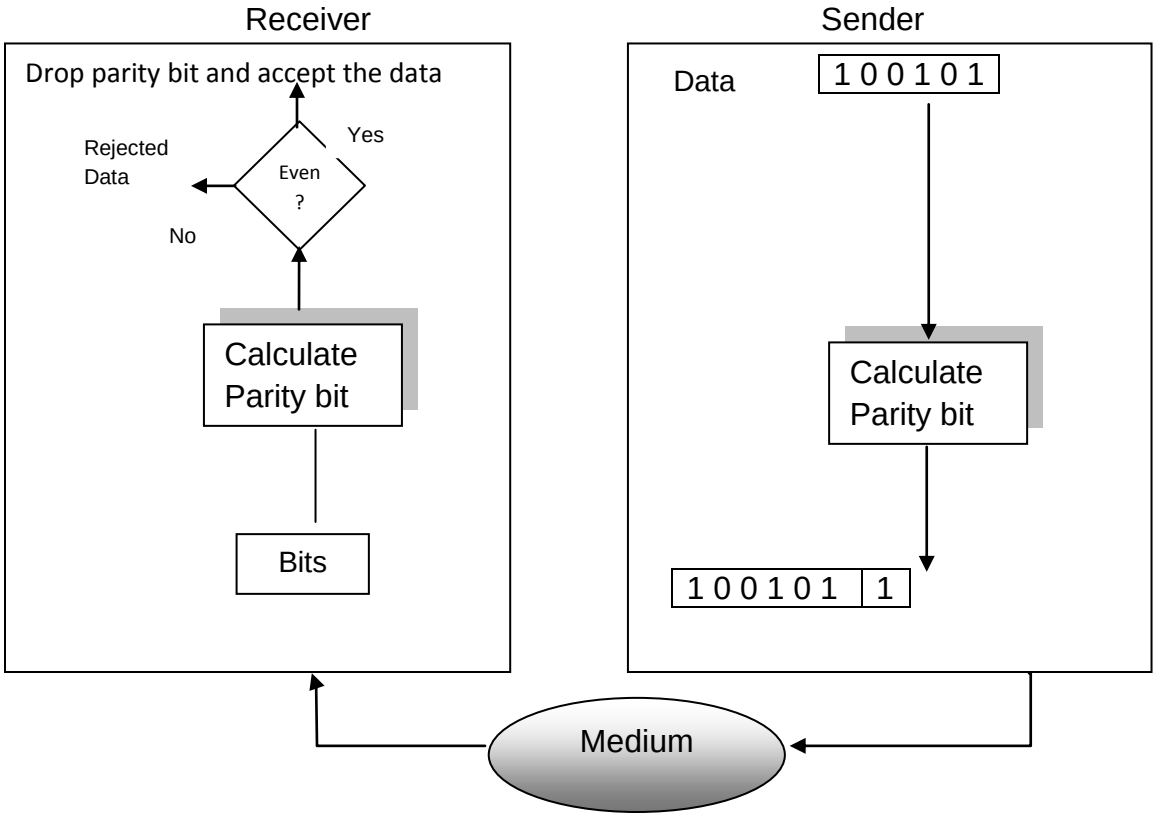


Figure 7.5

2. Parity Check:

In this technique, a redundant bit called a parity bit is addedto every data unit so that the total number of 1's in the unit(including the parity bit) becomes even (or odd). Following

Figure shows this concept when transmit the binary data unit110101.

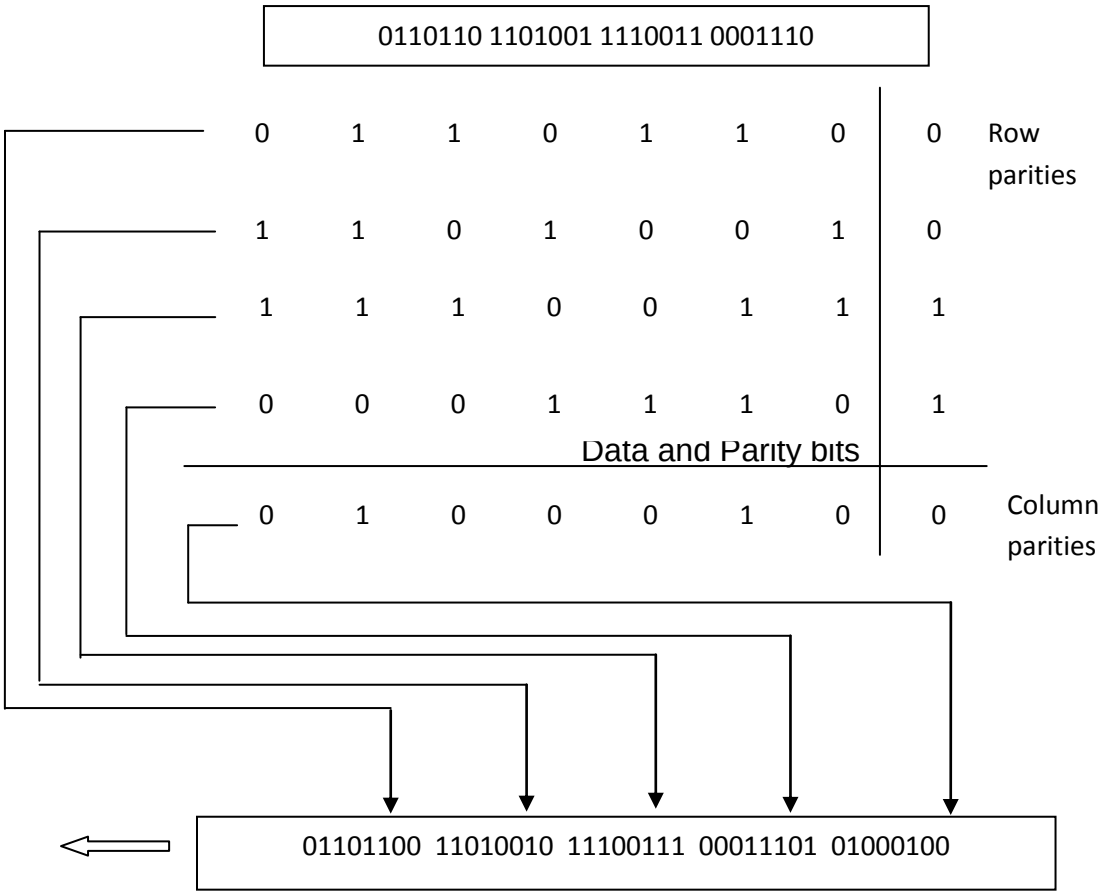


Simple parity check can detect all single-bit errors. It canalso detect burst errors as long as the total number of bitschanged is odd. This method cannot detect errors where thetotal number of bits changed is even.

Two-Dimensional Parity Check:

A better approach is the two dimensional parity checks. In this method, a block of bits is organized in a table (rows and columns). First we calculate the parity bit for each data unit. Then we organize them into a table. We then calculate the parity bit for eachcolumn and create a new row of 8 bits.

Consider the following example; we have four data units to send. They are organized in the tabular form as shown below.



We then calculate the parity bit for eachcolumn and create a new row of 8 bits; they are the parity bits for the whole block. Note that the first parity bit in thefifth row is calculated based on all first bits: the secondparity bit is calculated based on all second bits: and so on.We then attach the 8 parity bits to the original data andsend them to the receiver.

Two-dimensional parity check increases the likelihood ofdetecting burst errors.A burst error of more than ‘n’bits is also detected by thismethod with a very high probability.

3. Cyclic Redundancy Check (CRC)

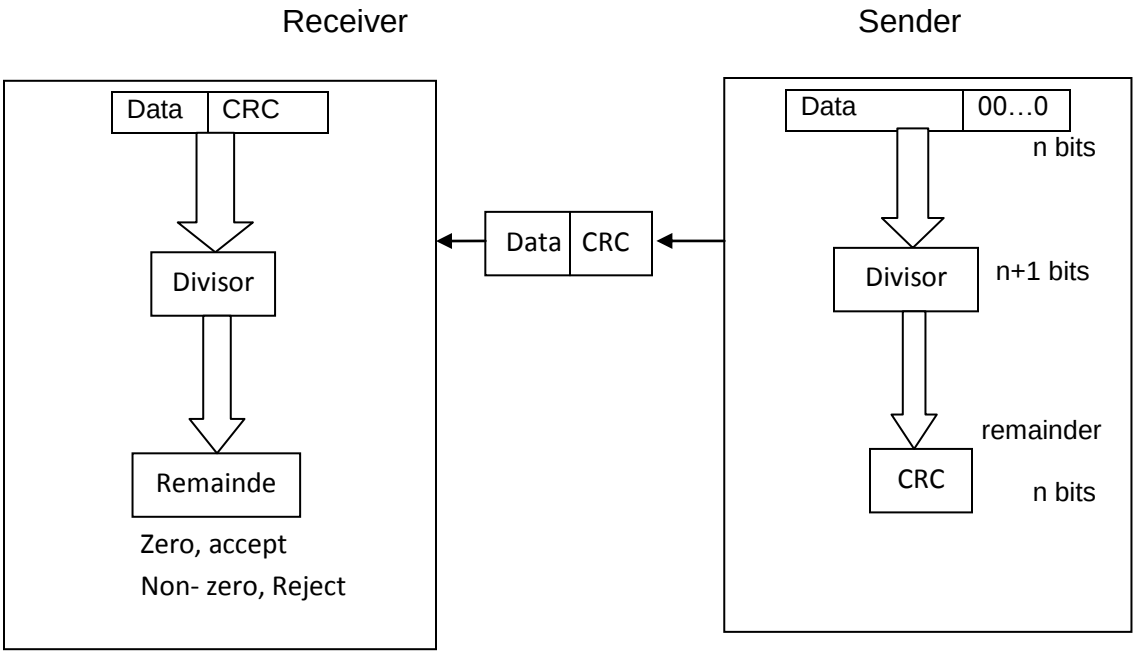
Most powerful of the redundancy checkingtechniques is the cyclic redundancy check (CRC). This method is based on the binary division. In CRC, the desired sequence of redundant bits are generated and is appended to the end of data unit. It is also called as CRC reminder. So that the resulting data unit becomes exactly divisible by a predetermined binary number.

At its destination, theincoming data unit is divided by the same number. If at thisstep there is no remainder then the data unit

is assumed to be correct and is therefore accepted.A remainder indicates thatthe data unit has been damaged in transit and thereforemust be rejected.

The redundancy bits used by CRC are derived by dividingthe data unit by a predetermined divisor; the remainder isthe CRC.To be valid, a CRC must have two qualities: It musthave exactly one less bit than the divisor, and appending itto the end of the data string must make the resulting bitsequence exactly divisible by the divisor.

The following figure shows the process:



Step1: A string of 0's is appended to the data unit. It is n bits long. The number n is 1 less if-number of bits in the predetermined divisor which is n + 1 bits.

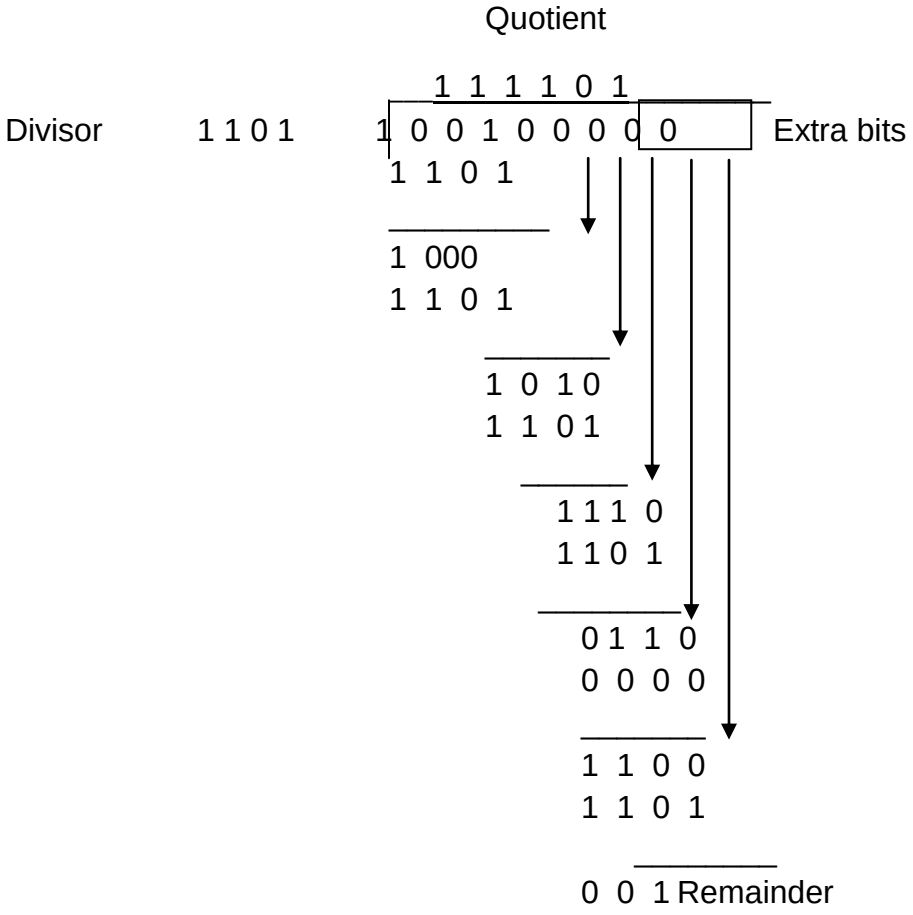
Step 2: The newly generated data unit is divided by the divisor, using a process called as binary division. The remainder resulting from this division is the CRC.

Step 3: the CRC of n bits derived in step 2 replaces the appended 0's at the data unit. Note that the CRC may consist of all 0's.

The data unit arrives at the receiver data first, followed by the CRC. The receiver treats the whole string as a unit and divides it by the same divisor that was used the CRC remainder. If the string arrives without error, the CRC checker yields a remainder of zero, the data unit passes. If the string has been changed in transit, the division yields zero remainder and the data unit does not pass.

Following figure shows the process of generating CRC reminder:

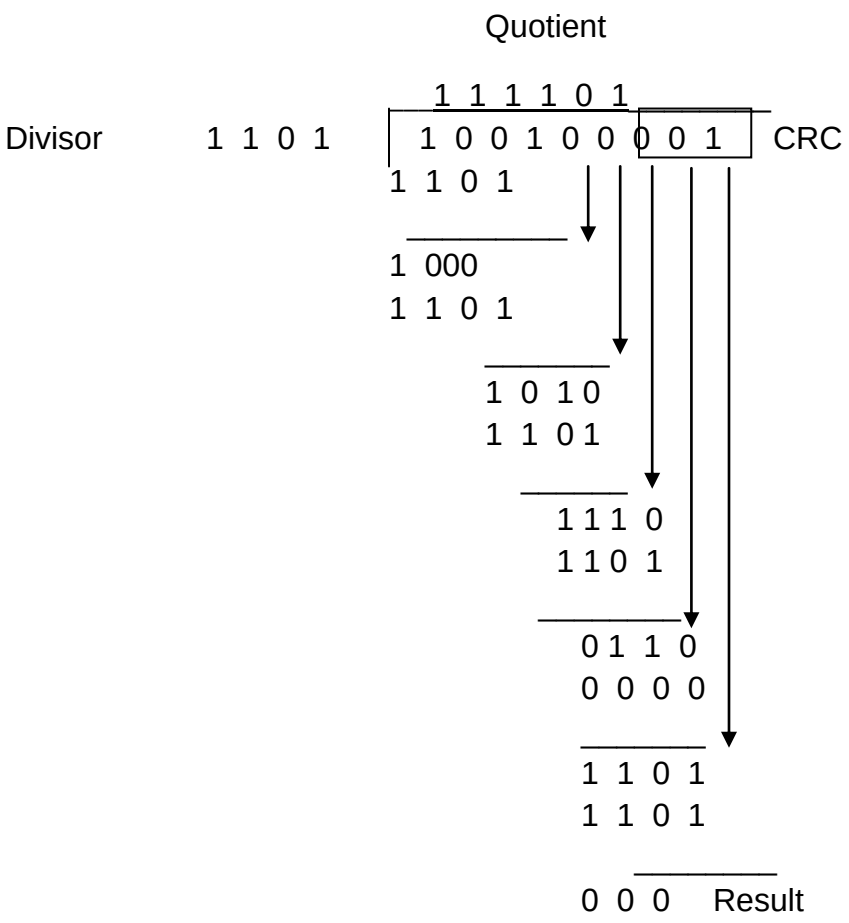
Figure :



A CRC checker functions does exactly as the generator does. After receiving the data appended with the CRC, it does the samemodulo-2 division. If the remainder is all 0's, the CRC isdropped and the data is accepted: otherwise, the receivedstream of bits is discarded and data is resent.

Following Figure shows the same process of division in the receiver.

Figure:



Performance:
CRC is a very effective error detection method. If the divisoris chosen according to the previously mentioned rules,

1.CRC can detect all burst errors that affect an odd numberof bits.

2.CRC can detect all burst errors of length less than or equalto the degree of the polynomial

3.CRC can detect, with a very high probability, burst errorsof length greater than the degree of the polynomial.

3. Checksum

A checksum is fixed length data that is the result of performing certain operations on the data to be sent from sender to the receiver. The sender runs the appropriate checksum algorithm to compute the checksum of the data, appends it as a field in the

packet that contains the data to be sent, as well as various headers.

When the receiver receives the data, the receiver runs the same checksum algorithm to compute a fresh checksum. The receiver compares this freshly computed checksum with the checksum that was computed by the sender. If the two checksum matches, the receiver of the data is assured that the data has not changed during the transit.

Hamming Code:

The Hamming code can be applied to data units of any length and uses the relationship between data and redundancy bits discussed above. For example, a 7-bit ASCII code requires 4 redundancy bits that can be added 10 the end of the data unit or interspersed with the original data bits. In following Figure,these bits are placed in positions 1, 2,4, and 8 (the positions in an 11-bit sequence that are powers of 2). For clarity in the examples below, we refer to these bits as r1, r2, r4, and r8.

11	10	9	8	7	6	5	4	3	2	1
d	d	d	*r8*	d	d	d	*r4*	d	*r2*	*r1*

In the Hamming code, each r bit is the parity bit for onecombination of data bits, is shown below:

r1 : bits 1,3,5,7,9,11
r2 : bits 2,3,6,7,10,11
r3 : bits 4,5,6,7
r4 : bits 8,9,10,11

r1will be assigned to these bits position:

11 9 7 5 3

d	d	d	*r8*	d	d	d	*r4*	d	*r2*	*r1*

1

r2will be assigned to these bits position:

11 10 7 6 3 2

d	d	d	*r8*	d	d	d	*r4*	d	*r2*	*r1*

r4will be assigned to these bits position:

7 6 5 4

d	d	d	*r8*	d	d	d	*r4*	d	*r2*	*r1*

r8will be assigned to these bits position:

11 10 9 8

d	d	d	*r8*	d	d	d	*r4*	d	*r2*	*r1*

Now suppose we have 1001101 data to be sent, then the redundant bits are calculated by the following method:

Data is: 1 0 0 1 1 0 1

11 10 9 8 7 6 5 4 3 2 1

1	0	0		1	1	0		1		

Adding r1:

11 10 9 8 7 6 5 4 3 2 1

1	0	0		1	1	0		1		1

Adding r2:

11	10	9		8	7	6		5	4	3	2	1
1	0	0		1	1	0		1	0	1		

Adding r4:

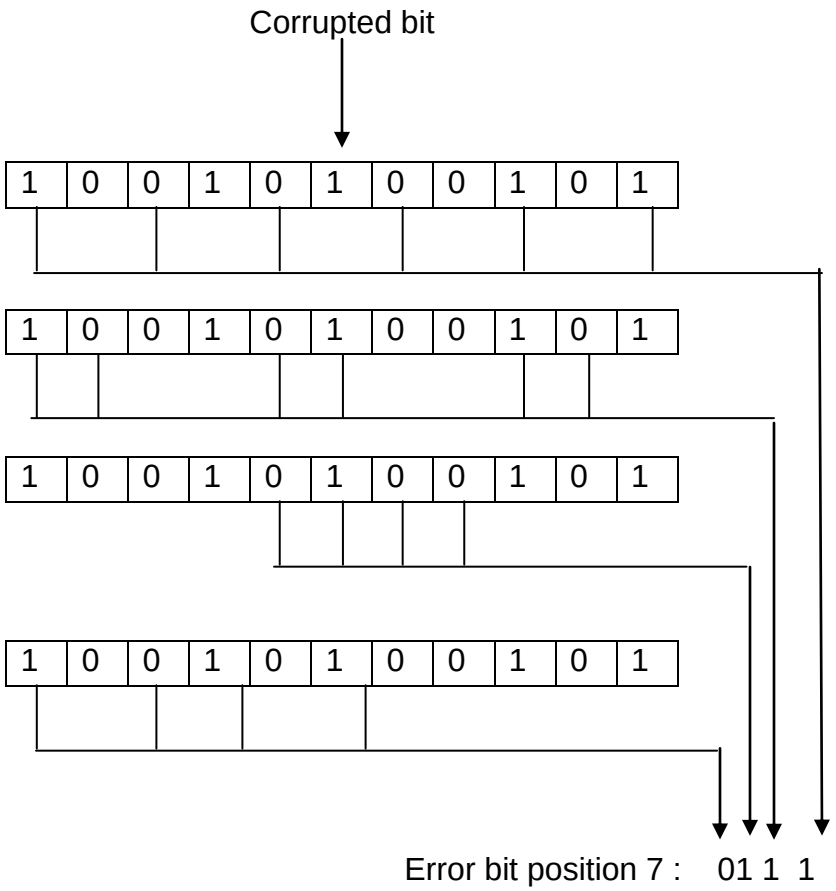
11	10	9		8	7	6	5	4	3	2	1
1	0	0		1	1	0	0	1	0	1	

Adding r8:

11	10	9	8	7	6	5	4	3	2	1
1	0	0	1	1	1	0	0	1	0	1

Code: 10011100101

Now imagine that by the time the above transmission isreceived, the number **7** bit has been changed from **1** to **0.**The receiver takes the transmission and recalculates 4 newparity bits, using the same sets of bits used by the senderplus the relevant parity r bit for each set (see following Fig.). Thenit assembles the new parity values into a binary number inorder of r position (r8 r4, r2, r1). In our example, this stepgives us the binary number 0111 (7 in decimal), which is theprecise location of the bit in error.



Once the bit is identified, the receiver can reverse itsvalue and correct the error. The beauty of the technique isthat it can easily be implemented in hardware and the codeis corrected before the receiver knows about it.

7.8 REVIEW QUESTIONS

1. Explain error classification?
2. Explain error type with example?
3. Discuss Hamming code?
4. explain CRC.
5. write a short note on Checksum?

7.9 REFERENCES

1. Data Communication & Networking – BehrouzForouzan

❖❖❖❖

8

SIGNAL ENCODING

Unit Structure

8.0 OBJECTIVES

The objectives of this chapter are:

1. Understand what is signal encoding
2. Different ways of converting analog signal to digital
3. Different ways of converting digital signal to analog
4. Modulation

8.1 INTRODUCTION TO SIGNAL ENCODING

- Data can be analog or digital, so can be the signal that represents it.
- **Signal encoding** is the conversion from analog/digital data to analog / digital signal.

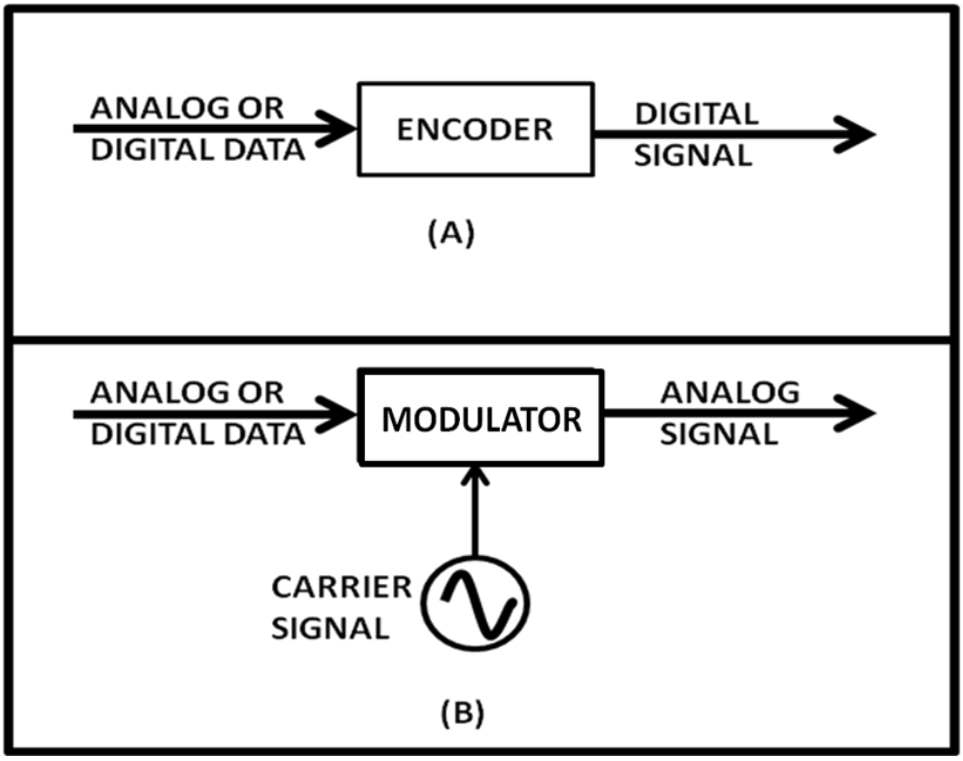


Figure: Signal Encoding

In the Figure above,

A) Demonstrates Digital Signaling where data from an analog/digital source is encoded into Digital Signal

B) Demonstrates Analog signaling in which the analog/digital source modulates a continuous carrier signal to produce an analog signal.

The possible encodings are:

1. Digital data to Digital Signal
2. Digital data to Analog Signal
3. Analog data to Digital Signal
4. Analog data to Analog Signal

8.2 SYNCHRONIZATION

- In order to receive the signals correctly, the receivers bit intervals must correspond exactly to the senders bit intervals.
- The clock frequency of the transmitter and receiver should be the same.
- If the clock frequency at the receiver is slower or faster than the bit intervals are not matched and the received signal is different than the transmitted one.

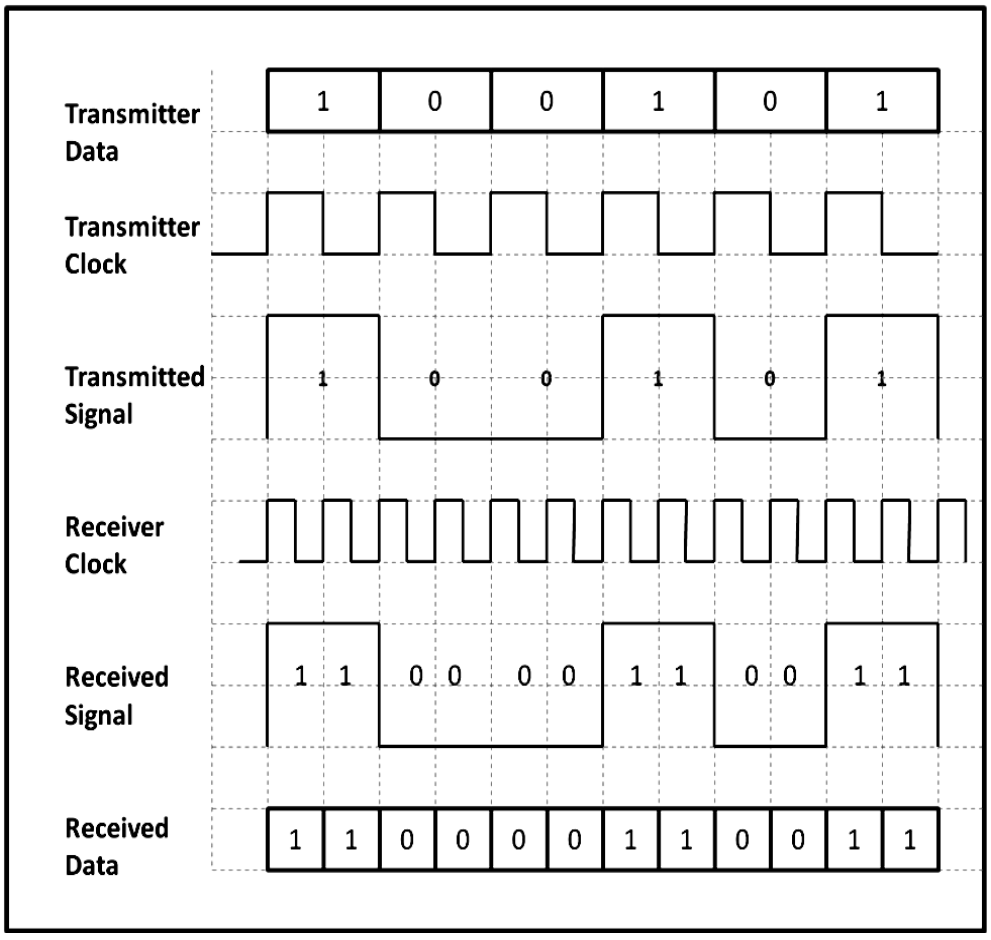


Figure : Synchronization

- In the above figure, the receiver clock frequency is twice that of the transmitter frequency. Hence the received data is totally different than the transmitted one

- To avoid this, receiver and transmitter clocks have to be **synchronized.**

- To achieve this the transmitted digital signal should include timing information which forces synchronization

8.3 Digital Data to Digital Signal

Coding methods Coding methods are used to convert digital data into digital signals.

There are two types of coding methods:

1 Line Coding
2 Block Coding

Scrambling is also one of the ways to convert digital data to digital signals but is not used.

8.3.1 Line Encoding

It is the process of converting Digital data into digital signal.

In other words, it is converting of binary data(i.e. A sequence of bits) into digital signal (i.e. a sequence of discrete, discontinuous voltage pulses)

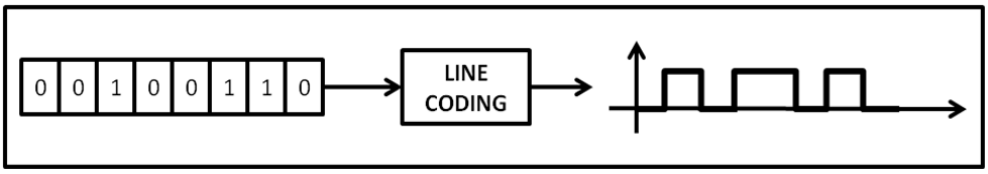


Figure: Line Coding

8.3.2 Classification of Line Codes

The following figure shows the classification of Line coding schemes:

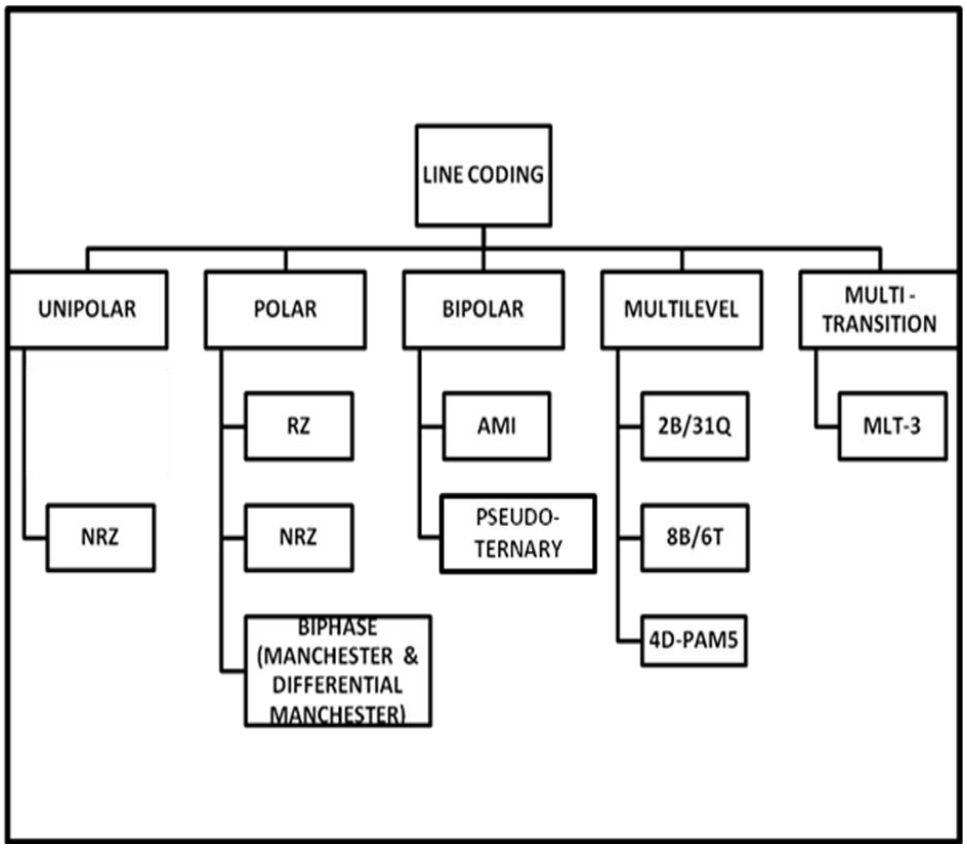


Figure : Classification of line coding schemes

8.3.2.A Unipolar

- All signal levels are either above or below the time axis.
- NRZ - Non Return to Zero scheme is an example of this code. The signal level does not return to zero during a symbol transmission.

8.3.2.B Polar

- **NRZ-voltages** are on both sides of the time axis.
- Polar NRZ scheme can be implemented with two voltages. E.g. +V for 1 and -V for 0.

- There are two variations:
    - **NZR - Level (NRZ-L)** - positive voltage for one symbol and negative for the other

    - **NRZ - Inversion (NRZ**-I) - the change or lack of change in polarity determines the value of a symbol. E.g. a “1” symbol inverts the polarity a “0” does not.

- **Polar – RZ**
    - The Return to Zero (RZ) scheme uses three voltage values. +, 0, -.

- Each symbol has a transition in the middle. Either from high to zero or from low to zero

- More complex as it uses three voltage level. It has no error detection capability

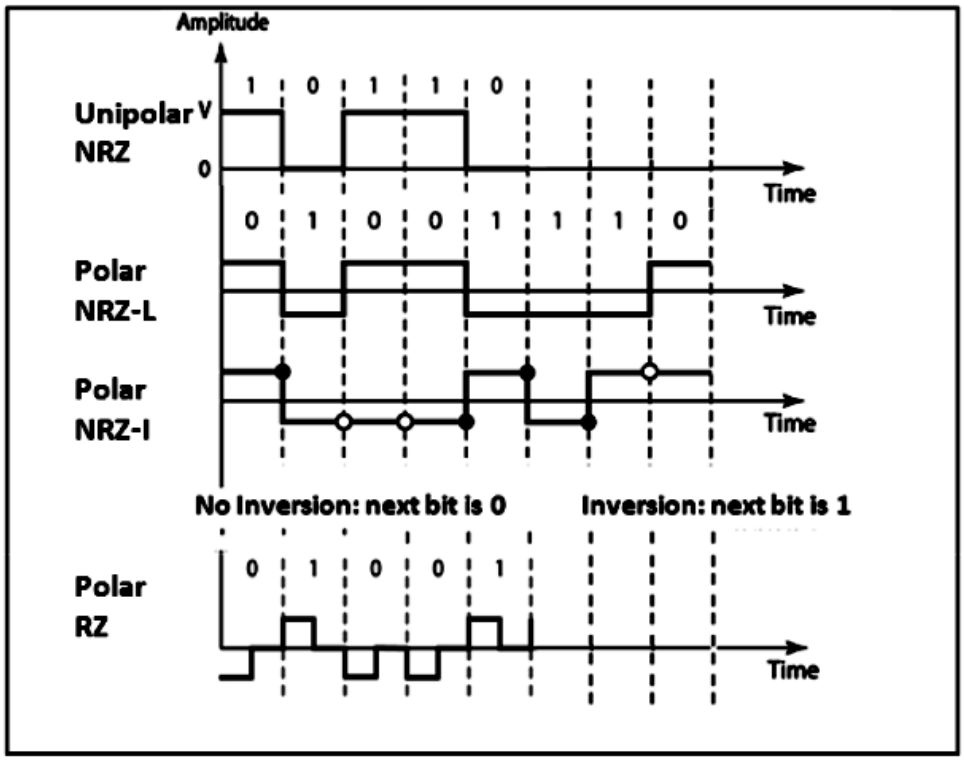


Figure : Unipolar(NRZ) & Polar(RZ & NRZ) Encoding

- **Polar - Biphase: Manchester and Differential Manchester**
  - **Manchester coding** is a combination of NRZ-L and RZ schemes.
    - Every symbol has a level transition in the middle: from high to low or low to high.
    - It uses only two voltage levels.

  - **Differential Manchester coding** consists of combining the NRZ-I and RZ schemes.
    - Every symbol has a level transition in the middle. But the level at the beginning of the symbol is determined by the symbol value. One symbol causes a level change the other does not.

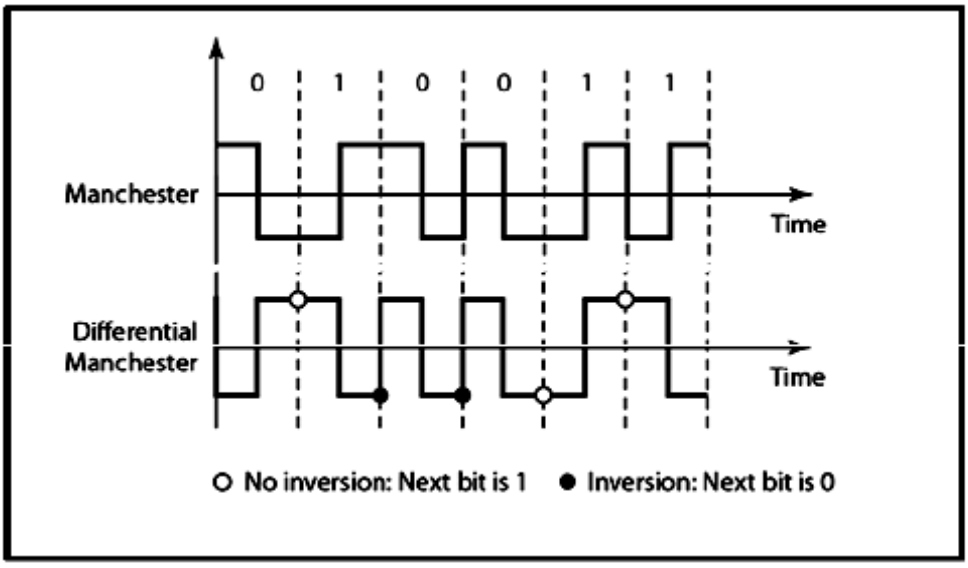


Figure : Polar biphase: Manchester and differential Manchester coding schemes

8.3.2.C Bipolar - AMI and Pseudoternary

- This coding scheme uses 3 voltage levels: - +, 0, -, to represent the symbols
- Voltage level for one symbol is at “0” and the other alternates between + & -.
- **Bipolar Alternate Mark Inversion (AMI)** - the “0” symbol is represented by zero voltage and the “1” symbol alternates between +V and -V.
- **Pseudoternary** is the reverse of AMI

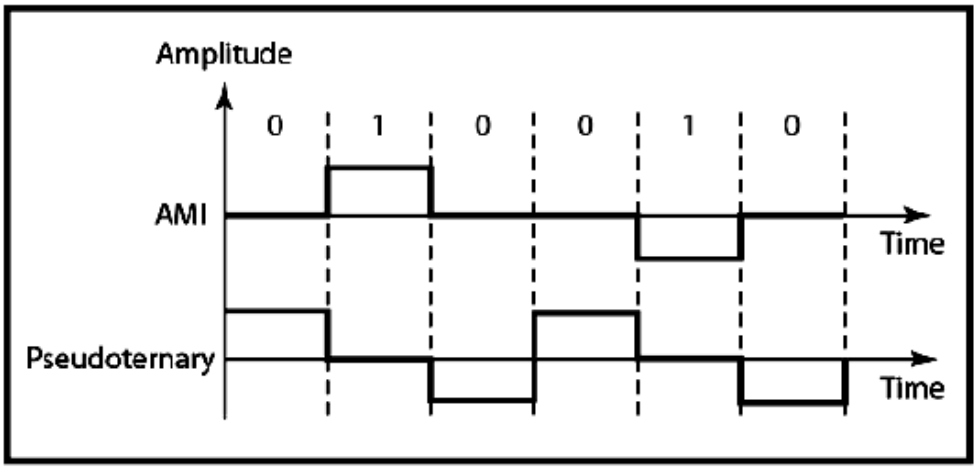


Figure: Bipolar coding scheme - AMI and Pseudoternary

8.3.2.D Multilevel

- Here the number of data bits is increased per symbol to increase the bit rate.
- 2 types of data element a 1 or a 0 are available, it can be combined into a pattern of ***n*** elements to create $\boldsymbol{2^m}$ symbols.
- Using L signal levels we can have ***n*** signal elements to create $\boldsymbol{L^n}$ signal elements. The following possibilities can occur:
- With 2^m symbols and L^n signals:
  - If $2^m > L^n$ then we cannot represent the data elements, we don't have enough signals.
  - If $2^m = L^n$ then we have an exact mapping of one symbol on one signal.
  - If $2^m < L^n$ then we have more signals than symbols and we can choose the signals that are more distinct to represent the symbols and therefore have better noise immunity and error detection as some signals are not valid
- These types of codings are classified as **mBnL** schemes. In ***mBnL*** schemes, a pattern of *m* data elements is encoded as a pattern of *n* signal elements in which $\boldsymbol{2^m \leq L^n}$.
- **2B1Q (**two binary, one quaternary**)**
- Here m = 2; n = 1 ; Q = 4. It uses data patterns of size 2 and encodes the 2-bit patterns as one signal element belonging to a four-level signal.

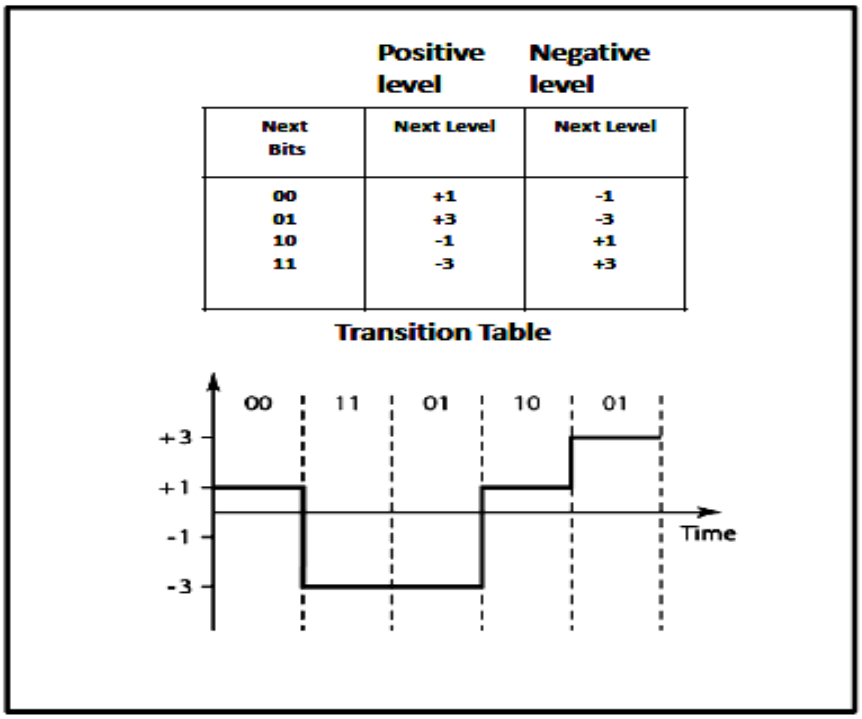


Figure: Multilevel coding scheme : 2B1Q

- **8B6T(**eight binary, six ternary**)**
  - Here a pattern of 8 bits is encoded a pattern of 6 signal elements, where the signal has three levels
  - Here m = 8; n = 6 ; T = 3
  - So we can have 2^8 =256 different data patterns and 3^6 =478 different signal patterns.

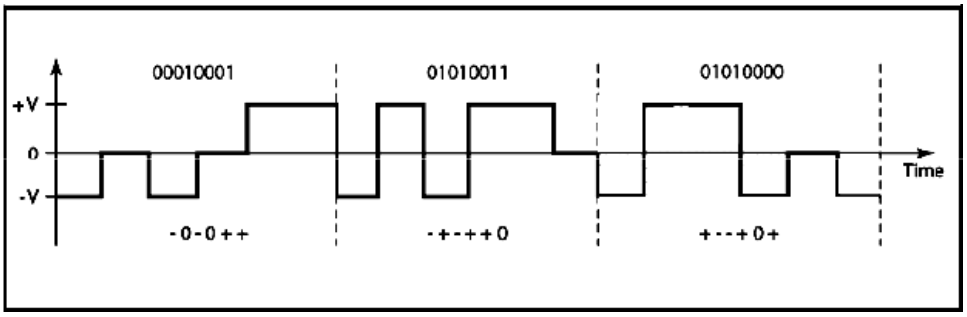


Figure : Multilevel coding scheme : 8B6T

- **4D-PAM5** (Four Dimensional Five-Level Pulse Amplitude Modulation)
  - **4D -**means that data is sent over four channels at the same time.
  - It uses five voltage levels, such as -2, -1, 0, 1, and 2.

8.3.2.E Multitransition

- Because of synchronization requirements we force transitions. This can result in very high bandwidth requirements -> more transitions than are bits (e.g. mid bit transition with inversion).

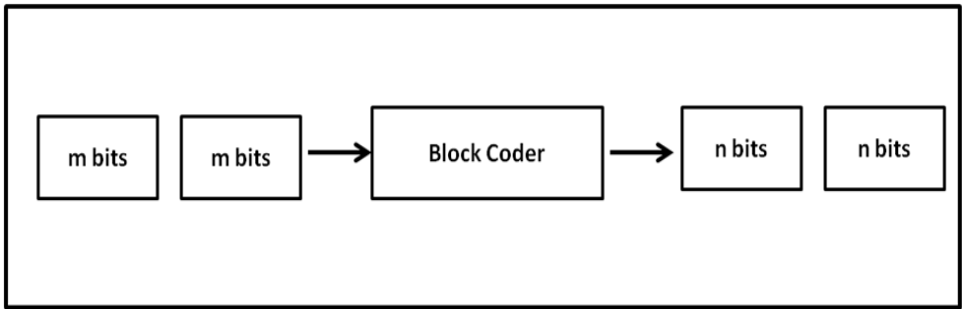
- Codes can be created that are differential at the bit level forcing transitions at bit boundaries. This results in a bandwidth requirement that is equivalent to the bit rate.

- In some instances, the bandwidth requirement may even be lower, due to repetitive patterns resulting in a periodic signal.

- **MLT-3**
  - Signal rate is same as NRZ-I
  - Uses three levels (+*v,* 0, and - *V)* and three transition rules to move between the levels.
    - If the next bit is 0, there is no transition.
    - If the next bit is 1 and the current level is not 0, the next level is 0.
    - If the next bit is 1 and the current level is 0, the next level is the opposite of the last nonzero level.

8.3.3 Block Coding

- Block coding adds redundancy to line coding so that error detection can be implemented.
- Block coding changes a block of *m* bits into a block of *n* bits, where *n* is larger than *m.*
- **Block coding is referred to as an *mB/nB* encoding technique.**
- The additional bits added to the original “m bits” are called parity **bits** or **check bits**



m : message bits

Figure : Block Coding

Example: 4B/5B encoding

Here a 4 bit code is converted into a 5 bit code

8.4 Analog data to analog signal

8.4.1 Modulation

- The Process of converting analog data to analog signal is called Modulation.
- Modulation is used to send an information bearing signal over long distances.
- Modulation is the process of varying some characteristic of a periodic wave with an external signal called carrier signal.
- These carrier signals are high frequency signals and can be transmitted over the air easily and are capable of traveling long distances.
- The characteristics (amplitude, frequency, or phase) of the carrier signal are varied in accordance with the information bearing signal(analog data).
- The information bearing signal is also known as the modulating signal.

- The modulating signal is a slowly varying – as opposed to the rapidly varying carrier frequency.

8.4.2 Types of Modulation:

Signal modulation can be divided into two broad categories:

- Analog modulation and
- Digital modulation.

- **Analog or digital** refers to how the data is modulated onto a sine wave.
- If analog audio data is modulated onto a carrier sine wave, then this is referred to as **analog modulation**.
- **Digital modulation** is used to convert digital data to analog signal. Ex ASK, FSK, PSK.

8.4.2.1 Analog Modulation can be accomplished in three ways:

1. Amplitude modulation (AM)
2. Frequency modulation (FM)
3. Phase modulation (PM).

8.4.2.1.1 Amplitude modulation (AM)

- Amplitude modulation is a type of modulation where the amplitude of the carrier signal is varied in accordance with modulating signal.
- The envelope, or boundary, of the amplitude modulated signal embeds modulating signal.
- Amplitude Modulation is abbreviated *AM*.

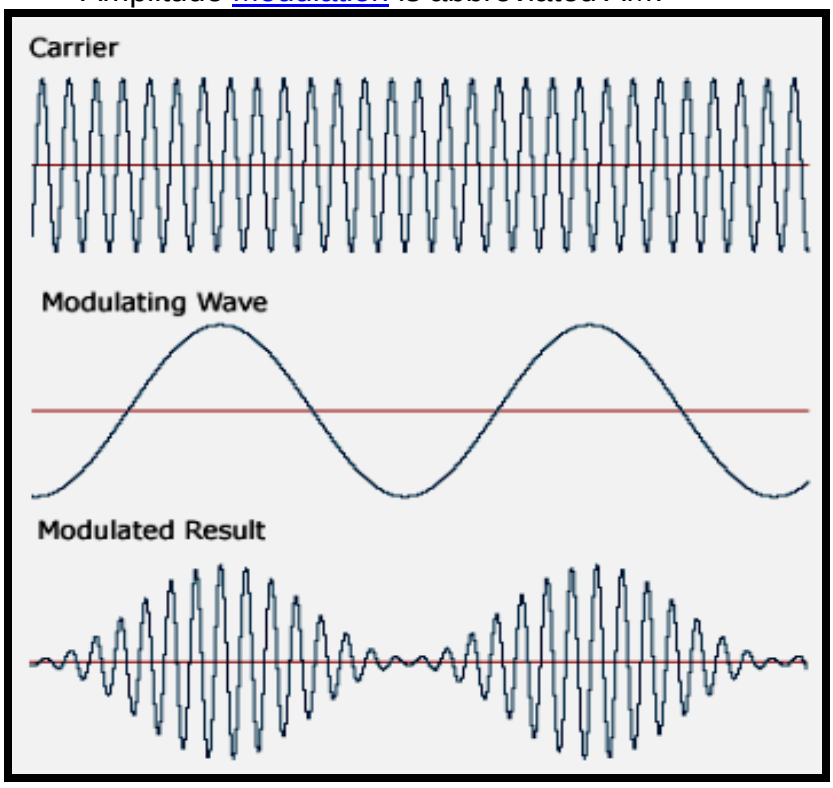


Figure : Amplitude modulation (AM)

8.4.2.1.2 Frequency modulation (FM)

- Frequency modulation is a type of modulation where the frequency of the carrier is varied in accordance with the modulating signal. The amplitude of the carrier remains constant.

- The information-bearing signal (the modulating signal) changes the instantaneous frequency of the carrier. Since the amplitude is kept constant, FM modulation is a low-noise process and provides a high quality modulation technique which is used for music and speech in hi-fidelity broadcasts.

- Frequency Modulation is abbreviated *FM*.

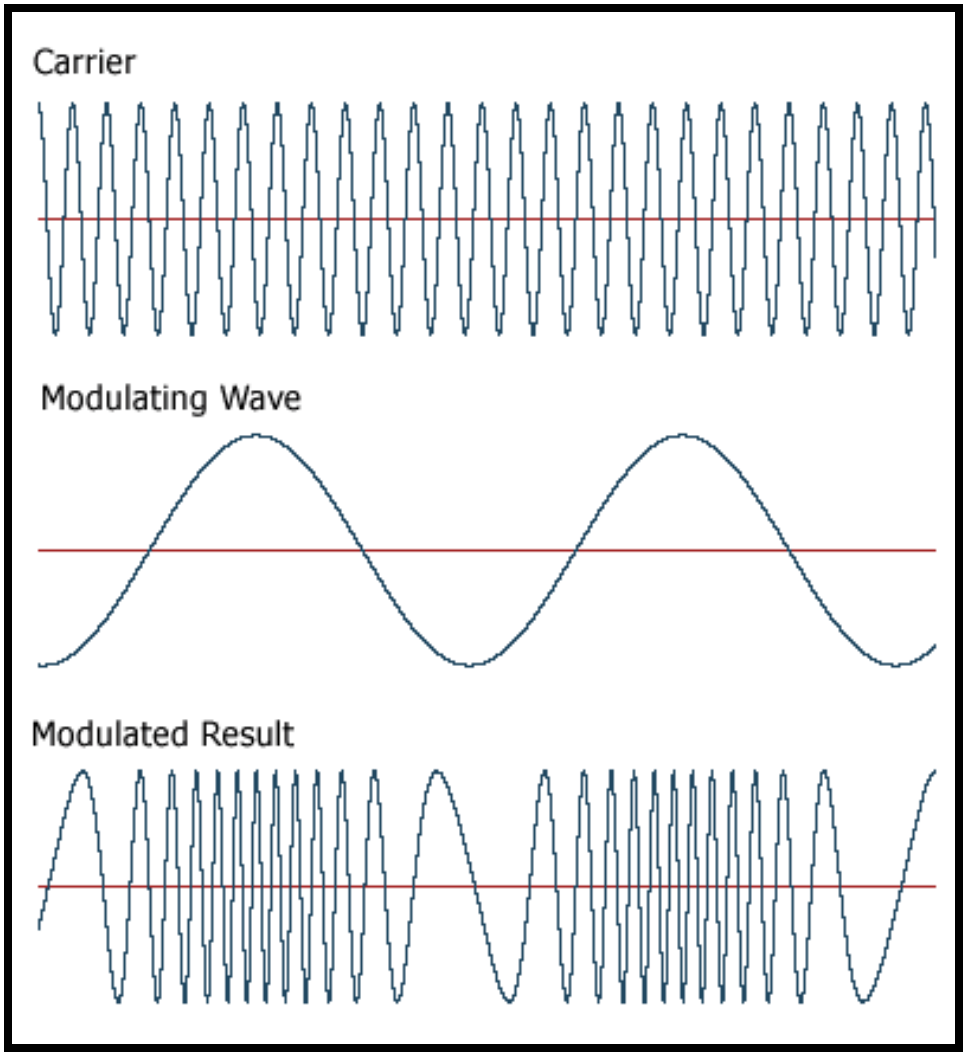


Figure : Frequency modulation (FM)

8.4.2.1.3 Phase modulation (PM).

- In phase modulation, the instantaneous phase of a carrier wave is varied from its reference value by an

amount proportional to the instantaneous amplitude of the modulating signal.

- Phase Modulation is abbreviated *PM*.

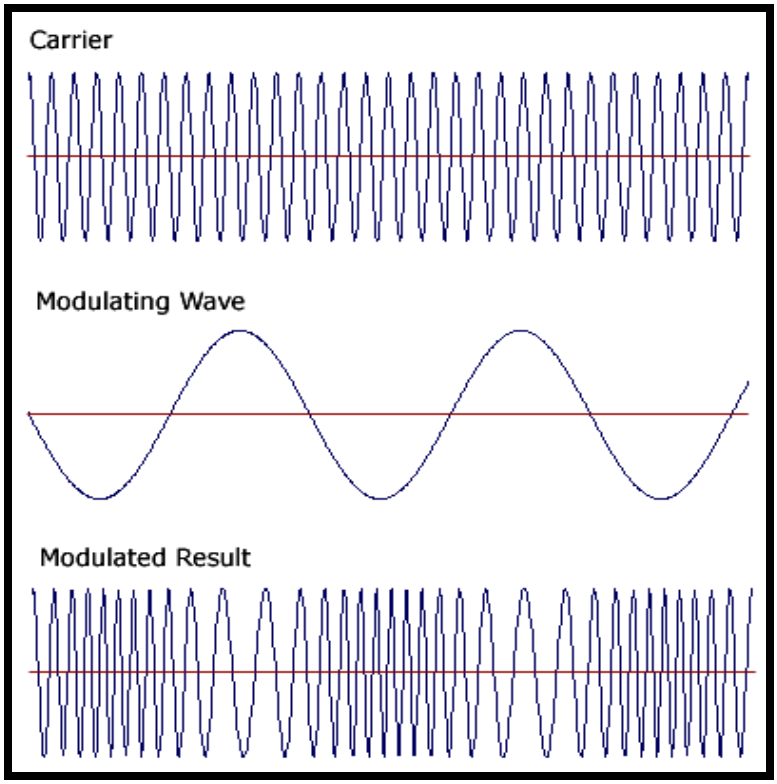


Figure : Phase modulation (PM).

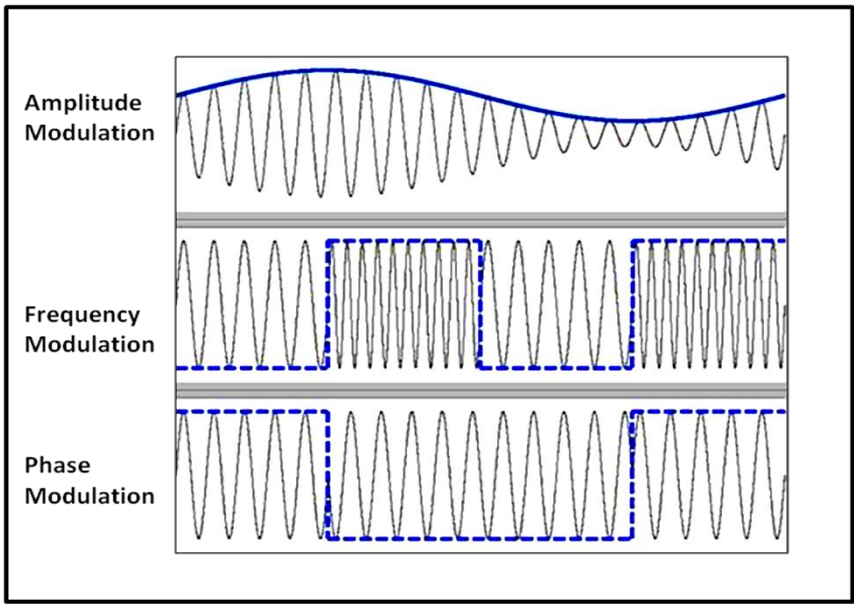


Figure : Comparison of AM, FM & PM

8.4.2.2 Digital Modulation Types(Digital to Analog signal conversion)

- **Digital modulation** is used to convert digital data to analog signal.
- **It can be accomplished in the following ways:**
  1. ASK
  2. FSK
  3. PSK
  4. QAM

8.4.2.2.1 Amplitude Shift Keying (ASK)

- In amplitude shift keying, the amplitude of the carrier signal is varied to create signal elements.
- Both frequency and phase remain constant while the amplitude changes.

- **Binary ASK (BASK)**
  ASK is normally implemented using only two levels and is hence called binary amplitude shift keying.
  Bit 1 is transmitted by a carrier of one particular amplitude.
  To transmit Bit 0 we change the amplitude keeping the frequency is kept constant

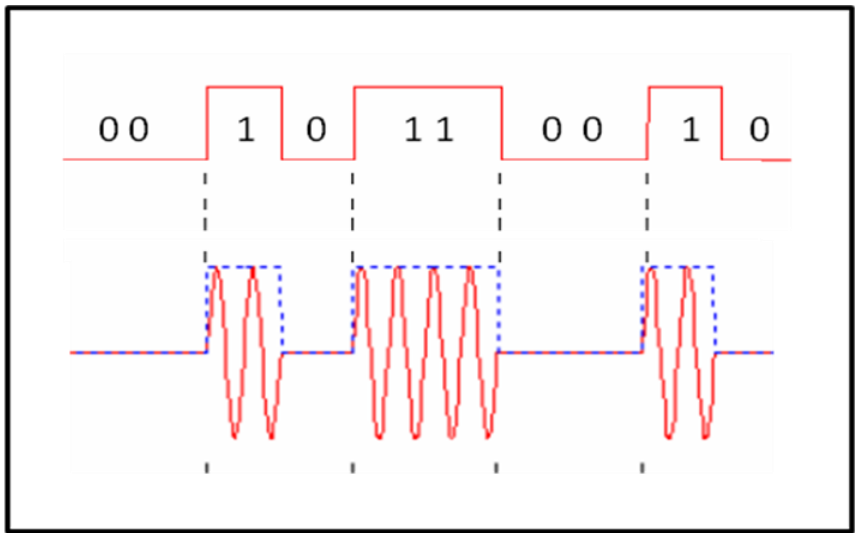


Figure : Amplitude Shift Keying (ASK

8.4.2.2.2 Frequency Shift Keying (FSK)

- In Frequency shift keying, we change the frequency of the carrier wave.
- Bit 0 is represented by a specific frequency, and bit 1 is represented by a different frequency.
- In the figure below frequency used for bit 1 is higher than frequency used for bit 0

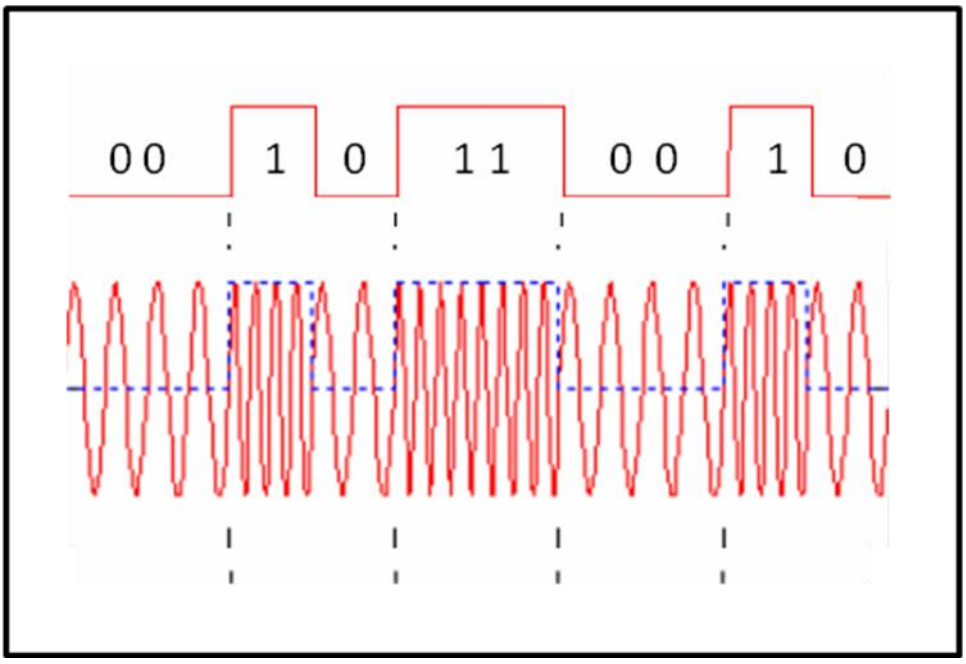


Figure : Frequency Shift Keying (FSK)

8.4.2.2.3. Phase Shift Keying (PSK)

- Phase shift keying (PSK) is a method of transmitting and receiving digital signals in which the phase of a transmitted signal is varied to convey information.
- Both amplitude and frequency remain constant as the phase changes.
- The simplest from of PSK has only two phases, 0 and 1.
- If the phase of the wave does not change, then the signal state stays the same (low or high).
- If the phase of the wave changes by 180 degrees, that is, if the phase reverses, then the signal state changes (from low to high or from high to low)

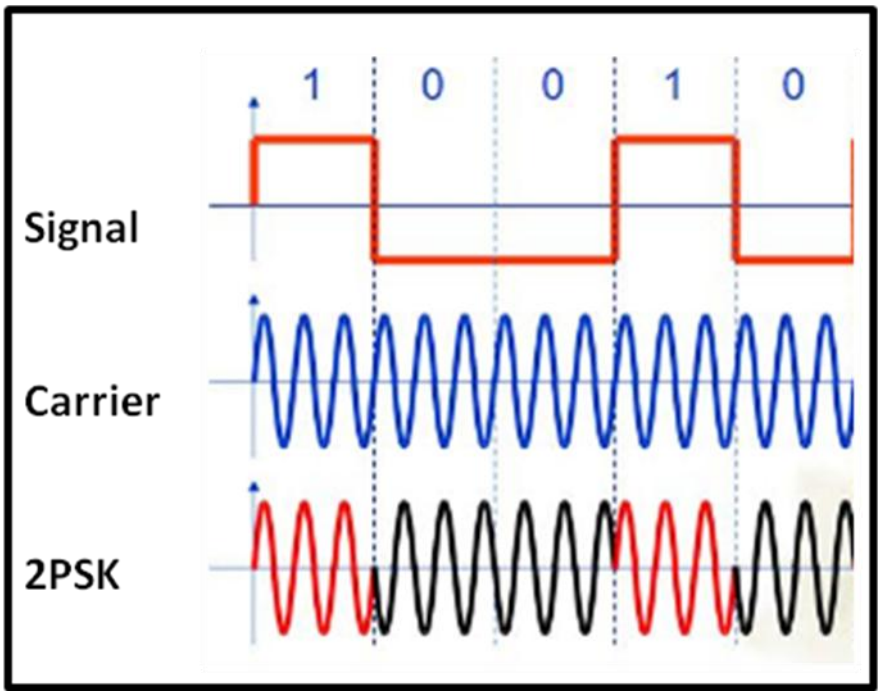


Figure: Phase Shift Keying (PSK)

8.4.2.2.4 QAM

- The concept of Quadrature Amplitude Modulation (QAM) involves use of two carriers, one for phase and the other for quadrature, with different amplitude levels for each carrier.
- It is a combination of ASK & PSK.

8.4.2.2 Analog to Digital Conversion using modulation

The definition of the term modulation is described in the next section. Here we discuss 3 modulation techniques:

1. PAM
2. PCM
3. PWM

8.4.2.3.1 PAM (Pulse Amplitude Modulation)

Pulse Amplitude Modulation refers to a method of carrying information on a train of pulses, the information being encoded in the amplitude of the pulses.

8.4.2.3.2 PCM (Pulse Code Modulation)

- PCM is a general scheme for transmitting analog data in a digital and binary way, independent of the complexity of the analog waveform. With PCM all forms of analog data like video, voice, music and telemetry can be transferred.

- To obtain PCM from an analog waveform at the source (transmitter), the analog signal amplitude is

sampled at regular time intervals. The sampling rate (number of samples per second), is several times the maximum frequency of the analog waveform. The amplitude of the analog signal at each sample is rounded off to the nearest binary level (quantization).

- The number of levels is always a power of 2 (4, 8, 16, 32, 64, ...). These numbers can be represented by two, three, four, five, six or more binary digits (bits) respectively.

- At the destination (receiver), a pulse code demodulator converts the binary numbers back into pulses having the same quantum levels as those in the modulator. These pulses are further processed to restore the original analog waveform.

8.4.2.3.3 PWM (Pulse Width Modulation)

- Pulse Width Modulation refers to a method of carrying information on a train of pulses, the information being encoded in the width of the pulses. In applications to motion control, it is not exactly information we are encoding, but a method of controlling power in motors without (significant) loss.

- There are several schemes to accomplish this technique. One is to switch voltage on and off, and let the current recirculate through diodes when the transistors have switched off. Another technique is to switch voltage polarity back and forth with a full-bridge switch arrangement, with 4 transistors.

- This technique may have better linearity, since it can go right down to an cycles, and may jitter between minimum duty cycles of positive and negative polarity.

- In battery systems PWM is the most effective way to achieve a constant voltage for battery charging by switching the system controller's power devices on and off.

The generation of exact working PWM circuitry is complicated, but it is extremely conceptually important since there is good reason to believe that neurons transmit information using PWM spike trains.

8.5 REVIEW QUESTIONS

1. Explain the different ways of converting data to signal and viceversa.
2. Eplain in detail what is signal encoding
3. What are the different ways of converting analog data to digital data?
4. What is modulation? What are its two types?

8.6 REFERENCES & FURTHER READING

1. Data Communication & Networking – Behrouz Forouzan.
2. Computer Networks – Andrew Tannenbaum

9

TRANSMISSION MODES & IMPAIRMENTS

Unit Structure

9.0 OBJECTIVES

After reading this chapter you will understand:

- Transmission Modes and their types
- Parallel transmission mode
- Serial Transmission mode.
- Asynchronous and synchronous transmission mode
- Transmission Impairment

9.1 INTRODUCTION

This chapter gives an understanding about the way in which data can be transmitted between two devices (parallel & Serial) and the impairments in transmission due to imperfections of the transmission medium.

9.2 TRANSMISSION MODES

- Data is transmitted between two digital devices on the network in the form of bits.

- Transmission mode refers to the mode used for transmitting the data. The transmission medium may be capable of sending only a single bit in unit time or multiple bits in unit time.

- When a single bit is transmitted in unit time the transmission mode used is Serial Transmission and when multiple bits are sent in unit time the transmission mode used is called Parallel transmission.

Types of Transmission Modes:

- There are two basic types of transmission modes Serial and Parallel as shown in the figure below.

- Serial transmission is further categorized into Synchronous and Asynchronous Serial transmission.

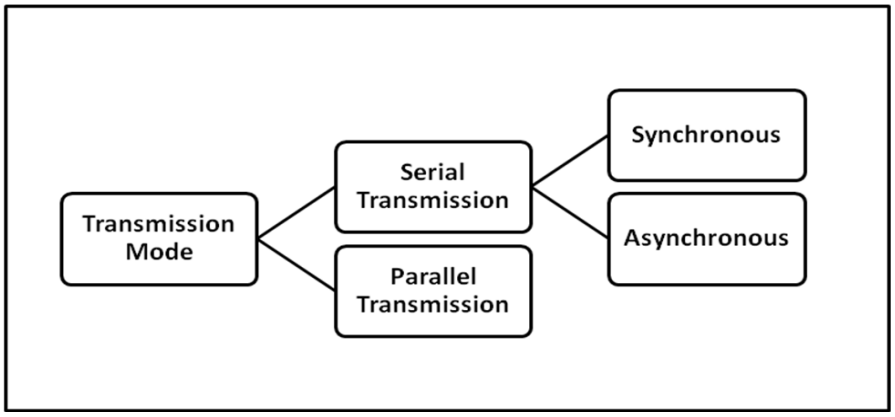


Fig. Types of Transmission Modes

9.2.1 Parallel Transmission

- It involves simultaneous transmission of *N* bits over *N* different channels
- Parallel Transmission increases transmission speed by a factor of *N* over serial transmission
- Disadvantage of parallel transmission is the cost involved, N channels have to be used, hence, it can be used for short distance communication only

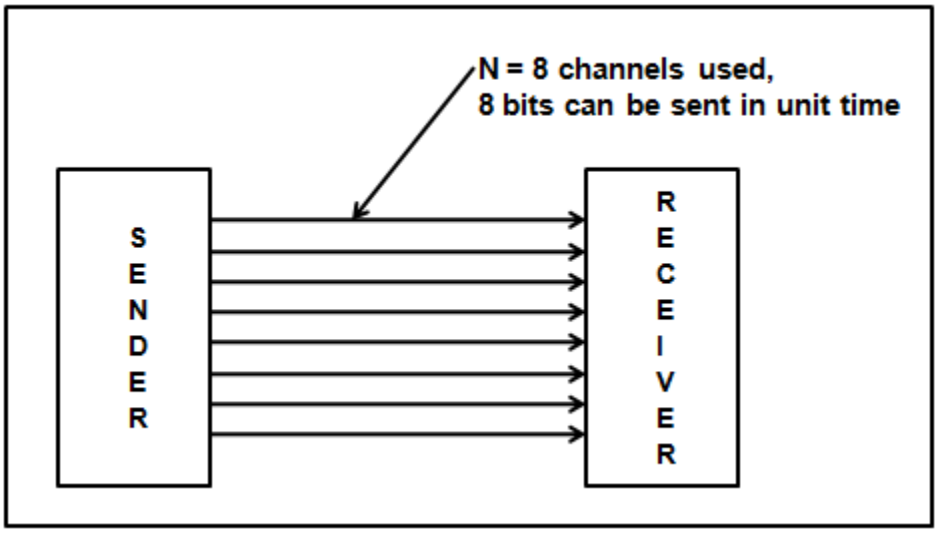


Fig. Parallel Transmission of Data over N = 8 channels

- Example of Parallel Transmission is the communication between CPU and the Projector.

9.2.2 Serial Transmission

- In Serial Transmission, as the name suggests data is transmitted serially, i.e. bit by bit, one bit at a time.
- Since only one bit has to be sent in unit time only a single channel is required.

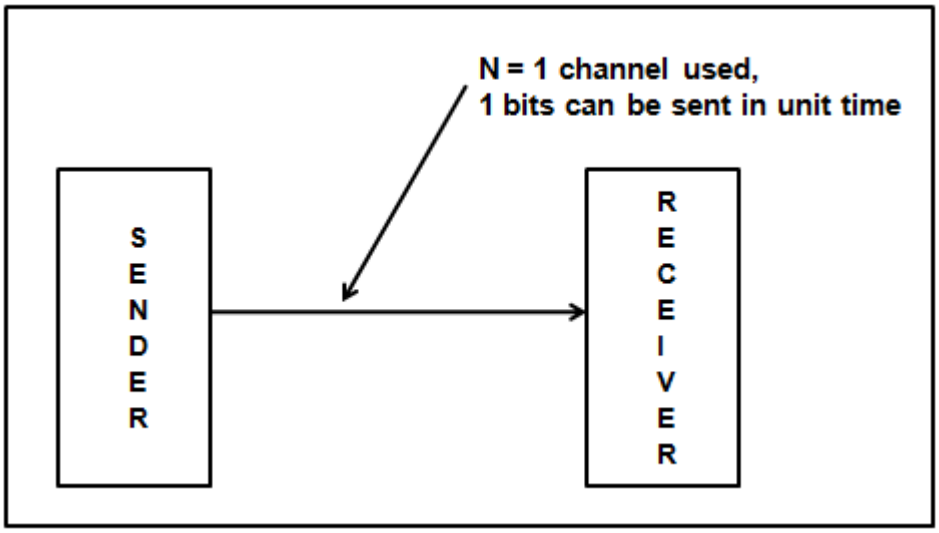


Fig. Serial Transmission of Data over N = 8 channels

Types of Serial Transmission:

Depending upon the timing of transmission of data there are two types of serial transmission as described below

9.2.2.1 ASynchronous Transmission

- In asynchronous serial transmission the sender and receiver are not synchronized.
- The data is sent in group of 8 bits i.e. in bytes.
- The sender can start data transmission at any time instant without informing the receiver.
- To avoid confusing the receiver while receiving the data, “start” and “stop” bits are inserted before and after every group of 8 bits as shown below

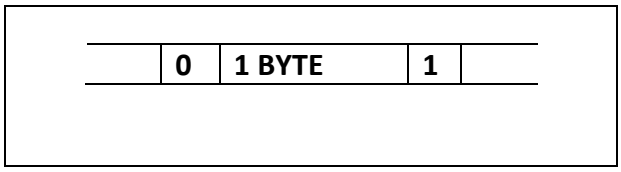


Fig: Start and Bit before and after every data byte

- The start bit is indicated by “0” and stop bit is indicated by “1”.
- The sender and receiver may not be synchronized as seen above but at the bit level they have to be synchronized i.e. the duration of one bit needs to be same for both sender and receiver for accurate data transmission.
- There may be gaps in between the data transmission indication that there is no data being transmitted from sender. Ex. Assume a user typing at uneven speeds, at times there is no data being transmitted from Keyboard to the CPU.
- Following is the Diagram for Asynchronous Serial Transmission.

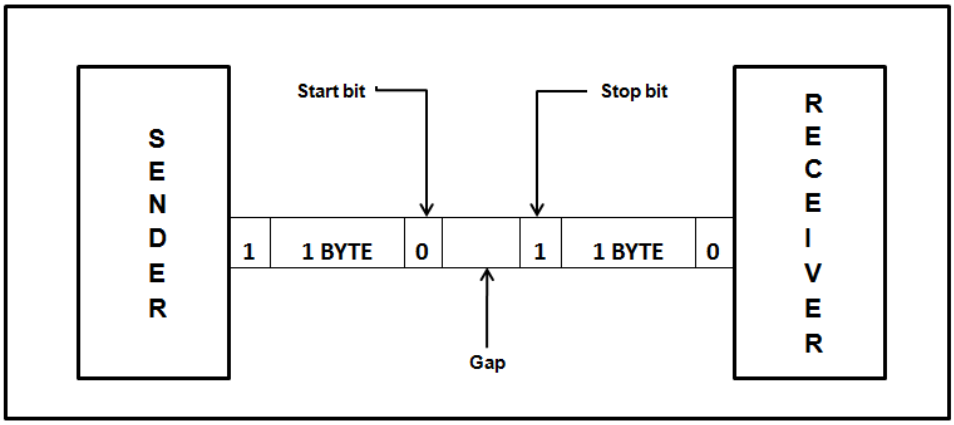


Fig: Asynchronous Serial Transmission

- **Advantages**
  1. Cheap and Effective implementation
  2. Can be used for low speed communication

- **Disadvantages**
  Insertion of start bits, stop bits and gaps make asynchronous transmission slow.

- **Application**
  Keyboard

9.2.2.2 Synchronous Transmission

- In Synchronous Serial Transmission, the sender and receiver are highly synchronized.
- No start, stop bits are used.
- Instead a common master clock is used for reference.
- The sender simply send stream of data bits in group of 8 bits to the receiver without any start or stop bit.
- It is the responsibility of the receiver to regroup the bits into units of 8 bits once they are received.
- When no data is being transmitted a sequence of 0's and 1's indicating IDLE is put on the transmission medium by the sender.

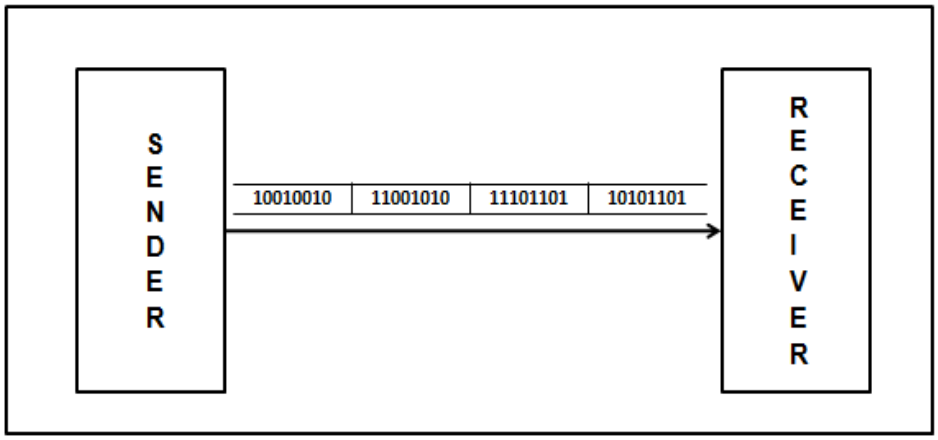


Fig: Asynchronous Serial Transmission

- **Advantage**
  1. There are no start bits, stop bits or gaps between data units
  2. Since the above are absent data transmission is faster.
  3. Due to synchronization there are no timing errors.

9.2.2.3 Comparison of serial and parallel transmission

Sr.no	Parameter	Parallel transmission	Serial transmission
1	Number of wire required to transmit N bits	N wire	1 wire
2	Number of bits transmitted simultaneously	N bits	1 bit
3	Speed of data transfer	False	Slow
4	Cost	Higher due to more number of conductor	Low, since only one wire is used
5	Application	Short distance communication such as computer to printer communication	Long distance computer to computer communication.

9.3 Transmission Impairments & Types

- Data is transmitted through transmission medium which are not perfect.
- The imperfection causes signal impairment.
- Due to the imperfection error is introduced in the transmitted data i.e. the original signal at the beginning of the transmission is not the same as the signal at the Receiver.
- There are three causes of impairment: attenuation, distortion, and noise as shown below:

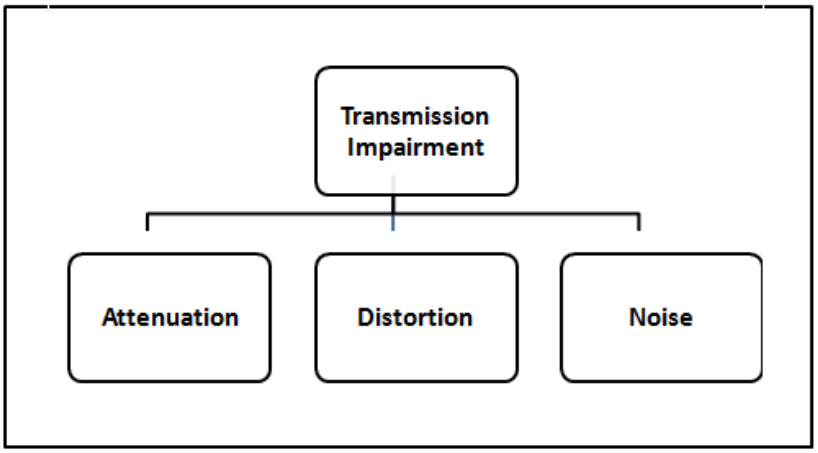


Fig: Transmission Impairment Types

9.3.1 Attenuation

- Attenuation results in loss of energy. When a signal travels through a medium, it loses some of its energy in overcoming the resistance of the medium.
- The electrical energy in the signal may converted to heat.
- To compensate for this loss, amplifiers are used to amplify the signal. Figure below shows the effect of attenuation and amplification.

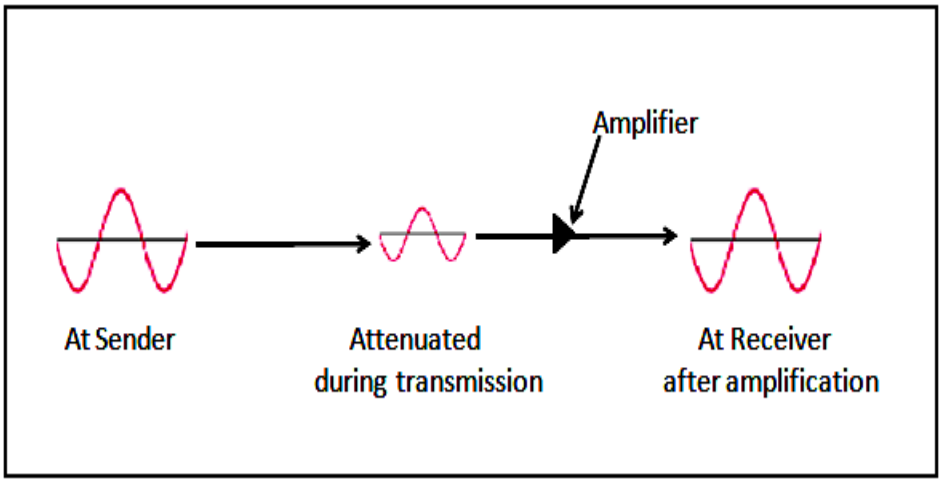


Fig. Attenuation

9.3.2 Distortion

- Distortion changes the shape of the signal as shown below

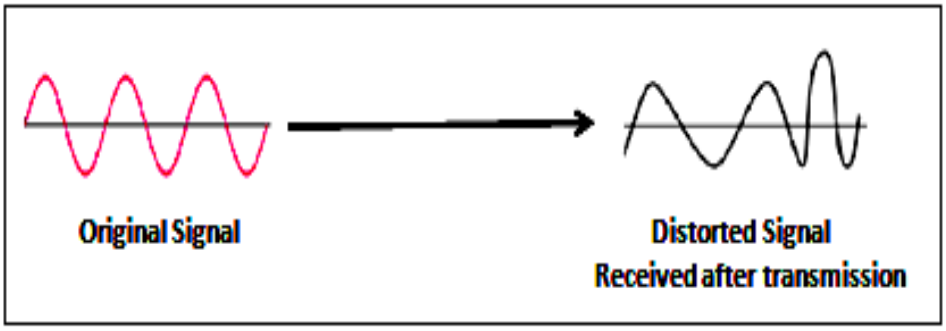


Fig. Distortion

9.3.3 Noise

- Noise is any unwanted signal that is mixed or combined with the original signal during transmission.
- Due to noise the original signal is altered and signal received is not same as the one sent.

9.4 REVIEW QUESTIONS

1. What is data transmission? What are the different possible ways of transmitting data?
2. Explain Parallel transmission mode
3. Explain Serial transmission mode and list its types
4. Explain Asynchronous serial transmission
5. Explain Synchronous Serial Transmission
6. Write a short note on transmission impairments
7. Differentiate between serial and parallel Transmission

9.5 REFERENCES & FURTHER READING

Data Communication & Networking – Behrouz Forouzan

10

TRANSMISSION MEDIUM

Unit Structure

10.0 OBJECTIVES

In this chapter, you will understand:

- Definition of Transmission Medium and its types
- Different types of Guided Transmission medium
- Different types of UnGuided Transmission medium
- Different ways in which wireless signals are transmitted

10.1 INTRODUCTION

In Data Communication networking, it is worth understanding the medium through which data passes and what are the available

mediums and their types. This chapter give a thorough understanding of the different types of transmission medium used for data communication

10.2 TRANSMISSION MEDIA

- Transmission media is ameans by which a communication signal is carried from one system to another
- A transmission medium can be defined as anything that can carry information from a source to a destination.
- The transmission medium is usually free space, metallic cable or fiber – optic cable.

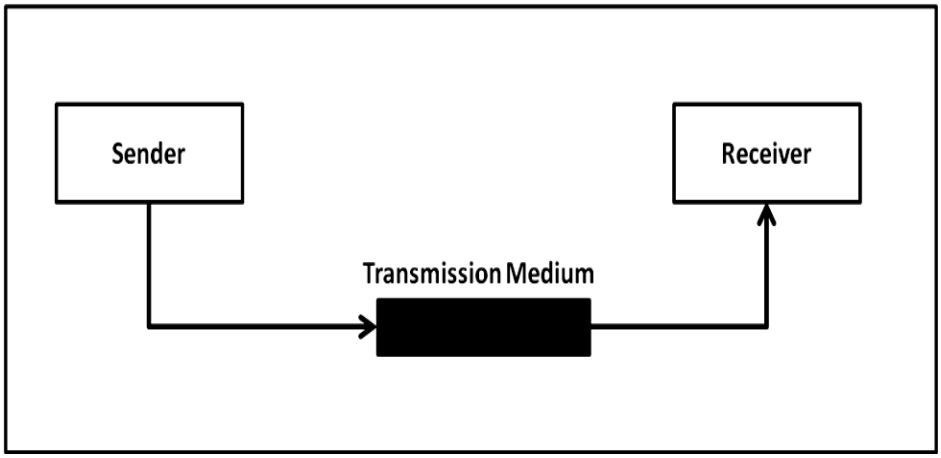


Figure: Transmission of data from sender to receiver through a medium

10.2.1 Categories of transmission media

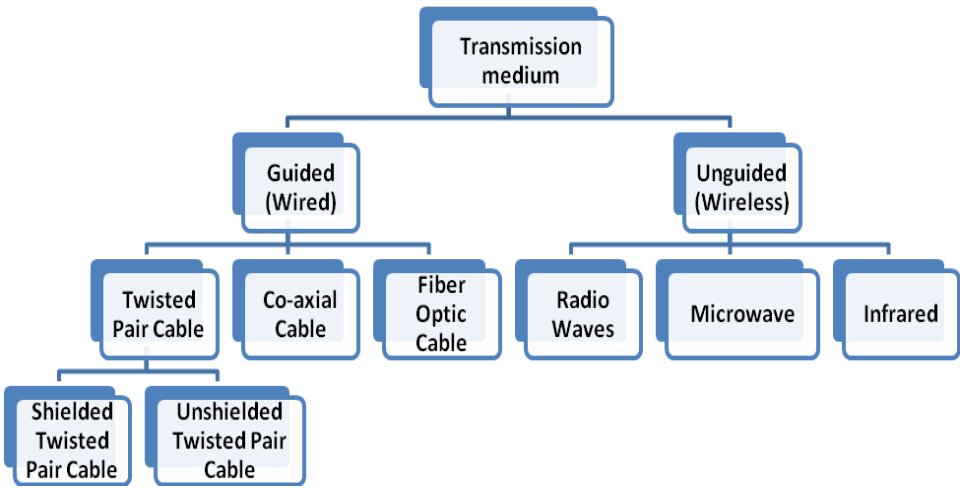


Figure : Categories of Transmission Medium

10.3 GUIDED MEDIA

- Guided Transmission media uses a cabling system that guides the data signals along a specific path.
- Guided media also known as Bounded media, which are those that provide a conduit from one device to another, include twisted-pair cable, coaxial cable, and fiber-optic cable.
- Out of these twisted-pair cable, coaxial cable transport signals in the form of electric signals and fiber-optic cable transport signals in the form of light.
- Types*:*
    1. Twisted-Pair Cable
    2. Coaxial Cable
    3. Fiber-OpticCable

10.3.1 Twisted-pair cable

Figure: Twisted Pair Cable

- The wires is twisted twisted together in pairs.
- Each pair would consist of wire used for the +ve data signal and a wire used for the —ve data signal. Any noise that appears on +ve/—ve wire of the pair would occur on the other wire.
- Because the wires are opposite polarities, they are 180 degrees out of phase (180 degree phases or definition of opposite polarity) when the noise appears on both wires, it cancels or nulls itself out at the receiving used.
- Twisted pair cables are most effectively used in a system that uses a balanced line method of transmission.

10.3.1.1 Unshielded Twisted Pair Cable (UTP)& Shielded Twisted Pair Cable (STP)

Fig. Unshielded Twisted Pair Cable

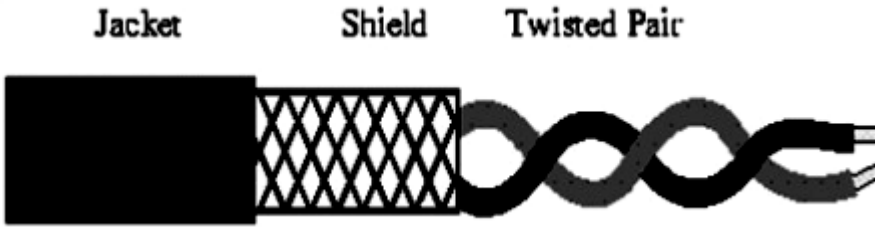


Fig. Shielded Twisted Pair Cable

- Cables with the shield are called shielded twisted pair and commonly abbreviated STP.
- Cables without a shield are called unshielded twisted pair or UTP.
- Twisting the wires together results in characteristics impedance for the cable.
- UTP or unshielded twisted pair cable is used on Ethernet
- UTP cables are used for Ethernet cabling where 4 twisted pair cables (a total of 8 wires are used)

- **10.3.2 Co-Axial Cable**

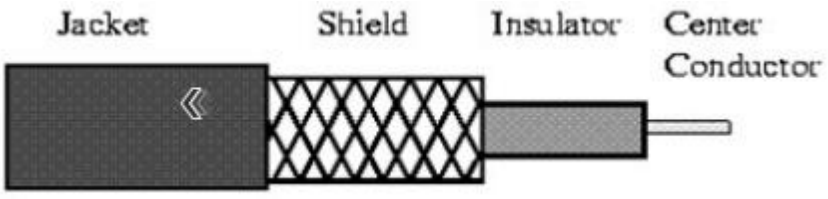


Figure: Co-axial cable

- Coaxial cable consists of 2 conductors.
- The inner conductor is contained inside the insulator with the other conductor weaves around it providing a shield.
- An insulating protective coating called a jacket covers the outer conductor.

- The outer shield protects the inner conductor from outside electrical signals.
- The distance between the outer conductor (Shield) and inner conductor plus the type of material used for insulating the inner conductor determine the cable properties or impedance. The excellent control of the impedance characteristics of the cable allow higher data rates to be transferred than twisted pair cable.

10.3.3 Fibre Optic Cable

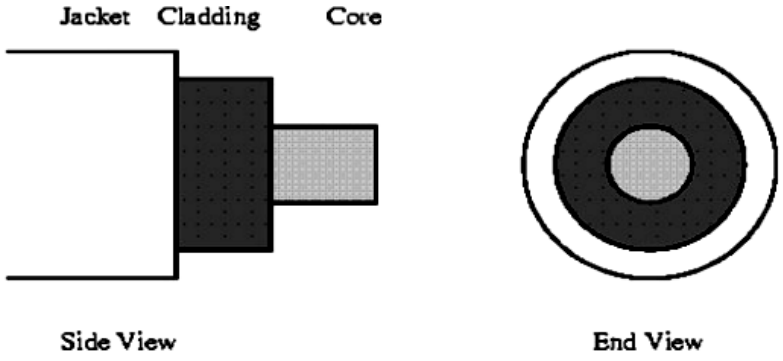


Figure Fiber Optic Cable

- Optical fiber consists of thin glass fiber that can carry information at frequencies in the visible light spectrum.
- The typical optical fiber consists of a very narrow strand of glass called the cladding.
- A typical core diameter is 62.5 microns.
- Typically cladding has a diameter of 125 minors. Coating the cladding is a protective coating consisting of plastic, it is called the jacket.
- The device generating the message has it in electromagnetic form (electrical signal); this has to be converted into light (i.e. optical signal) to send it on optic fiber cable. The process of converting light to electric signal is done on the receiving side.

Advantages:

1. **Small size and light weight**: The size of the optical fibers is very small.Therefore a large number of optical fibers can fit into a cable of small diameter.

2. **Easy availability and low cost**: The material used for the manufacturing of optical fibers is “Silica glass”. this material is easily available. So the optical fibers cost lower than the cables with metallic conductors.

3. **No electrical or electromagnetic interference:** Since the transmission takes place in the form of light rays the signal is not affected due to any electrical or electromagnetic Interference.

4. **Large Bandwidth**: As the light rays have a very high frequency in GHz range, the bandwidth of the optical fiber is extremely large.

5. **Other advantages**: - No cross talk inside the optical fiber cable. Signal can be sent up to 100 times faster.

10.4 UNGUIDED (WIRELESS) TRANSMISSION MEDIUM

- Unguided media transport data without using a physical conductor. This type of communication is often referred to as wireless communication.

- It uses wireless electromagnetic signals to send data.

- There are three types of Unguided Media
  (i) Radio waves
  (ii) Micro waves
  (iii) Infrared.

- Before understanding the different types of wireless transmission medium, let us first understand the ways in which wireless signals travel. These signals can be sent or propagated in the following three ways:
  1. Ground-wave propagation
  2. Sky-wave propagation
  3. Line-of-sight propagation

1. Ground-wave propagation

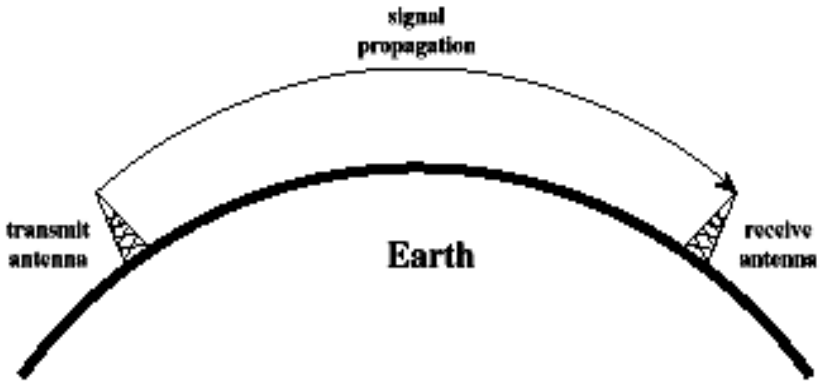


Figure : Ground Propagation of waves

Characteristics of Ground-wave propagation are as follows:

i. Follows contour of the earth
ii. Can Propagate considerable distances
iii. Frequencies up to 2 MHz
iv. Example
    a. AM radio

2. Sky-wave propagation

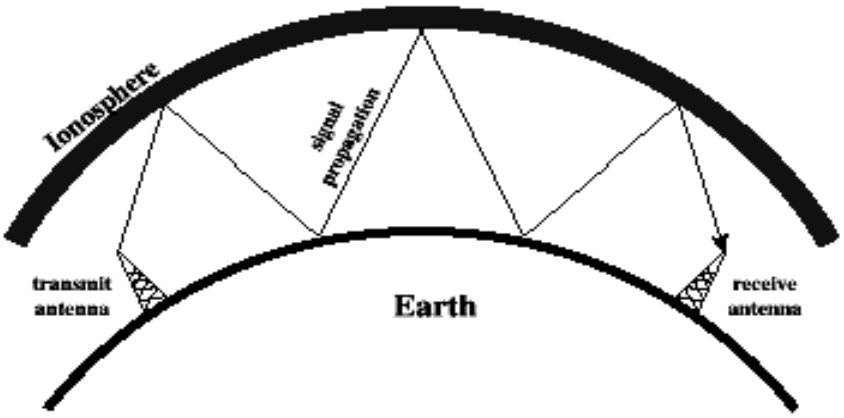


Figure :of waves

Characteristics of Sky Propagation are as follows:

i. Signal reflected from ionized layer of atmosphere back down to earth
ii. Signal can travel a number of hops, back and forth between ionosphere and earth's surface
iii. Reflection effect caused by refraction
iv. Examples
    a. Amateur radio
    b. CB radio

3. Line-of-sight propagation

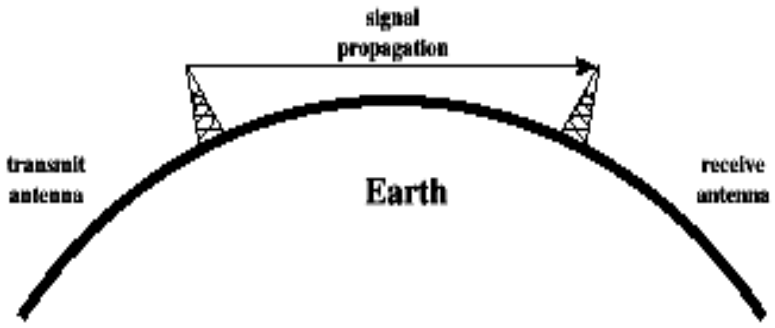


Figure : Line of Sight Propagation of waves

Characteristics of Line of Sight Propagation are as follows:

i. Transmitting and receiving antennas must be within line of sight
    a. Satellite communication – signal above 30 MHz not reflected by ionosphere
    b. Ground communication – antennas within *effective* line of site due to refraction

1. **Radio waves:**

- Electromagnetic wave ranging in frequencies between 3 KHz and 1GHz are normally called radio waves.

- Radio waves are omni-directional when an antenna transmits radio waves they are propagated in all directions. This means that sending and receiving antenna do not have to he aligned. A sending antenna can send waves that can be received by any receiving antenna.

- Radio waves particularly those waves that propagate in sky mode, can travel long distances. This makes radio waves a good candidate for long-distance broadcasting such as AM radio.

- Radio waves particularly those of low and medium frequencies can penetrate walls. It is an advantage because; an AM radio can receive signals inside a building. It is the disadvantage because we cannot isolate a communication to first inside or outside a building.

2. Microwaves:

- Electromagnetic waves having frequencies between 1 and 300 GHz are called microwaves.

- Microwaves are unidirectional; when an antenna transmits microwaves they can be narrowly focused. This means that the sending and receiving antennas need to be aligned. The unidirectional property has an obvious advantage. A pair of antennas can be aligned without interfering with another pair of aligned antennas.

- Microwaves propagation is line-of-sight. Since the towers with the mounted antennas needs to be in direct sight of each other, towers that are far apart need to be very tall, the curvature of the earth as well as other blocking obstacles do not allow two short towers to communicate using microwaves, Repeaters are often needed for long distance communication very high frequency microwaves cannot penetrate walls.

- Parabolic dish antenna and horn antenna are used for this means of transmission

3. Infrared

- Infrared signals with frequencies ranges from 300 GHz to 400 GHz can be used for short range communication.

- Infrared signals, having high frequencies, cannot penetrate walls. This helps to prevent interference between one system and another. Infrared Transmissionin one room cannot be affected by the infrared transmission in another room.

- Infrared band, has an excellent potential for data transmission. Transfer digital data is possible with a high speed with a very high frequency. There are number of computer devices which are used to send the data through infrared medium e.g. keyboard mice, PCs and printers. There are some manufacturers provide a special part called the IrDA port that allows a wireless keyboard to communicate with a PC.

10.5 COMPARISON BETWEEN WIRED AND WIRELESS MEDIA

Wired media	Wireless media
The signal energy is contained and guided within a solid medium	The signal energy propagates in the form of unguided electromagnetic waves.
Twisted pair wires, coaxial cable, optical fiber cables are the examples of wired media.	Radio and infrared lights are the examples of wireless media.
Used for point to point communication	Used for radio broadcasting in all direction
Wired media lead to discrete network topology	Wireless media leads to continuous network topology
Additional transmission capacity can be procured by adding more wire	It is not possible procure additional capacity.
Installation is costly and time consuming	Installation needs less time and money
Attenuation depends exponentially on the distance	Attenuation is proportional to square of the distance.

10.6 COMPARISON BETWEEN TWISTED PAIR CABLE, CO-AXIAL CABLE AND OPTICAL FIBER

Twisted pair cable	Co-axial cable	Optical fiber
Transmission of signals take place in the electrical form over the metallic conducting wires.	Transmission of signals take place in the inner conductor of the cable	Signal transmission takes place in an optical form over a glass fiber.
Noise immunity is low. Therefore more distortion	Higher noise immunity than the twisted pair cable due to the presence of shielding conductor	Higher noise immunity as the light rays are unaffected by the electrical noise.
Affected due to external magnetic field	Less affected due to external magnetic field	Not affected by the external magnetic field.
Short circuit between the two conductor is possible	Short circuit between the two conductor is possible	Short circuit is not possible
Cheapest	Moderately expensive	Expensive
Can support low data rates	Moderately high data rate	Very high data rates.
Low bandwidth	Moderately high bandwidth	Very high bandwidth
Easy to installed	Installation is fairly easy	Installation is difficult

10.7 REVIEW QUESTIONS

1. Write short note on transmission medium and explain its different types.
2. Explain Twisted Pair Cables in detail
3. Explain Fiber Optic Cables with its advantages
4. Explain the different ways is which wireless signals propogate.

5. Write short notes on :
   a) Radio waves
   b) Microwaves
   c) Infrared

10.8 REFERENCES & FURTHER READING

a) Data Communication & Networking – BehrouzForouzan
b) Computer Networks – Andrew Tannenbaum

11

NETWORK TOPOLOGIES

Unit Structure

11.0 OBJECTIVES

- To understand various network strategies and topologies, you will:
- Examine three common strategies used to connect nodes on a network.
- Explore network processing strategies and establish the differences between centralized and distributed processing.
- Identify and compare three common network classifications.
- Identify and define three common network topologies.

11.1 INTRODUCTION

This chapter presents an outline on Network topology is the layout pattern of interconnections of the various elements (links, nodes, etc.) of a computer or biological network. Network topologies may be physical or logical.

Physical topology refers to the physical design of a network including the devices, location and cable installation. Logical topology refers to how data is actually transferred in a network as opposed to its physical design. In general physical topology relates to a core network whereas logical topology relates to basic network. This chapter also presents an insight into the various networking strategies and the platform needed for networking.

11.2 AN OVERVIEW OF NETWORK

A network consists of two or more computers that are linked in order to share resources (such as printers and CDs), exchange files, or allow electronic communications. The computers on a network may be linked through cables, telephone lines, radio waves, satellites, or infrared light beams.

Two very common types of networks include:

- Local Area Network (LAN)
- Wide Area Network (WAN)

You may also see references to a Metropolitan Area Networks (MAN), a Wireless LAN (WLAN), or a Wireless WAN (WWAN).

11.3 WHAT IS A NETWORK TYPE

11.3.1 Local Area Network

A Local Area Network (LAN) is a network that is confined to a relatively small area. It is generally limited to a geographic area such as a writing lab, school, or building.

Computers connected to a network are broadly categorized as servers or workstations. Servers are generally not used by humans directly, but rather run continuously to provide "services" to the other computers (and their human users) on the network. Services provided can include printing and faxing, software hosting, file storage and sharing, messaging, data storage and retrieval, complete access control (security) for the network's resources, and many others.

Workstations are called such because they typically do have a human user which interacts with the network through them. Workstations were traditionally considered a desktop, consisting of a computer, keyboard, display, and mouse, or a laptop, with with integrated keyboard, display, and touchpad. With the advent of the tablet computer, and the touch screen devices such as iPad and iPhone, our definition of workstation is quickly evolving to include those devices, because of their ability to interact with the network and utilize network services.

Servers tend to be more powerful than workstations, although configurations are guided by needs. For example, a group of servers might be located in a secure area, away from humans, and only accessed through the network. In such cases, it would be common for the servers to operate without a dedicated display or keyboard. However, the size and speed of the server's processor(s), hard drive, and main memory might add dramatically to the cost of the system. On the other hand, a workstation might not need as much storage or working memory, but might require an expensive display to accommodate the needs of its user. Every computer on a network should be appropriately configured for its use.

11.3.2 Wide Area Network

Wide Area Networks (WANs) connect networks in larger geographic areas, such as Maharashtra, India, or the world. Dedicated transoceanic cabling or satellite uplinks may be used to connect this type of global network.

Using a WAN, schools in Maharashtra can communicate with places like Tokyo in a matter of seconds, without paying enormous phone bills. Two users a half-world apart with workstations equipped with microphones and a webcams might teleconference in real time. A WAN is complicated. It uses multiplexers, bridges, and routers to connect local and metropolitan networks to global communications networks like the Internet. To users, however, a WAN will not appear to be much different than a LAN.

11.4Comparing types of network coverage

The table below compares the three types of networks:

LAN	MAN	WAN
Relatively small.	Can incorporate multiple LANs.	Uses data transmission networks to incorporate LANs and MANs.
Contained within a single building or campus.	Contained within a single city or metropolitan area.	Essentially unlimited geographic area.
Generally inexpensive to implement and maintain.	Expensive to implement and maintain.	Cost varies widely, depending on how it is configured.
Typically owned privately.	Typically owned by private providers.	

11.5 AN ILLUSTRATED EXAMPLE OF A UNIVERSITY NETWORK

Advantages of Installing a Network

- *User access control.*
  Modern networks almost always have one or more servers which allows centralized management for users and for network resources to which they have access. User credentials on a privately-owned and operated network may be as simple as a user name and password, but with ever-increasing attention to computing security issues, these servers are critical to ensuring that sensitive information is only available to authorized users.

- *Information storing and sharing.*
  Computers allow users to create and manipulate information. Information takes on a life of its own on a network. The network provides both a place to store the information and mechanisms to share that information with other network users.

- *Connections.*
  Administrators, instructors, and even students and guests can be connected using the campus network.

- *Services.*
  The institution can provide services, such as registration, college directories, course schedules, access to research, and email accounts, and many others. (Remember, network services are generally provided by servers).

- *Internet.*
  The institution can provide network users with access to the internet, via an internet gateway.

- *Computing resources.*
  The institution can provide access to special purpose computing devices which individual users would not normally own. For example, an institution network might have high-speed high quality printers strategically located around a campus for instructor or student use.

- *Flexible Access.*
  Institution networks allow students to access their information from connected devices throughout the school. Students can begin an assignment in their classroom, save part of it on a public access area of the network, then go to the media center after school to finish their work. Students can also work cooperatively through the network.

- *Workgroup Computing.*

Collaborative software allows many users to work on a document or project concurrently. For example, educators located at various institution within a county could simultaneously contribute their ideas about new curriculum standards to the same document, spreadsheets, or website.

Disadvantages of Installing a Network

- *Expensive to Install.*
  Large campus networks can carry hefty price tags. Cabling, network cards, routers, bridges, firewalls, wireless access points, and software can get expensive, and the installation would certainly require the services of technicians. But, with the ease of setup of home networks, a simple network with internet access can be setup for a small campus in an afternoon.

- *Requires Administrative Time.*
  Proper maintenance of a network requires considerable time and expertise. Many schools have installed a network, only to find that they did not budget for the necessary administrative support.

- *Servers Fail.*
  Although a network server is no more susceptible to failure than any other computer, when the files server "goes down" the entire network may come to a halt. Good network design practices say that critical network services (provided by servers) should be redundant on the network whenever possible.

- *Cables May Break.*
  The Topology chapter presents information about the various configurations of cables. Some of the configurations are designed to minimize the inconvenience of a broken cable; with other configurations, one broken cable can stop the entire network.

- *Security and compliance.*
  Network security is expensive. It is also very important. An institution network would possibly be subject to more

stringent security requirements than a similarly-sized corporate network, because of its likelihood of storing personal and confidential information of network users, the danger of which can be compounded if any network users are minors. A great deal of attention must be paid to network services to ensure all network content is appropriate for the network community it serves.

11.6 WHAT IS A TOPOLOGY?

A *topology* is a description of the layout of a specific region or area. A *network topology* is a description of the layout of the region or area covered by that network.

There are two types of connections that describe how many devices connect to a single cable or segment of transmission media. They are: point-to-point and multi-point.

Point-to-point connections provide a direct link between two devices; for example, a computer connected directly to a printer, or a modem to a mainframe.

Multi-point connections provide a link between three or more devices on a network. All computer networks rely upon point-to-point and multi-point connections.

11.6.1 The Technical Connotation of Topology

The virtual shape or structure of a network is referred as topology.

The pattern or layout of interconnections of different elements or nodes of a computer network is a network topology that might be logical or physical.

However, the complete physical structure of the cable (or transmission media) is called the *physical topology*. The physical topology of a network refers to the configuration of cables, computers, and other peripherals.

The way data flows through the network (or transmission media) is called the *logical topology*. A logical topology is the method used to pass information between workstations.

11.6.2What are the Basic Types of Topology?

There are seven basic topologies in the study of network topology:

1. Point-to-point topology,
2. Bus (point-to-multipoint) topology,
3. Ring topology,
4. Star topology,
5. Hybrid topology,
6. Mesh topology and
7. Tree topology.

The interconnections between computers whether logical or physical are the foundation of this classification.

Logical topology is the way a computer in a given network transmits information, not the way it looks or connected, along with the varying speeds of cables used from one network to another.

On the other hand the **physical topology** is affected by a number of factors:

- Troubleshooting technique,
- Installation cost,
- Office layout and
- Cables' types.

The physical topology is figured out on the basis of a network's capability to access media and devices, the fault tolerance desired and the cost of telecommunications circuits.

The classification of networks by the virtue of their physical span is as follows: Local Area Networks (LAN), Wide Area Internetworks (WAN) and Metropolitan Area Networks or campus or building internetworks.

11.6.3 How Is the Physical Topology Classified?

- **Point-to-Point Network Topology**

It is the basic model of typical telephony. The simplest topology is a permanent connection between two points. The value

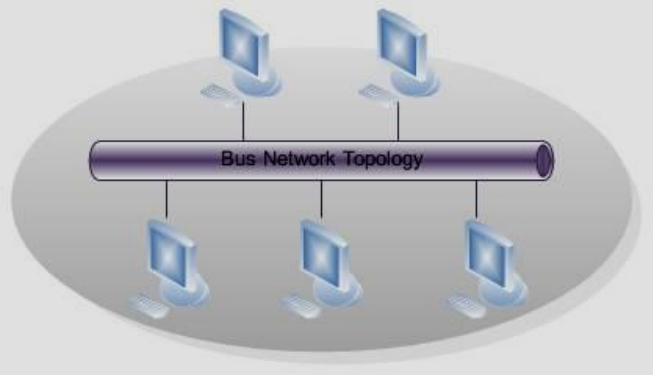
of a demanding point-to-point network is proportionate to the number of subscribers' potential pairs. It is possible to establish a permanent circuit within many switched telecommunication systems: the telephone present in a lobby would always connect to the same port, no matter what number is being dialed. A switch connection would save the cost between two points where the resources could be released when no longer required.

- **Bus Network Topology**

LANs that make use of bus topology connects each node to a single cable. Some connector connects each computer or server to the bus cable. For avoiding the bouncing of signal a terminator is used at each end of the bus cable. The source transmits a signal that travels in both directions and passes all machines unless it finds the system with IP address, the intended recipient. The data is ignored in case the address is unmatched. The installation of one cable makes bus topology an inexpensive solution as compared to other topologies; however the maintenance cost is high. If the cable is broken all systems would collapse.

- **Linear Bus:** If all network nodes are connected to a combine transmission medium that has two endpoints the Bus is Linear. The data transmitted between these nodes is transmitted over the combine medium and received by all nodes simultaneously.

- **Distributed Bus:** If all network nodes are connected to a combine transmission medium that has more than two endpoints created by branching the main section of the transmitting medium.



A linear bus topology consists of a main run of cable with a terminator at each end (See fig. 1). All nodes (file server, workstations, and peripherals) are connected to the linear cable. *A bus topology* uses one long cable (backbone) to which network

devices are either directly attached or are attached by using short drop cables. Because all workstations share this bus, a workstation checks for any information that might be coming down the backbone before sending their messages. All messages pass the other workstations on the way to their destinations. Each workstation then checks the address of each message to see if it matches its own. Note that bus network topologies, the backbone must be terminated at both ends to remove the signal from the wire after it has passed all devices on the network.

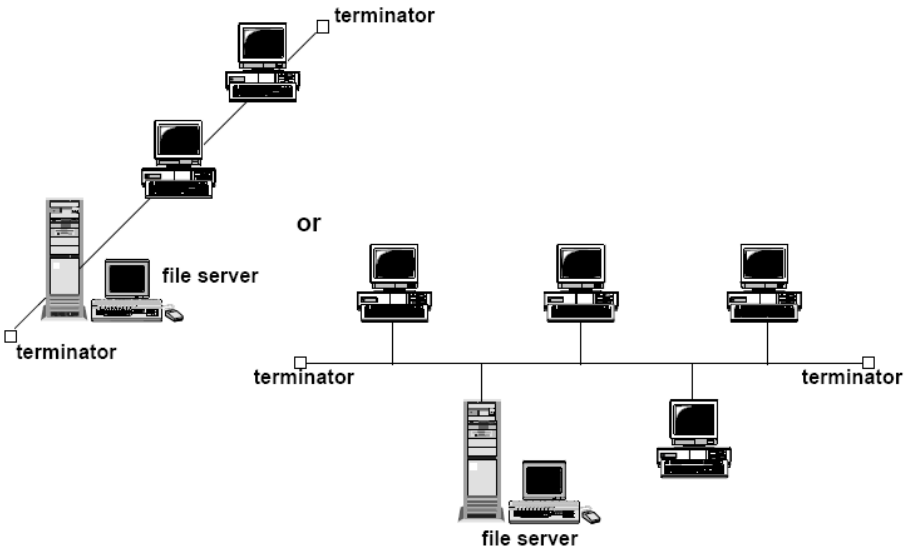


Fig. 1. Linear Bus topology

Advantages of a Linear Bus Topology

- Easy to connect a computer or peripheral to a linear bus.
- Requires less cable length than a star topology.

Disadvantages of a Linear Bus Topology

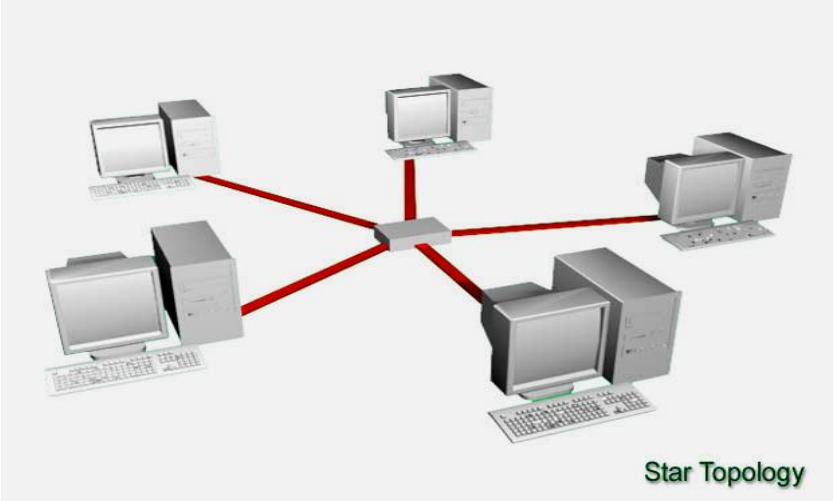
- Entire network shuts down if there is a break in the main cable.
- Terminators are required at both ends of the backbone cable.
- Difficult to identify the problem if the entire network shuts down.
- Not meant to be used as a stand-alone solution in a large building.

- **Star Network Topology**

The topology when each network host is connected to a central hub in LAN is called Star. Each node is connected to the hub with a point-to-point connection. All traffic passes through the hub that serves as a repeater or signal booster. The easiest topology to install is hailed for its simplicity to add more nodes but criticized for making hub the single point of failure. The network could be BMA (broadcast multi-access) or NBMA (non-broadcast multi-access) depending on whether the signal is automatically propagated at the hub to all spokes or individually spokes with those who are addressed.

- **Extended Star:** A network that keeps one or more than one repeaters between the central node or hub and the peripheral or the spoke node, supported by the transmitter power of the hub and beyond that supported by the standard of the physical layer of the network.

- **Distributed Star:** The topology is based on the linear connectivity that is Daisy Chained with no top or centre level connection points.



Advantages of a Star Topology

- Easy to install and wire.
- No disruptions to the network when connecting or removing devices.
- Easy to detect faults and to remove parts.

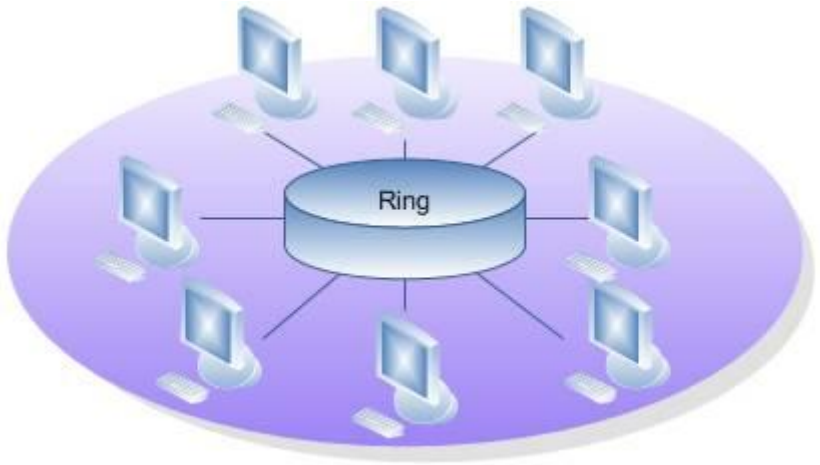
Disadvantages of a Star Topology

- Requires more cable length than a linear topology.
- If the hub, switch, or concentrator fails, nodes attached are disabled.
- More expensive than linear bus topologies because of the cost of the hubs, etc.

Ring Network Topology

Ring topology is one of the old ways of building computer network design and it is pretty much obsolete. FDDI, SONET or Token Ring technologies are used to build ring technology. It is not widely popular in terms of usability but incase if you find it anywhere it will mostly be in schools or office buildings.

Such physical setting sets up nodes in a circular manner where the data could travel in one direction where each device on the right serves as a repeater to strengthen the signal as it moves ahead.

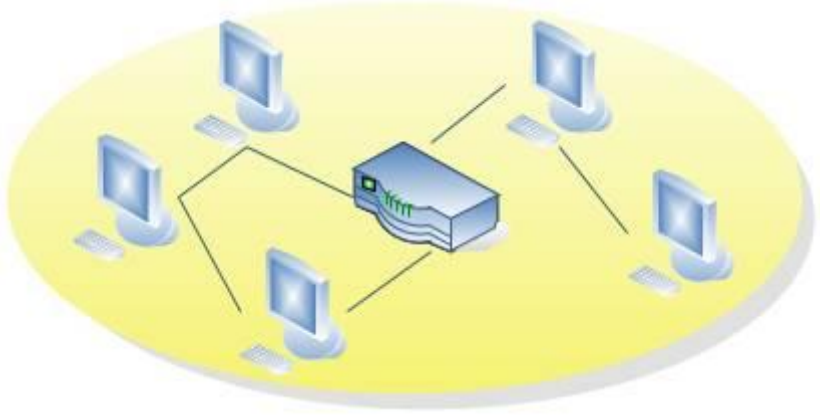


- **Mesh Network Topology**

The exponent of the number of subscribers is proportionate to the value of the fully meshed networks.

- **Fully Connected:** For practical networks such topology is too complex and costly but highly recommended for small number of interconnected nodes.

- **Partially Connected:** This set up involves the connection of some nodes to more than one nodes in the network via point-to-point link. In such connection it is possible to take advantage of the redundancy without any complexity or expense of establishing a connection between each node.

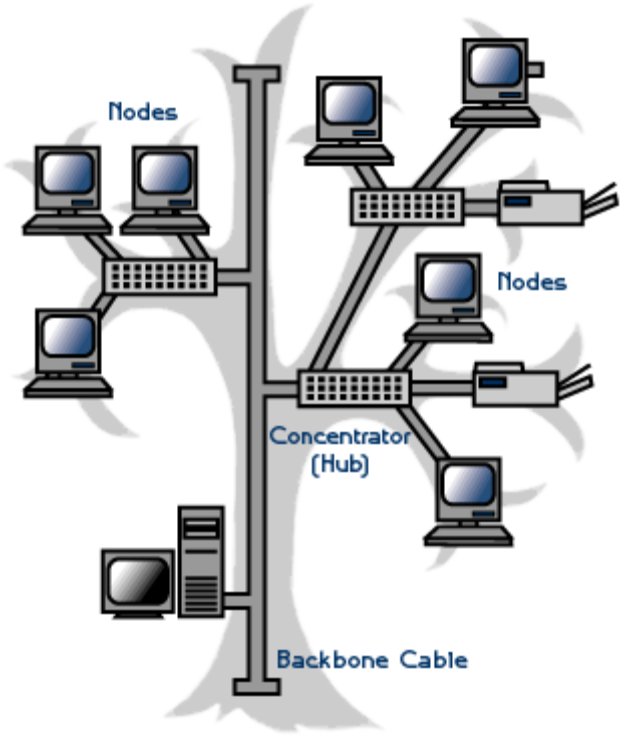


- **Hybrid Topology**

  Hybrid topologies are a combination of two or more different topologies. WANs sometimes have hybrid topologies because they connect a variety of LAN topologies. The big advantage of hybrid topologies is that they connect disparate topologies. However, the disadvantage of hybrid topologies is that they are potentially complex to establish and manage.

- **Tree Network Topology**

The top level of the hierarchy, the central root node is connected to some nodes that are a level low in the hierarchy by a point-to-point link where the second level nodes that are already connected to central root would be connected to the nodes in the third level by a point-to-point link. The central root would be the only node having no higher node in the hierarchy. The tree hierarchy is symmetrical. The BRANCHING FACTOR is the fixed number of nodes connected to the next level in the hierarchy. Such network must have at least three levels. Physical Linear Tree Topology would be of a network whose Branching Factor is one.



Advantages of a Tree Topology

- Point-to-point wiring for individual segments.
- Supported by several hardware and software venders.

Disadvantages of a Tree Topology

- Overall length of each segment is limited by the type of cabling used.
- If the backbone line breaks, the entire segment goes down.
- More difficult to configure and wire than other topologies.

Considerations When Choosing a Topology

- **Money**. A linear bus network may be the least expensive way to install a network; you do not have to purchase concentrators.

- **Length of cable needed**. The linear bus network uses shorter lengths of cable.

- **Future growth**. With a star topology, expanding a network is easily done by adding another concentrator.

- C**able type**. The most common cable in schools is unshielded twisted pair,

11.7 SUMMARY

- Knowledge of networking topologies is of core importance of computer networking design. Computer networks can only be developed using the knowledge about these topologies and decide to which topology design is best suited according to the requirement.

- A computer **network** consists of **nodes** and communication **links** whichimplement its **protocols**. It interconnects a set of **hosts** which conform to thenetwork protocols.

- A network may be classified as a **LAN**, **MAN**, or **WAN**, depending on itsgeographic spread, and as **private** or **public**, depending on its accessrestrictions.

- It may employ a **point-to-point** or a **broadcast** communication model. A point-to-point model may be based on **circuit switching** or **packetswitching**.

Exercises:

1. Classify the networks operated and/or utilized by your organization as LAN, MAN,WAN, private, public, point-to-point, broadcast, circuit-switched, or packetswitched.

2. Discuss and compare the advantages and disadvantages of circuit switching versuspacket switching. Name at least one well-known network which is based on eithertype of switching.

11.8 REVIEW QUESTION

1. What is a Network?

2. Explain Lan, Man, Wan?

3. Write a short note on Network coverage?

4. ExplainAn Illustrated Example of a University Network

5. *What is a Topology?*

6.What are the Basic Types of Topology?

7.How Is the Physical Topology Classified?

11.9 REFERENCES

http://www.garymgordon.com/misc/tutorials/networking/Lesson2.pdf

http://networkworld.com/ns/books/ciscopress/samples/0735700745.pdf

http://pages.cs.wisc.edu/~tvrdik/7/html/Section7.html

http://fcit.usf.edu/network/chap2/chap2.htm

www.pragsoft.com Chapter 4: The Network Layer 57

12

INTRODUCTION TO ROUTING

Unit Structure

12.0 OBJECTIVE

- Introduction to switches and router
- Routing concept
- Concept of switching
- Routing algorithms
- Static and dynamic routing
- Routing metrics

12.0 WHAT IS ROUTING?

Routing is the act of moving information across an internetwork from a source to a destination. Along the way, at least one intermediate node typically is encountered. Routing is often contrasted with bridging, which might seem to accomplish precisely the same thing to the casual observer. The primary difference between the two is that bridging occurs at Layer 2 (the link layer) of the OSI reference model, whereas routing occurs at Layer 3 (the network layer). This distinction provides routing and bridging with different information to use in the process of moving information from source to destination, so the two functions accomplish their tasks in different ways.

12.1.1 Routing Components

Routing involves two basic activities: determining optimal routing paths and transporting information groups (typically called packets) through an internetwork. In the context of the routing process, the latter of these is referred to as packet switching. Although packet switching is relatively straightforward, path determination can be very complex.

12.1.2 Path Determination

Routing protocols use metrics to evaluate what path will be the best for a packet to travel. A metric is a standard of measurement, such as path bandwidth, that is used by routing algorithms to determine the optimal path to a destination. To aid the process of path determination, routing algorithms initialize and maintain routing tables, which contain route information. Route information varies depending on the routing algorithm used.

Routing algorithms fill routing tables with a variety of information. Destination/next hop associations tell a router that a particular destination can be reached optimally by sending the packet to a particular router representing the “next hop” on the way to the final destination. When a router receives an incoming packet, it checks the destination address and attempts to associate this address with a next hop.

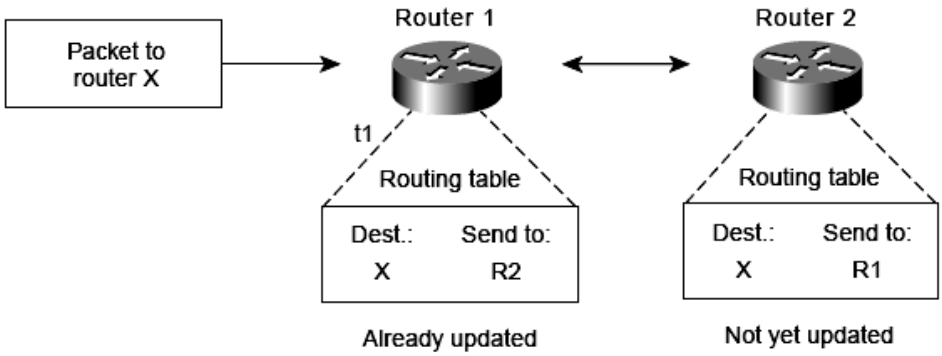


Fig. Destination/Next Hop Associations Determine the Data's Optimal Path

Routing tables also can contain other information, such as data about the desirability of a path. Routers compare metrics to determine optimal routes, and these metrics differ depending on the design of the routing algorithm used.

Routers communicate with one another and maintain their routing tables through the transmission of a variety of messages. The routing update message is one such message that generally consists of all or a portion of a routing table. By analyzing routing updates from all other routers, a router can build a detailed picture of network topology. A link-state advertisement, another example of a message sent between routers, informs other routers of the state of the sender's links. Link information also can be used to build a complete picture of network topology to enable routers to determine optimal routes to network destinations.

12.2 SWITCHING

Switching algorithms is relatively simple; it is the same for most routing protocols. In most cases, a host determines that it must send a packet to another host. Having acquired a router's address by some means, the source host sends a packet addressed specifically to a router's physical (Media Access Control [MAC]-layer) address, this time with the protocol (network layer) address of the destination host.

As it examines the packet's destination protocol address, the router determines that it either knows or does not know how to forward the packet to the next hop. If the router does not know how to forward the packet, it typically drops the packet. If the router knows how to forward the packet, however, it changes the destination physical address to that of the next hop and transmits the packet. The next hop may be the ultimate destination host. If not, the next hop is usually another router, which executes the same switching decision process. As the packet moves through the

internetwork, its physical address changes, but its protocol address remains constant.

The preceding discussion describes switching between a source and a destination end system. The International Organization for Standardization (ISO) has developed a hierarchical terminology that is useful in describing this process. Using this terminology, network devices without the capability to forward packets between subnetworks are called *end systems (ESs)*, whereas network devices with these capabilities are called *intermediate systems (ISs)*. ISs are further divided into those that can communicate within routing domains (*intradomain ISs*) and those that communicate both within and between routing domains (*interdomain ISs*). A routing domain generally is considered a portion of an internetwork under common administrative authority that is regulated by a particular set of administrative guidelines. Routing domains are also called autonomous systems.With certain protocols, routing domains can be divided into routing areas, but intradomain routing protocols are still used for switching both within and between areas.

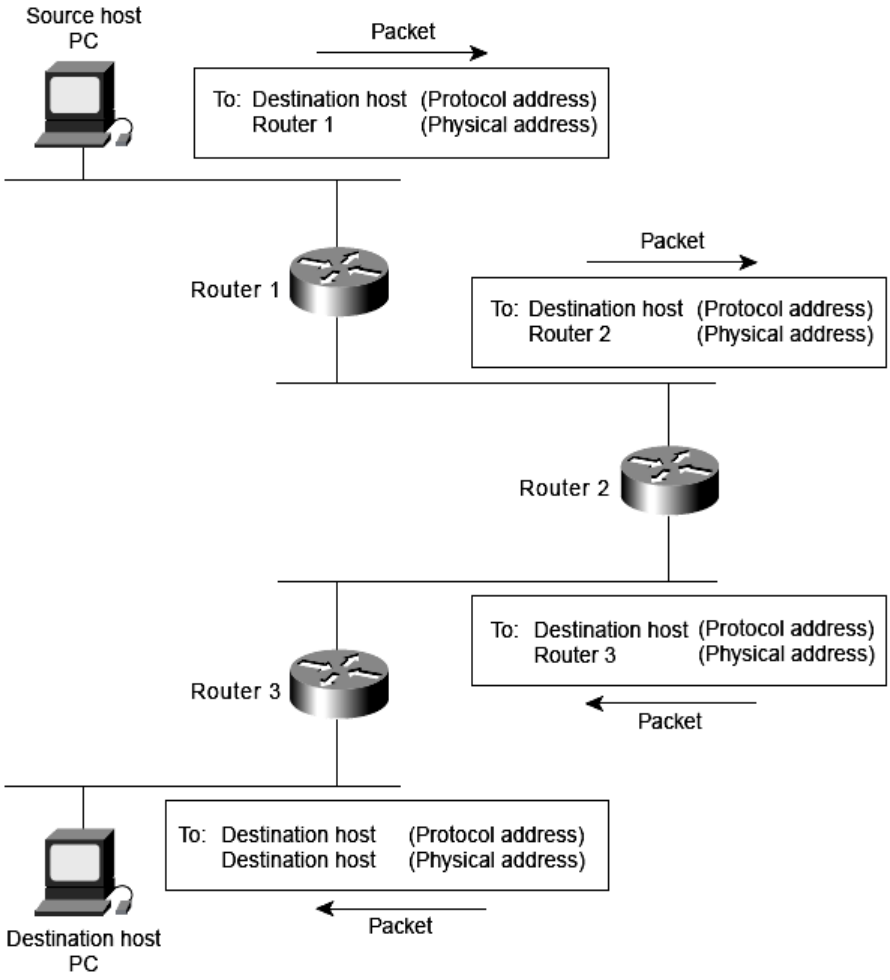


Fig.of switching

12.3 INTRODUCTION TO ALGORITHM

Routing Algorithms- Introduction

Routing algorithms can be differentiated based on several key characteristics. First, the particular goals of the algorithm designer affect the operation of the resulting routing protocol. Second, various types of routing algorithms exist, and each algorithm has a different impact on network and router resources.

Finally, routing algorithms use a variety of metrics that affect calculation of optimal routes. The following sections analyze these routing algorithm attributes.

*12.3.1*Design Goals

Routing algorithms often have one or more of the following design goals:

• Optimality

• Simplicity and low overhead

• Robustness and stability

• Rapid convergence

• Flexibility

- *Optimality* refers to the capability of the routing algorithm to select the best route, which depends on the metrics and metric weightings used to make the calculation.

For example, one routing algorithm may use a number of hops and delays, but it may weigh delay more heavily in the calculation. Naturally, routing protocols must define their metric calculation algorithms strictly.

- Routing algorithms also are designed to be as simple as possible. In other words, the routing algorithm must offer its functionality efficiently, with a minimum of software and utilization overhead. Efficiency is particularly important when the software implementing the routing algorithm must run on a computer with limited physical resources.

- Routing algorithms must be *robust*, which means that they should perform correctly in the face of unusual or unforeseen circumstances, such as hardware failures, high load conditions, and incorrect implementations. Because routers are located at network junction points, they can cause considerable problems when they fail. The best routing

algorithms are often those that have withstood the test of time and that have proven stable under a variety of network conditions.

- In addition, routing algorithms must converge rapidly. *Convergence* is the process of agreement, by all routers, on optimal routes. When a network event causes routes to either go down or become available, routers distribute routing update messages that permeate networks, stimulating recalculation of optimal routes and eventually causing all routers to agree on these routes. Routing algorithms that converge slowly can cause routing loops or network outages.

- Routing algorithms should also be flexible, which means that they should quickly and accurately adapt to a variety of network circumstances. Assume, for example, that a network segment has gone down. As many routing algorithms become aware of the problem, they will quickly select the next-best path for all routes normally using that segment. Routing algorithms can be programmed to adapt to changes in network bandwidth, router queue size, and network delay, among other variables.

➢ 12.3.2 Routing Algorithm Types

Routing algorithms can be classified by type. Key differentiators include these:

1. Static versus dynamic
2. Single-path versus multipath
3. Flat versus hierarchical
4. Host-intelligent versus router-intelligent
5. Intradomain versus interdomain
6. Link-state versus distance vector

1. Static Versus Dynamic

Static routing algorithms are hardly algorithms at all, but are table mappings established by the network administrator before the beginning of routing. These mappings do not change unless the network administrator alters them. Algorithms that use static routes are simple to design and work well in environments where network traffic is relatively predictable and where network design is relatively simple.

Because static routing systems cannot react to network changes, they generally are considered unsuitable for today's large, constantly changing networks. Most of the dominant routing algorithms today are

Dynamic routing algorithms, which adjust to changing network circumstances by analyzing incoming routing update messages. If the message indicates that a network change has occurred, the routing software recalculates routes and sends out new routing update messages. These messages permeate the network, stimulating routers to rerun their algorithms and change their routing tables accordingly.

Dynamic routing algorithms can be supplemented with static routes where appropriate. A router of last resort (a router to which all unroutable packets are sent), for example, can be designated to act as a repository for all unroutable packets, ensuring that all messages are at least handled in some way.

2. Single-Path Versus Multipath

Some sophisticated routing protocols support multiple paths to the same destination. Unlike single-path algorithms, these multipath algorithms permit traffic multiplexing over multiple lines. The advantages of multipath algorithms are obvious: They can provide substantially better throughput and reliability. This is generally called load sharing.

3. Flat Versus Hierarchical

Some routing algorithms operate in a flat space, while others use routing hierarchies.

In a *flat routing system*, the routers are peers of all others. In a hierarchical routing system, some routers form whatamounts to a routing backbone. Packets from nonbackbone routers travel to the backbone routers, wherethey are sent through the backbone until they reach the general area of the destination. At this point, theytravel from the last backbone router through one or more nonbackbone routers to the final destination.Routing systems often designate logical groups of nodes, called domains, autonomous systems, or areas.

In *hierarchical systems*, some routers in a domain can communicate with routers in other domains, while others can communicate only with routers within their domain. In very large networks, additional hierarchical levels may exist, with routers at the highest hierarchical level forming the routing backbone.

The primary advantage of hierarchical routing is that it mimics the organization of most companies and therefore supports their traffic patterns well. Most network communication occurs within small company groups (domains). Because intradomain routers need to know only about other routers within their domain, their routing algorithms can be simplified, and, depending on the routing algorithm being used, routing update traffic can be reduced accordingly.

4. Host-Intelligent Versus Router-Intelligent

Some routing algorithms assume that the source end node will determine the entire route. This is usually referred to as *source routing*. In source-routing systems, routers merely act as store-and-forward devices, mindlessly sending the packet to the next stop.

Other algorithms assume that hosts know nothing about routes. In these algorithms, routers determine the path through the internetwork based on their own calculations. In the first system, the hosts have the routing intelligence. In the latter system, routers have the routing intelligence.

5. Intradomain Versus Interdomain

Some routing algorithms work only within domains; others work within and between domains. The nature of these two algorithm types is different. It stands to reason, therefore, that an optimal intradomain-routing algorithm would not necessarily be an optimal interdomain-routing algorithm.

6. Link-State Versus Distance Vector

Link-state algorithms (also known as shortest path first algorithms) flood routing information to all nodes in the internetwork. Each router, however, sends only the portion of the routing table that describes the state of its own links. In link-state algorithms, each router builds a picture of the entire network in its routing tables.

Distance vector algorithms (also known as Bellman-Ford algorithms) call for each outer to send all or some portion of its routing table, but only to its neighbors. In essence, link-state algorithms send small updates everywhere, while distance vector algorithms send larger updates only to neighboring routers. *Distance vector* algorithms know only about their neighbors. Because they converge more quickly, link-state algorithms are somewhat less prone to routing loops than distance vector algorithms. On the other hand, link-state algorithms require more CPU power and memory than distance vector algorithms. Link-state algorithms, therefore, can be more expensive to implement and support. Link-state protocols are generally more scalable than distance vector protocols.

12.4 ROUTING METRICS

Routing tables contain information used by switching software to select the best route. Routing algorithms have used many different metrics to determine the best route. Sophisticated routing algorithms can base route selection on multiple metrics, combining them in a single (hybrid) metric. All the following metrics have been used:

- Path length
- Reliability
- Delay
- Bandwidth Load
- Communication cost

- *Path length* is the most common routing metric. Some routing protocols allow network administrators to assign arbitrary costs to each network link. In this case, path length is the sum of the costs associated with each link traversed. Other routing protocols define hop count, a metric that specifies the number of passes through internetworking products, such as routers, that a packet must take en route from a source to a destination.

- *Reliability*, in the context of routing algorithms, refers to the dependability (usually described in terms of the bit-error rate) of each network link. Some network links might go down more often than others. After a network fails, certain network links might be repaired more easily or more quickly than other links. Any reliability factors can be taken into account in the assignment of the reliability ratings, which are arbitrary numeric values usually assigned to network links by network administrators.

- *Routing delay* refers to the length of time required to move a packet from source to destination through the internetwork. Delay depends on many factors, including the bandwidth of intermediate network links, the port queues at each router along the way, network congestion on all intermediate network links, and the physical distance to be travelled. Because delay is a conglomeration of several important variables, it is a common and useful metric.

- *Bandwidth* refers to the available traffic capacity of a link. All other things being equal, a 10-Mbps Ethernet link would be preferable to a 64-kbps leased line. Although bandwidth is a rating of the maximum attainable throughput on a link, routes through links with greater bandwidth do not necessarily provide better routes than routes through slower links. For example, if a faster link is busier, the actual time required to send a packet to the destination could be greater.

- *Load* refers to the degree to which a network resource, such as a router, is busy. Load can be calculated in a variety of ways, including CPU utilization and packets processed per second. Monitoring these parameters on a continual basis can be resource-intensive itself.

- *Communication cost* is another important metric, especially because some companies may not care about performance as much as they care about operating expenditures. Although line delay may be longer, they will send packets over their own lines rather than through the public lines that cost money for usage time.

12.5 SUMMARY

- *Routing* is the act of moving information across an internetwork from a source to a destination.
- Routing involves two basic activities: determining optimal routing paths and transporting information groups (typically called packets) through an internetwork.
- Routing protocols use metrics to evaluate what path will be the best for a packet to travel.
- Routing protocols use metrics to evaluate what path will be the best for a packet to travel.
- Switching algorithms is relatively simple; it is the same for most routing protocols.
- Routing algorithms often have one or more of the following design goals:
    - Optimality
    - Simplicity and low overhead
    - Robustness and stability
    - Rapid convergence
    - Flexibility

- Routing tables contain information used by switching software to select the best route.

12.6 REVIEW QUESTION

1. Explain routing concept?
2. Explain switching concept?
3. Discuss design goal?
4. Explain routing algorithms?
5. Explain Routing metrics?

12.6 LIST OF REFERENCES

- Stamper, D. (1993) *Local Area Networks*, Addison-Wesley, Reading, MA.
- Stamper, D. (1991) *Business Data Communications*, Third Edition, Addison-Wesley, Reading, MA.
- Stone, H. (1982), *Microcomputer Interfacing*, Addison-Wesley, Reading, MA.
- Tanenbaum, A. (1989), *Computer Networks*, Second Edition, Prentice Hall, Englewood Cliffs, NJ.
- Van Duuren, J., Schoute, F., and Kastelein, P. (1992) *Telecommunications*
- *Networks and Services*, Addison-Wesley, Reading, MA.
- Viniotis Y. and Onvural R. (editors) (1993) *Asynchronous Transfer Mode, Networks*, Plenum, New York, NY.
- White, G. (1992) *Internetworking and Addressing*, McGraw-Hill, NY.
- Zitsen, W. (1990) 'Metropolitan Area Networks: Taking LANs into the Public, Network,' *Telecommunications*, pp. 53-60.

13

SWITCHING CONCEPTS

Unit Structure

13.0 OBJECTIVE

- ✓ Introduce switching concept
- ✓ Define switching node
- ✓ Define packet switching
- ✓ Switching mode

13.1 INTRODUCTION

Switching is the generic method for establishing a path for point-to-point communication in a network. It involves the nodes in

the network utilizing their direct communication lines to other nodes so that a path is established in a piecewise fashion. Each node has the capability to ‘switch’ to a neighbouring node (i.e., a node to which it is directly connected) to further stretch the path until it is completed.

One of the most important functions of the network layer is to employ theswitching capability of the nodes in order to route messages across the network.There are two basic methods of switching circuit switching and packet switching.

13.2.1. Circuit Switching

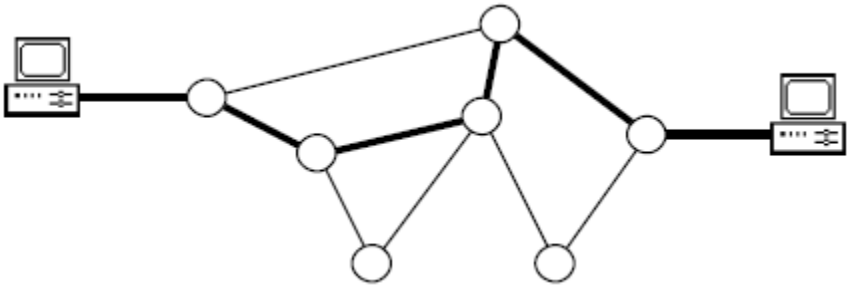


Figure 13.2.1 A ‘switched’ path.

In circuit switching, two communicating stations are connected by a *dedicated* communication path which consists of intermediate nodes in the network and thelinks that connect these nodes.

Figure **13.2.1** shows a simple circuit switch which consists of a 3×3 matrix, capable of connecting any of its inlets (*a*, *b*, and *c*) to any of its outlets (*d*, *e*, and *f*). Each crosspoint appears as a circle. A hollow circle means that the crosspoint is *off* (i.e., the two crossing wires are not connected). A solid circles means that the crosspoint is *on* (i.e., the crossing wires are connected).

Switches may also have more inlets than outlets, or more outlets than inlets.)

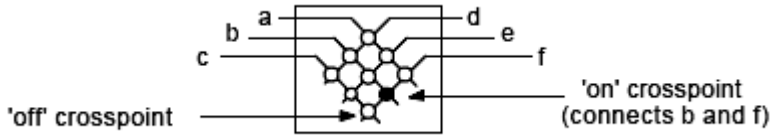


Figure 13.2.2 A simple circuit switch.

When the two hosts shown in the figure initiate a connection, the network determines a path through the intermediate switches and establishes a circuit which is maintained for the duration of the connection. When the hosts disconnect, the network releases the circuit.

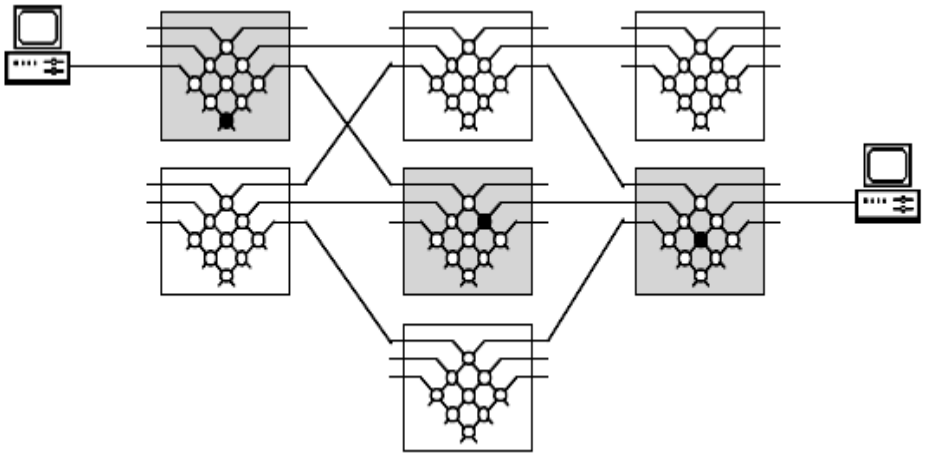


Fig 13.2.3 Circuit switching.

Communication via circuit switching implies that there is a dedicated communication path between the two stations. The path is a connected through a sequence of links between network nodes. On each physical link, a logical channel is dedicated to the connection. Circuit switching is commonly used technique in telephony, where the caller sends a special message with the address of the callee (i.e. by dialling a number) to state its destination. It involved the following three distinct steps, as shown in Fig. **13.2.3**

Circuit Establishment: To establish an end-to-end connection before any transfer of data.

Some segments of the circuit may be a dedicated link, while some other segments may be shared.

Data transfer:

Transfer data is from the source to the destination.
The data may be analog or digital, depending on the nature of the network.
The connection is generally full-duplex.

Circuit disconnect:

Terminate connection at the end of data transfer.
• Signals must be propagated to deallocate the dedicated resources.

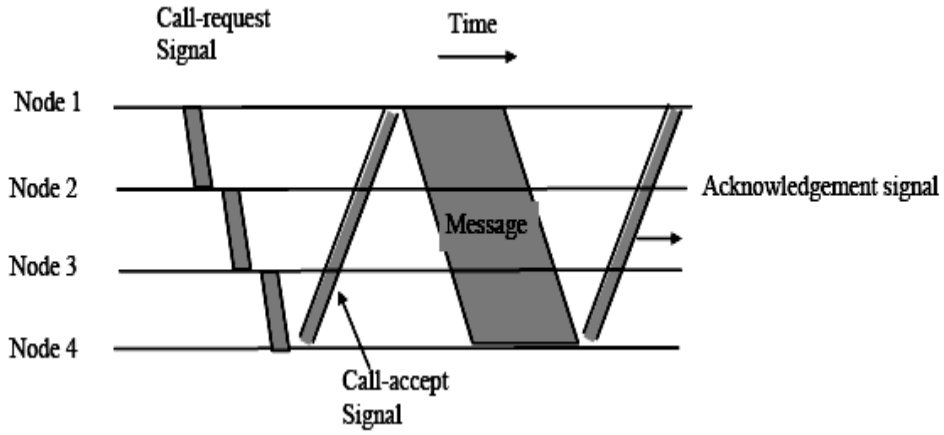


Fig: 13.2.4 Circuit Switching technique

13.2.2 Switching Node

Let us consider the operation of a single circuit switched node comprising a collection of stations attached to a central switching unit, which establishes a dedicated path between any two devices that wish to communicate.

Major elements of a single-node network are summarized below:

***Digital switch*:** That provides a transparent (full-duplex) signal path between any pair of attached devices.

Network interface:That represents the functions and hardware needed to connect digital devices to the network (like telephones).

Control unit:That establishes, maintains, and tears down a connection.

An important characteristic of a circuit-switch node is whether it is *blocking* or *non-blocking*.

A blocking network is one, which may be unable to connect two stations because all possible paths between them are already in use. A non-blocking network permits all stations to be connected (in pairs) at once and grants all possible connection requests as long as the called party is free. For a network that supports only voice traffic, a blocking configuration may be acceptable, since most phone calls are of short duration. For data applications, where a connection may remain active for hours, non-blocking configuration is desirable.

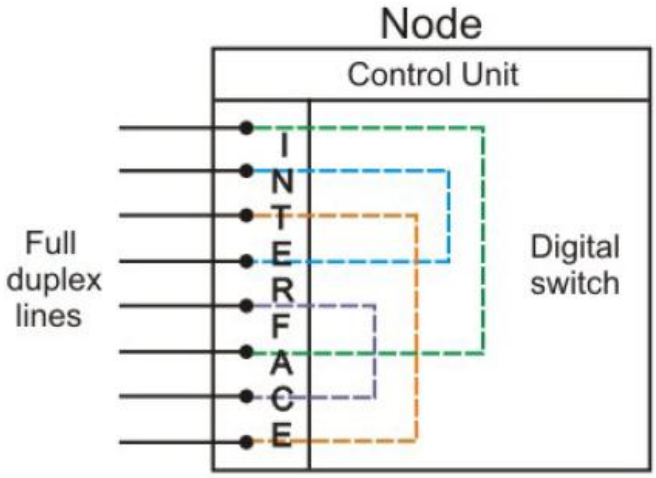


Fig 13.2.5 Schematic Diagram of a Switching node

Circuit switching uses any of the three technologies: **Space-division** switches, **Time-division** switches or a **combination of both**. In Space-division switching, the paths in the circuit are separated with each other spatially, i.e. different ongoing connections, at a same instant of time, uses different switching paths, which are separated spatially. This was originally developed

for the analog environment, and has been carried over to the digital domain. Some of the space switches are crossbar switches, Multi-stage switches (e.g. Omega Switches). A **crossbar** switch is shown in Fig. **Fig 13.2.6 .** Basic building block of the switch is a metallic crosspoint or semiconductor gate that can be enabled or disabled by a control unit.

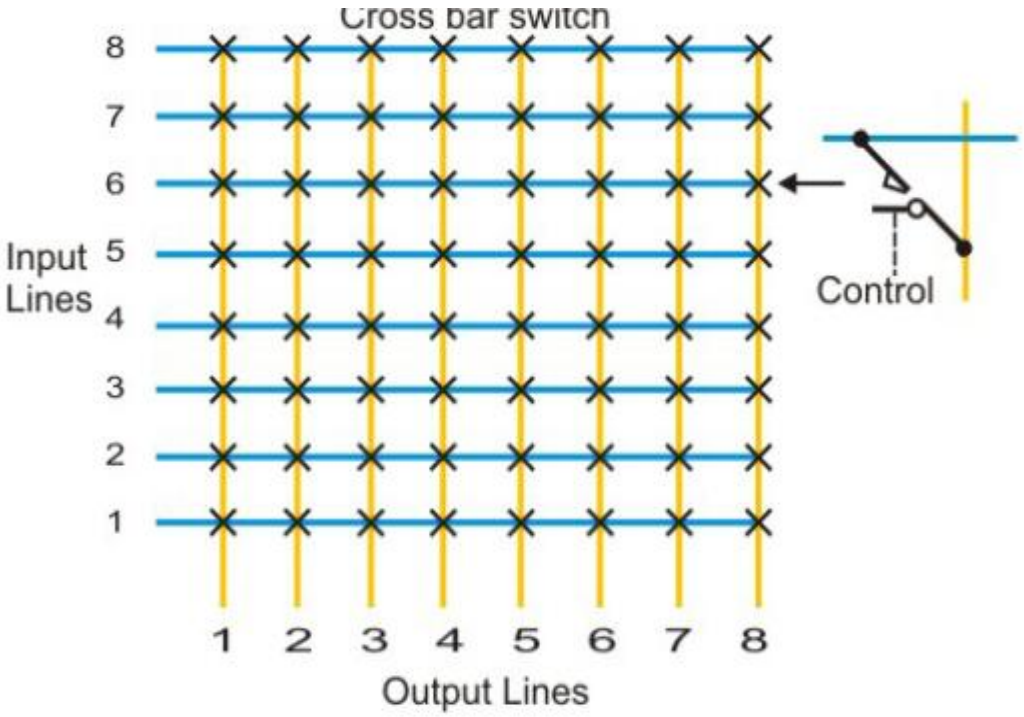


Fig 13.2.6 Schematic diagram of a crossbar switch

Limitations of crossbar switches are as follows:

- The number of crosspoints grows with the square of the number of attached stations.
- Costly for a large switch.
- The failure of a crosspoint prevents connection between the two devices whose lines intersect at that crosspoint.
- The crosspoints are inefficiently utilized.
- Only a small fraction of crosspoints are engaged even if all of the attached devices are active.

Some of the above problems can be overcome with the help of *multistage space division* switches. By splitting the crossbar switch into smaller units and interconnecting them, it is possible to build multistage switches with fewer crosspoints.

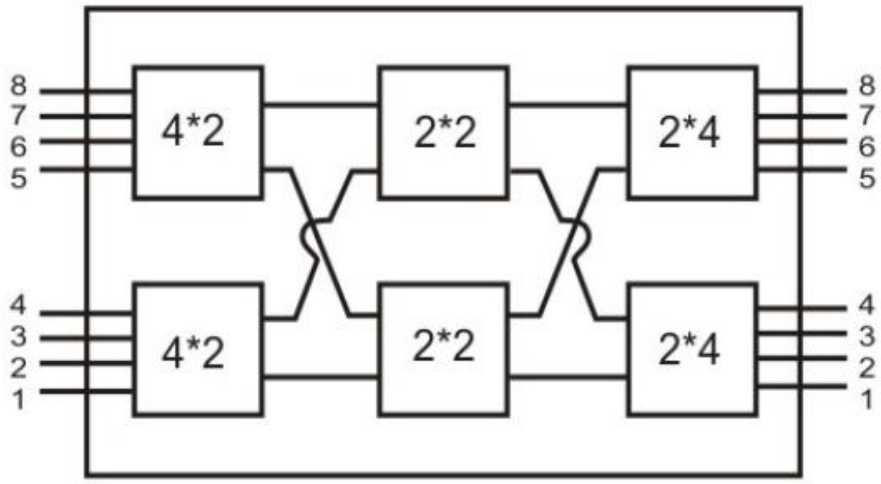


Fig 13.2.6 A three-stage space division switch

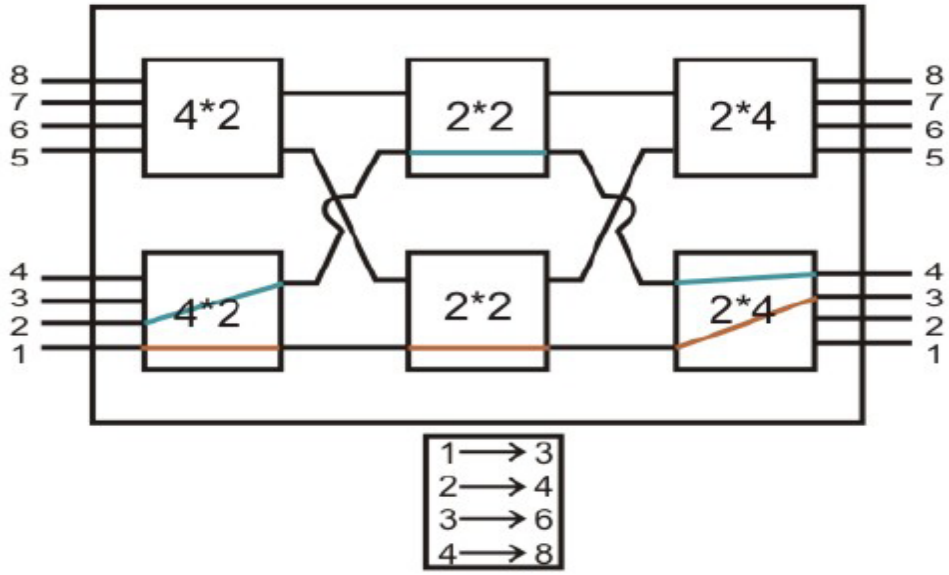
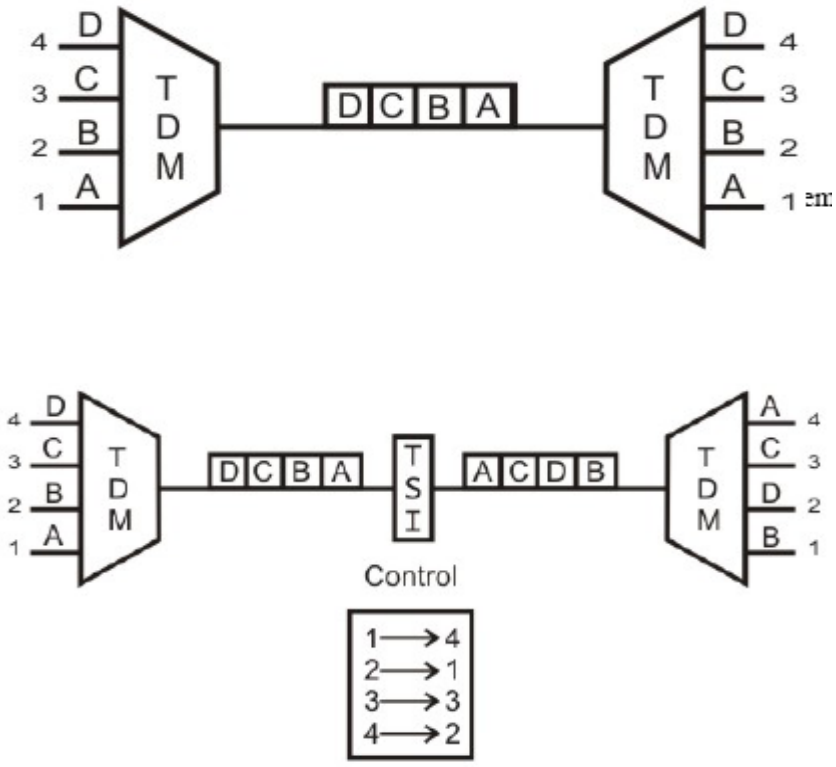


Fig 13.2.7 Block nature of the switch

Figure **Fig 13.2.6** shows a three-stage space division switch. In this case the number of crosspoints needed goes down from 64 to 40. There is more than one path through the network to connect two endpoints, thereby increasing reliability. Multistage switches may lead to *blocking*. The problem may be tackled by increasing the number or size of the intermediate switches, which also increases the cost. The blocking feature is illustrated in Fig. **Fig 13.2.7** . As shown in Fig. **Fig 13.2.7** , after setting up connections for 1-to-3 and 2-to-4, the switch cannot establish connections for 3-to-6 and 4-to-5.

13.2.3 Time Division Switching

Both voice and data can be transmitted using digital signals through the same switches. All modern circuit switches use digital time-division multiplexing (TDM) technique for establishing and maintaining circuits. Synchronous TDM allows multiple low-speed bit streams to share a high-speed line.



13.2.8 TIME Division Multiplexing

Time-division switching uses time-division multiplexing to achieve switching, i.e. different ongoing connections can use same switching path but at different interleaved time intervals.

13.2.4 Packet Switching

Packet switching was designed to address the shortcomings of circuit switching in dealing with data communication. Unlike circuit switching where communication is continuous along a dedicated circuit, in packet switching, communication is discrete in form of packets. Each packet is of a limited size and can hold up to a certain number of octets of user data. Larger messages are broken into smaller chunks so that they can be fitted into packets. In addition to user data, each packet carries additional information

(in form of a header) to enable the network to route it to its final destination.

A packet is handed over from node to node across the network. Each receiving node temporarily stores the packet, until the next node is ready to receive it, and then passes it onto the next node. This technique is called **store-and-forward** and overcomes one of the limitations of circuit switching. A packet-switched network has a much higher capacity for accepting further connections. Additional connections are usually not blocked but simply slow down existing connections, because they increase the overall number of packets in the network and hence increase the delivery time of each packet. Figure 13.2.9 shows a simple packet switch with six I/O channels (*a* through *f*). Each channel has an associated buffer which it uses to store packets in transit. The operation of the switch is controlled by a microprocessor. A packet received on anynof the channels can be passed onto any of the other channels by the microprocessor moving it to the corresponding buffer.

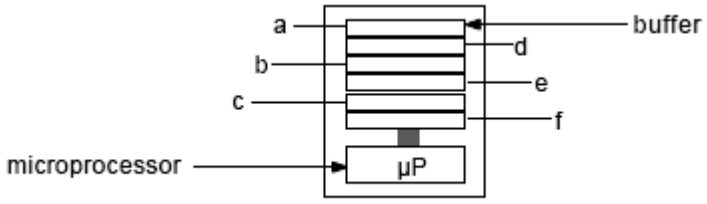


Figure 13.2.9 A simple packet switch.

Two variations of packet switching exist: virtual circuit and datagram. The **virtual circuit** method (also known as **connection-oriented**) is closer to circuit switching. Here a complete route is worked out prior to sending data packets. The route is established by sending a connection request packet along the route to the intended destination. This packet informs the intermediate nodes about the connection and the established route so that they will know how to route subsequent packets. The result is a circuit somewhat similar to those in circuit switching, except that it uses packets as its basic unit of communication. Hence it is called a virtual circuit.

Each packet carries a virtual circuit identifier which enables a node to determine to which virtual circuit it belongs and hence how

it should be handled. (The virtual circuit identifier is essential because multiple virtual circuits may pass through the same node at the same time.) Because the route is fixed for the duration of the call, the nodes spend no effort in determining how to route packets.

Fig 13.2.10 illustrates the virtual circuit method using the switch. When the two hosts initiate a connection, the network layer establishes a virtual circuit (denoted by shaded switches) which is maintained for the duration of the connection. When the hosts disconnect, the network layer releases the circuit. The packets in transit are displayed as dark boxes within the buffers. These packets travel only along the designatedvirtual circuit.

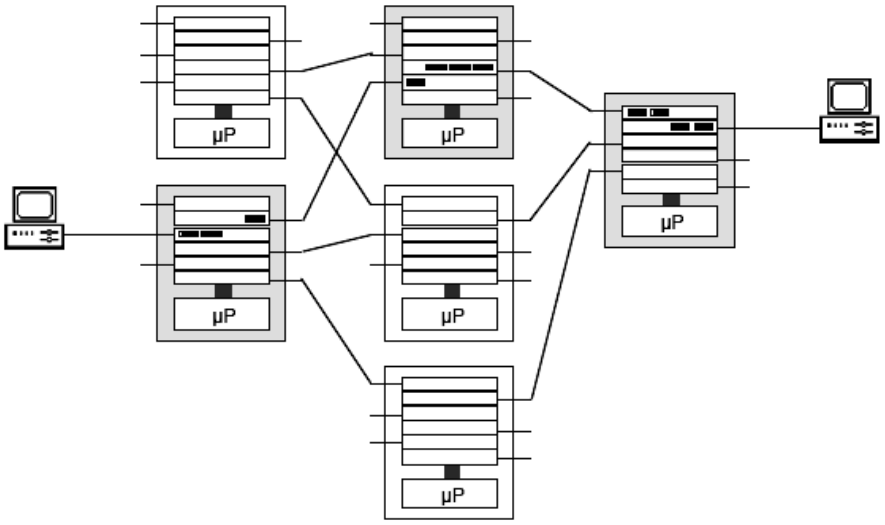
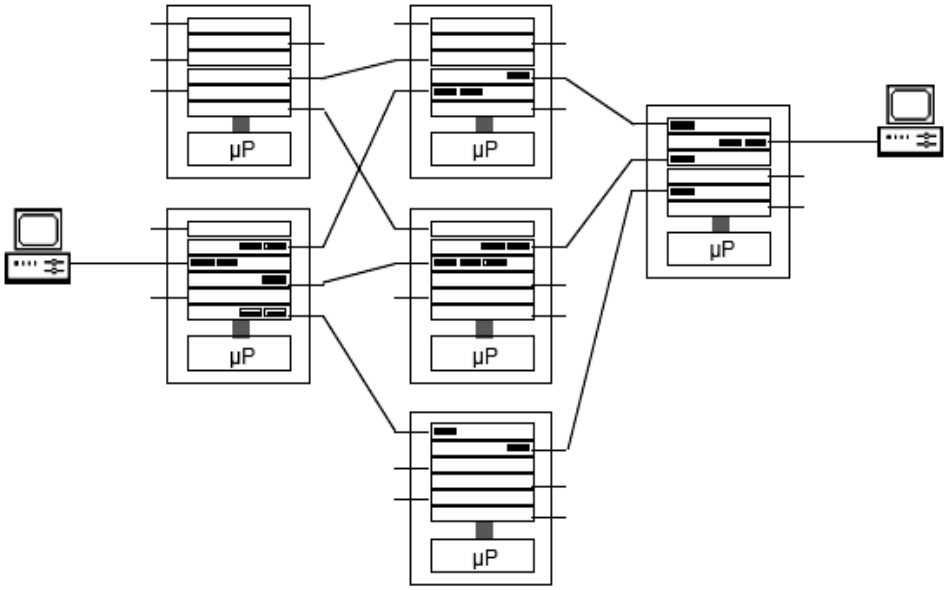


Figure 13.2.10 **Packet switching with virtual circuits.**

The **datagram** method (also known as **connectionless**) does not rely on a pre-established route, instead each packet is treated independently. Therefore, it ispossible for different packets to travel along different routes in the network to reach the same final destination. As a result, packets may arrive out of order, or even never arrive (due to node failure). It is up to the network user to deal with lost packets, and to rearrange packets to their original order. Because of the absence of a preestablishedcircuit, each packet must carry enough information in its header to enable the nodes to route it correctly.

Figure 13.2.11 illustrates the datagram method. Note how the packets exercise different routes.



13.2.11 Packet switching with datagrams.

The advantage of the datagram approach is that because there is no circuit, congestion and faulty nodes can be avoided by choosing a different route. Also, connections can be established more quickly because of reduced overheads. This makes datagrams better suited than virtual circuits for brief connections. For example, database transactions in banking systems are of this nature, where each transaction involves only a few packets.

The advantage of the virtual circuit approach is that because no separate routing is required for each packet, they are likely to reach their destination more quickly; this leads to improved throughput. Furthermore, packets always arrive in order.

Virtual circuits are better suited to long connections that involve the transfer of large amounts of data (e.g., transfer of large files). Because packet switching is the more dominant form of switching for data communication, we will focus our attention on this form of switching from now on.

13.2.5 Switching Modes

Any delay in passing traffic is known as latency. Switches offer three ways to switch the traffic depending upon how thoroughly you want the frame to be checked before it is passed on. The more checking you want, the more latency you will introduce to the switch.

The three switching modes to choose from are:

- Cut-through
- Store-and-forward
- Fragment-free

➢ **Cut-through Mode**

Cut-through switching is the fastest switching method meaning it has the lowest latency. The incoming frame is read up to the destination MAC address. Once it reaches the destination MAC address, the switch then checks its CAM table for the correct port to forward the frame out of and sends it on its way. There is no error checking, so this method gives you the lowest latency. The price, however, is that the switch will forward any frames containing errors.

The process of switching modes can best be described by using a metaphor.

➢ **Store-and-forward** Mode

Here the switch reads the entire frame and copies it into its buffers. A cyclic redundancy check (CRC) takes place to check the frame for any errors. If errors are found, the frame is dropped. Otherwise the switching table is examined and the frame forwarded.

➢ **Fragment-free (modified cut-through/runt-free) Mode**

Since cut-through can ensure that all frames are good and store-and-forward takes too long, we need a method that is both quick and reliable. Using our example of the nightclub security, imagine you are asked to make sure that everyone has an ID and that the picture matches the person. With this method you have made sure everyone is who they say they are, but you do not have to take down all the information. In switching we accomplish this by using the fragment-free method of switching.

13.3 SUMMARY

The generic method for establishing a path for point-to-point communication in anetwork is called **switching**. There are two general switching methods: circuitswitching and packet switching.

In **circuit switching** two communicating stations are connected by a dedicatedcommunication path.

In **packet switching** communication is discrete in form of packets. The packetsare handled by the intermediate nodes in a store-and-forward fashion. Packetswitching is either based on **virtual circuits** or on **datagrams**.

The task of selecting a path for the transport of packets across the network iscalled **routing**. The three classes of routing algorithms are: **flooding**, **staticrouting**, and **dynamic routing**.

13.4 REVIEW QUESTIONS

1. Define Circuit Switching
2. Explain Switching Node
3. Explain Time Division Switching

13.5 LIST OF REFERENCES

- Black, U. (1989), *Data Networks: Concepts, Theory, and Practice*, Prentice
- Hall, Englewood Cliffs, NJ.
- Gitlin, R. D., Hayes, J. F., and Weinstein, S. B. (1992) *Data Communication*
- *Principles*, Plenum, New York, NY.
- Hughes, L. (1992) *Data Communications*, McGraw-Hill, NY.
- Kessler, G. and Train, D. (1992) *Metropolitan Area Networks: Concepts,*
- *Standards, and Service*, McGraw-Hill, NY.
- Martin, J. and Leben, J. (1988), *Principles of Data Communication*, Prentice
- Hall, Englewood Cliffs, NJ.
- Spohn, D. (1993) *Data Network Design*, McGraw-Hill, NY.
- Stallings, W. (1990), *Handbook of Computer Communications Standards*, *Volumes I and II*, Howard Sams and Company, Carmel.
- Stallings, W. (1992), *ISDN and Broadband ISDN*, Second Edition, Macmillan, NY.
- Stallings, W. (1994), *Data and Computer Communications*, Fourth Edition, Macmillan, NY.

14

INTRODUCTION TO IP Version 6 (IPv6)

Unit Structure

14.0 OBJECTIVES:

- To briefly understand the major functions of addresses in the Internetworking arena
- To discusses the problems and the drawbacks of IPv4 addressing.
- It discusses how IPv6 can be used in the future of internetworking
- It reviews the areas that IPv6 technology covers the technologies and design approaches used when used with the latest technology.

14.0 INTRODUCTION

This chapter provides a brief introduction on the IP networking Protocol that uses the internet to connect to Learn the characteristics of IPv6 routing Learn the properties of IPSec security, It emphasises on the the various standards and optional headers supported by IPv6 packets.

14.2 AN OVERVIEW

14.2.1 IPv6 Introduction

Definition : What is IPv6?

Internet Protocol version 6 (IPv6) is a networking protocol that allows Windows users to communicate with other users over the Internet. It interacts with Windows naming services such as Domain Name System (DNS) and uses security technologies such as Internet Protocol security (IPSec), because they help facilitate the successful andsecure transfer of IP packets between computers.IPv6 supplants IPv4, pure IPv6 across the Internet will become more prevalent and will eventually replace IPv4.

Specifically, IPv6 contains addressing and control information to route packets for the next generation Internet.We believe that the expansion of the Internet is important and upgrades are sometimes warranted. Gathering information concerning every aspects of IPv6 we would hope to provide knowledge about this technology so everyone can benefit. It is therefore also called the *Next Generation Internet Protocol or IPng .*

Why we need IPv6 Addressing

IPv6 addresses the main problem of IPv4, that is, the exhaustion of addresses to connect computers or host in a packet-switched network. IPv6 has a very large address space and consists of 128 bits as compared to 32 bits in IPv4. The extended address length offered by IPv6 eliminates the need to use techniques such as network address translation to avoid running out of the available address space. IPv6 contains addressing and control information to route packets for the next generation Internet.

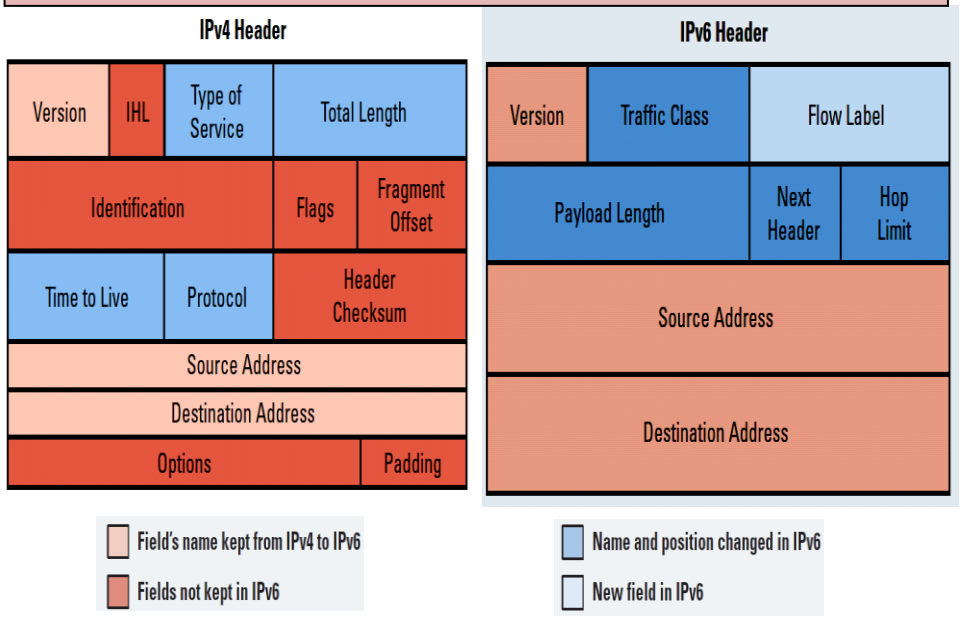
Therefore, it is now possible to support 2^128 unique IP addresses, a substantial increase in number of computers that can be addressed with the help of IPv6 addressing scheme.

14.2.2IPv6 Packet

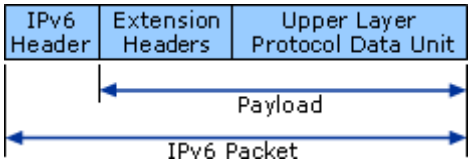
Like IPv4, IPv6 is a connectionless, unreliable datagram protocol that is primarily responsible for addressing and routing packets between hosts. An IPv6 packet consists of an IPv6 header and an IPv6 payload. The IPv6 payload consists of zero or more IPv6 extension headers and an upper layer protocol data unit, such as an ICMPv6 message, a TCP segment, or a UDP message. The following figure shows the structure of an IPv6 packet.

IPv4 Header vs. IPv6 Header

• The IPv4 header has 20 octets	• The IPv6 header has 40 octets
• Three of these fields are identical in	
• Other fields serve similar functions as in	
• The remaining IPv4 fields no longer exist	



Structure of an IPv6 Packet



Key Fields in the IPv6 Header

IP HEADER FIELD FUNCTION

- Source Address: Identifies the IPv6 address of the original source of the packet.

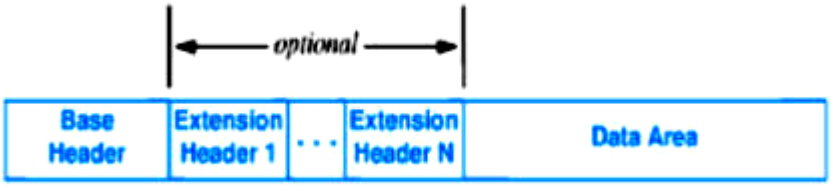
- Destination Address: Identifies the IPv6 address of the intermediate or final destination of the packet.

- Next Header: Identifies either the extension header immediately following the IPv6 header or an upper layer protocol, such as ICMPv6, TCP, or UDP.
- Hop Limit: Designates the number of subnets, or network segments, on which the packet is allowed to travel before being discarded by a router. The sending host sets the Hop

Limit, which prevents packets from endlessly circulating on an IPv6 internetwork. When forwarding an IPv6 packet, routers decrement the Hop Limit by one.

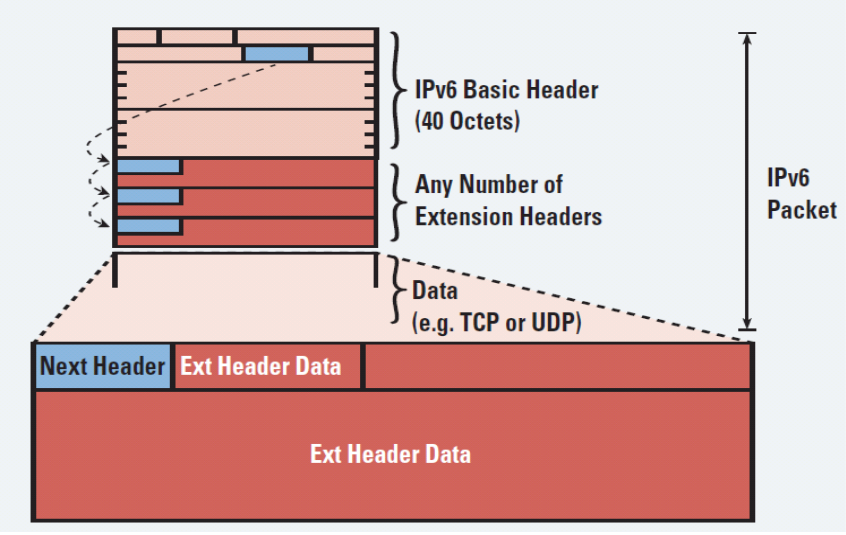
14.2.3 IPv6 Extension Headers

Zero or more extension headers of varying lengths can be present. If any extension headers are present, a Next Header field in the IPv6 header indicates the first extension header. In a typical IPv6 packet, no extension headers are present. The sending host adds one or more extension headers only if either an intermediate router or the destination requires special handling.



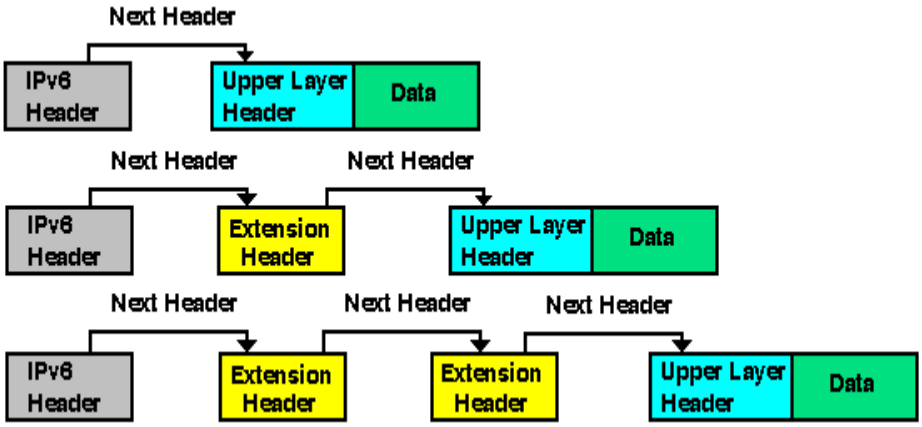
Within each extension header is another Next Header field that indicates the next extension header. The last extension header indicates the upper layer protocol (such as TCP, UDP, or ICMPv6) contained within the upper layer protocol data unit.

The IPv6 header and extension headers replace the existing IPv4 header with options. The new extension header format allows IPv6 to be augmented to support future needs and capabilities. Unlike options in the IPv4 header, IPv6 extension headers have no maximum size and can expand to accommodate all the extension data needed for IPv6 communication.



Following extension headers, which all IPv6 nodes must support:

- Hop-by-Hop Options header
- Destination Options header
- Routing header
- Fragment header
- Authentication header
- Encapsulating Security Payload header



14.2.4 Fragmentation in IPv6

If an IPv4 router receives a packet that is too large for the network segment to which the packet is being forwarded and fragmentation of the packet is allowed, IPv4 fragments the original packet into smaller packets that fit on the downstream network segment. In IPv6, only the sending host performs fragmentation. If an IPv6 router cannot forward a packet because it is too large, the router sends an ICMPv6 Packet Too Big message to the sending host and discards the packet.

Use of the Fragment extension header facilitates sending host fragmentation and destination host reassembly.

14.2.5 IPv6 Addressing

- **IPv6 Address Syntax**
- **Types of IPv6 Addresses**

The most obvious distinguishing feature of IPv6 is its use of much larger addresses. The size of an address in IPv6 is 128 bits, which is four times larger than an IPv4 address. A 32-bit address space includes 232 or 4,294,967,296 possible addresses. A 128-bit address space includes 2^{128} or 340,282, 366, 920, 938, 463, 463, 374, 607, 431, 768, 211, 456 (or 3.4×10^{38}) possible addresses.

It is even harder to conceive that the IPv6 address space will be consumed. A 128-bit address space provides 655, 570, 793, 348, 866, 943, 898, 599 (6.5×10^{23}) addresses for every square meter of the Earth's surface.

It is important to remember that the decision to make the IPv6 address 128 bits long was not so that every square meter of the Earth could have 6.5×10^{23} addresses. Rather, the relatively large size of the IPv6 address is designed to be subdivided into hierarchical routing domains that reflect the topology of the modern Internet. The use of 128 bits allows multiple levels of hierarchy and flexibility in designing hierarchical addressing and routing that is currently lacking in the IPv4-based Internet.

- **IPv6 Address Syntax**

IPv4 addresses are represented in dotted-decimal format. These 32-bit addresses are divided along 8-bit boundaries. Each set of 8 bits is converted to its decimal equivalent and separated from the other sets by periods. For IPv6, the 128-bit address is divided along 16-bit boundaries. Each 16-bit block is converted to a 4-digit hexadecimal number and separated by colons. The resulting representation is known as colon-hexadecimal.

The following is an IPv6 address in binary form:

```
00100001110110100000000011010011000000000000000001011
11100111011

00000010101010100000000011111111111111100010100010011
10001011010
```

The 128-bit address is divided along 16-bit boundaries:

```
0010000111011010  0000000011010011  0000000000000000
0010111100111011  0000001010101010  0000000011111111
1111111000101000  1001110001011010
```

Each 16-bit block is converted to hexadecimal and delimited with colons. The result is:

```
21DA:00D3:0000:2F3B:02AA:00FF:FE28:9C5A
```

IPv6 representation can be further simplified by removing the leading zeros within each 16-bit block. However, each block must have at least a single digit. With leading zero suppression, the address representation becomes:

```
21DA:D3:0:2F3B:2AA:FF:FE28:9C5A
```

- **Compressing Zeros**

Some types of addresses contain long sequences of zeros. To further simplify the representation of IPv6 addresses, a contiguous sequence of 16-bit blocks set to 0 in the colon-hexadecimal format can be compressed to "::," known as double-colon.

For example, the link-local address of FE80:0:0:0:2AA:FF:FE9A:4CA2 can be compressed to FE80::2AA:FF:FE9A:4CA2. The multicast address FF02:0:0:0:0:0:0:2 can be compressed to FF02::2.
Zero compression can be used to compress only a single contiguous series of 16-bit blocks expressed in colon-hexadecimal notation. You cannot use zero compression to include part of a 16-bit block. For example, you cannot express FF02:30:0:0:0:0:0:5 as FF02:3::5.

To determine how many 0 bits are represented by the double colon, you can count the number of blocks in the compressed address, subtract this number from 8, and then multiply the result by 16. For example, the address FF02::2 has two blocks (the "FF02" block and the "2" block.) The number of 0 bits expressed by the double colon is 96 (96 = (8 – 2)×16).

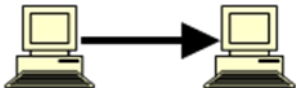
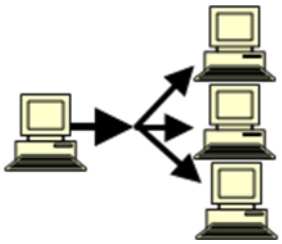
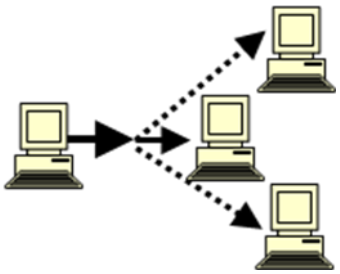
Zero compression can be used only once in a given address. Otherwise, you could not determine the number of 0 bits represented by each double colon.

- **Types of IPv6 Addresses**
  IPv6 supports three types of addresses:

- **Unicast**
  A unicast address identifies a single interface within the scope of the type of unicast address. With the appropriate unicast routing topology, packets addressed to a unicast address are delivered to a single interface. To accommodate load-balancing systems, allows multiple interfaces to use the same address as long as they appear as a single interface to the IPv6 implementation on the host.

- **Multicast**
  A multicast address identifies multiple interfaces. With the appropriate multicast routing topology, packets addressed to a multicast address are delivered to all interfaces that are identified by the address. A multicast address is used for one-to-many communication, with delivery to multiple interfaces.

- **Anycast**
  An anycast address identifies multiple interfaces. With the appropriate routing topology, packets addressed to an anycast address are delivered to a single interface, the nearest interface that is identified by the address. The nearest interface is defined as being closest in terms of routing distance. An anycast address is used for one-to-one-of-many communication, with delivery to a single interface.

ADDRESS TYPE	DESCRIPTION	TOPOLOGY
UNICAST	**"One to One"** • An address destined for a single interface. • A packet sent to a unicast address is delivered to the interface identified by that address.	
MULTICAST	**"One to Many"** • An address for a set of interfaces (typically belonging to different nodes). • A packet sent to a multicast address will be delivered to all interfaces identified by that address.	
ANYCAST	"***One to Nearest***" (Allocated from Unicast) • An address for a set of interfaces. • In most cases these interfaces belong to different nodes. • A packet sent to an anycast address is delivered to the	

	closest interface as determined by the IGP.	

14.3 SUMMARY

In this chapter we have examined the future of the Internet Protocol (IP) as embodied by IPv6, which is the next-generation IP. The areas in which IPv6 introduces the greatest changes are the areas of security, addressing, routing, and quality of service. Much of the detail in IPv6 differs from IPv4, and these changes were covered in detail in this chapter.

The chapter focused on the following topics:

- The issues associated with IPv4.
- The features of IPv6, including: larger address space, elimination of NAT and broadcast addresses, simplified header for improved router efficiency, support for mobility and security, and transition richness
- The features of IPv6 addresses, including: stateless autoconfiguration, prefix renumbering, multiple addresses per interface, link-local addresses, and the ability to use provider-dependent or provider-independent addressing.
    - The 40-octet IPv6 header, with its 8 fields plus extension headers to handle options
    - The 128-bit IPv6 addresses written in the format x:x:x:x:x:x:x:x
    - The IPv6 address interface ID
- The IPv6 address types including unicast (including global, link-local, and the deprecated site-local), multicast (for one-

to-many), and anycast (for one-to-nearest). There are no broadcast addresses.

- The ability to summarize IPv6 addresses, similar to IPv4 address summarization.

14.4 REVIEW QUESTIONS

1. Explain IP6 protocol?

2. Define IP6 **Packet Format ?**

3. IPv6 Addressing ?

14.5 REFERENCES

- Davies, J. *Understanding IPv6*. Redmond, WA: Microsoft Press, 2002.
- Huitema, C. *IPv6: The New Internet Protocol.* Second edition. Upper Saddle River, NJ: Prentice Hall, 1998.
- Miller, M. *Implementing IPv6: Supporting the Next Generation of Protocols.* Second edition. Foster City, CA: M&T Books, 2000.
- [RFC1884] R. Hinden and S. Deering, 1995, *IP version 6 addressing architecture*.
- [RFC1886] S. Thomson and C. Huitema, 1995, *DNS Extensions to support IP version 6*.
- [RFC2463] S. Thomson and T. Narten, 1998, *IPv6 Stateless Address Autoconfiguration*.
- [ngtrans-broker-05] Alain Durand, Paolo Fasano, and Domenico Lento, 2000, *IPv6 Tunnel broker, draft-ietf-ngtrans-broker-05*.

15

IP Version 6 (IPv6) CONFIGURATIONS AND TRANSISTIONS

Unit Structure

15.0 CHAPTER OBJECTIVES:

- List and describe the different types of IPv4 and IPv6 nodes.
- List and describe the types of tunneling configurations.
- Define the differences between configured and automatic tunneling.

15.1 CHAPTER INTRODUCTION

- The IPv6/IPv4 nodes with a dual stack or dual IP architecture, DNS infrastructure, and IPv6 over IPv4 tunneling are used to coexist with an IPv4 infrastructure and to provide eventual migration to anIPv6-only infrastructure. The chapter presents the IPv6 address configuration and verification commands. It also lists and describe the different types of IPv4 and IPv6 nodes, mechanisms for IPv4 to IPv6 transition, types of tunneling configurations.

15.2.1 Address Auto configuration

A highly useful aspect of IPv6 is its ability to automatically configure itself without the use of a stateful configuration protocol, such as Dynamic Host Configuration Protocol for IPv6 (DHCPv6). By default, an IPv6 host can configure a link-local address for each interface. By using router discovery, a host can also determine the addresses of routers, additional addresses, and other configuration parameters. The Router Advertisement message indicates whether a stateful address configuration protocol should be used.

Address autoconfiguration can be performed only on multicast-capable interfaces.

Autoconfigured Address States

Autoconfigured addresses are in one or more of the following states:

- **Tentative**
  The address is in the process of being verified as unique. Verification occurs through duplicate address detection.

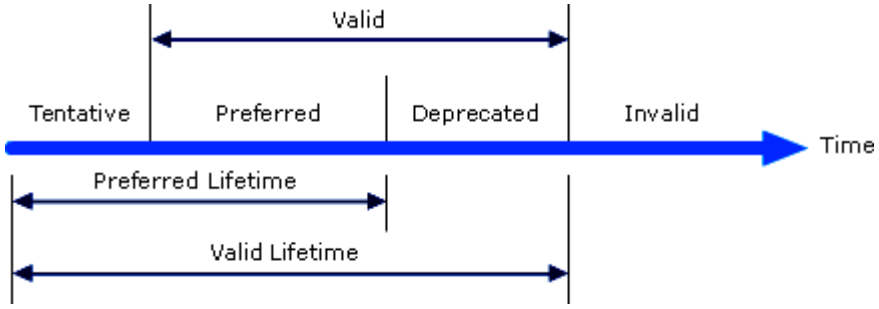
- **Preferred**
  An address for which uniqueness has been verified. A node can send and receive unicast traffic to and from a preferred address. Router Advertisement messages include the period of time that an address can remain in the tentative and preferred states.

- **Deprecated**
  An address that is still valid but whose use is discouraged for new communication. Existing communication sessions can continue to use a deprecated address. Nodes can send and receive unicast traffic to and from deprecated addresses.

- **Valid**
  An address from which unicast traffic can be sent and received. The valid state covers both the preferred and deprecated states. Router Advertisement messages include the amount of time that an address remains in the valid state. The valid lifetime must be longer than or equal to the preferred lifetime.

- **Invalid**
  An address for which a node can no longer send or receive unicast traffic. An address enters the invalid state after the valid lifetime expires.

The following figure shows the relationship between the states of an autoconfigured address and the preferred and valid lifetimes.



With the exception of link-local addresses, address autoconfiguration is specified only for hosts. Routers must obtain address and configuration parameters through another means (for example, manual configuration).

15.2.2 Types of Autoconfiguration

Autoconfiguration falls into three types:

1. **Stateless**
   Configuration is based on Router Advertisement messages. These messages include stateless address prefixes and require that hosts not use a stateful address configuration protocol.

2. **Stateful**
   Configuration is based on a stateful address configuration protocol, such as DHCPv6, to obtain addresses and other configuration options. Hosts use stateful address configuration when they receive Router Advertisement messages that do not include address prefixes and that require the hosts to use a stateful address configuration protocol. A host will also use a

stateful address configuration protocol when no routers are present on the local link.

3. **Both**
   Configuration is based on Router Advertisement messages. These messages include stateless address prefixes but require hosts to use a stateful address configuration protocol.

For all autoconfiguration types, a link-local address is always configured.

15.2.3 Autoconfiguration Process

Address autoconfiguration for an IPv6 node occurs as follows:

1. A tentative link-local address is derived, based on the link-local prefix of FE80::/64 and the 64-bit interface identifier.

2. Duplicate address detection is performed to verify the uniqueness of the tentative link-local address. If the address is already in use, the node must be configured manually.

3. If the address is not already in use, the tentative link-local address is assumed to be unique and valid. The link-local address is initialized for the interface. The corresponding solicited-node multicast link-layer address is registered with the network adapter.

4. The host sends a Router Solicitation message.

5. If the host receives no Router Advertisement messages, then it uses a stateful address configuration protocol to obtain addresses and other configuration parameters. Windows Server 2003 does not support the use of a stateful address configuration protocol for IPv6.

6. If the host receives a Router Advertisement message, the host is configured based on the information in the message

7. For each stateless address prefix that the message includes:
   A tentative address is derived from the address prefix and the appropriate 64-bit interface identifier.

8. The uniqueness of the tentative address is verified. If the tentative address is in use, the address is not initialized for the interface.

   If the tentative address is not in use, the address is initialized. Initialization includes setting the valid and preferred lifetimes based on information in the Router Advertisement message. Initialization also includes registering the corresponding solicited-node multicast link-layer address with the network adapter.

9. If specified in the Router Advertisement message, the host uses a stateful address configuration protocol to obtain additional addresses or configuration parameters.

15.2.4IPv6 Transition Technologies

1. **Node Types**
2. **Address Compatibility**

Protocol transitions are not easy, and the transition from IPv4 to IPv6 is no exception. Protocol transitions are typically deployed by installing and configuring the new protocol on all nodes within the network and verifying that all node and router operations work. Although this might be possible in a small- or medium-sized organization, the challenge of making a rapid protocol transition in a large organization is very difficult. Additionally, given the scope of the Internet, rapid protocol transition becomes an impossible task.

The transition from IPv4 to IPv6 will take years, and organizations or hosts within organizations might continue to use IPv4 indefinitely. Therefore, although migration is the long-term goal, equal consideration must be given to the interim coexistence of IPv4 and IPv6 nodes.

It defines the following transition criteria:

1. Existing IPv4 hosts can be upgraded at any time, independent of the upgrade of other hosts or routers.

2. Hosts that use only IPv6 can be added at any time, without dependencies on other hosts or routing infrastructure.

3. IPv4 hosts on which IPv6 is installed can continue to use their IPv4 addresses and do not need additional addresses.

4. Little preparation is required to either upgrade IPv4 nodes to IPv6 or deploy new IPv6 nodes.

The inherent lack of dependencies between IPv4 and IPv6 hosts, IPv4 routing infrastructure, and IPv6 routing infrastructure requires several mechanisms that allow seamless coexistence.

1. Node Types

Defines the following node types:

- **IPv4-only Node**

A node that implements only IPv4 (and has only IPv4 addresses). This node does not support IPv6. Most hosts and routers installed today are IPv4-only nodes.

- **IPv6-only Node**

A node that implements only IPv6 (and has only IPv6 addresses). This node is able to communicate only with IPv6 nodes and applications. This type of node is not common today, but it might become more prevalent as smaller devices such as cellular phones and handheld computing devices include the IPv6 protocol.

- **IPv6/IPv4 Node**

A node that implements both IPv4 and IPv6. This node is IPv6-enabled if it has an IPv6 interface configured.

For coexistence to occur, the largest number of nodes (IPv4 or IPv6 nodes) can communicate using an IPv4 infrastructure, an IPv6 infrastructure, or an infrastructure that is a combination of IPv4 and IPv6. True migration is achieved when all IPv4 nodes are converted to IPv6-only nodes. However, for the foreseeable future, practical migration is achieved when as many IPv4-only nodes as possible are converted to IPv6/IPv4 nodes. IPv4-only nodes can communicate with IPv6-only nodes only through an IPv4-to-IPv6 proxy or translation gateway.

2. Address Compatibility

The following addresses are defined to help IPv4 and IPv6 nodes coexist:

- **IPv4-compatible Addresses**

IPv6/IPv4 nodes that are communicating with IPv6 over an IPv4 infrastructure use IPv4-compatible addresses, 0:0:0:0:0:0:w.x.y.z or ::w.x.y.z (where w.x.y.z is the dotted decimal representation of a public IPv4 address). When an IPv4-compatible address is used as an IPv6 destination, IPv6 traffic is automatically encapsulated with IPv4 headers and sent to their destinations using the IPv4 infrastructure.

- **IPv4-mapped Addresses**

The IPv4-mapped address, 0:0:0:0:0:FFFF:w.x.y.z or ::FFFF:w.x.y.z, is used to represent an IPv4-only node to an IPv6 node. It is used only for internal representation. The IPv4-mapped address is never used as a source or destination address of an IPv6 packet. The IPv6 protocol for Windows Server 2003 does not support IPv4-mapped addresses. Some IPv6 implementations use IPv4-mapped addresses when translating traffic between IPv4-only and IPv6-only nodes.

- **6over4 Addresses**

Each 6over4 address comprises a valid 64-bit unicast address prefix and the interface identifier ::WWXX:YYZZ (where WWXX:YYZZ is the colon-hexadecimal representation of w.x.y.z, a unicast IPv4 address assigned to an interface). An example of a link-local 6over4 address based on the IPv4 address of 131.107.4.92 is FE80::836B:45C. 6over4 addresses represent a host that use the automatic tunneling mechanism.

- **6to4 Addresses**

6to4 addresses are based on the prefix 2002:WWXX:YYZZ::/48 (where WWXX:YYZZ is the colon-hexadecimal representation of w.x.y.z, a public IPv4 address assigned to an interface). 6to4 addresses represent sites that use the automatic tunneling mechanism

- **ISATAP Addresses**

Each Intra-site Automatic Tunnel Addressing Protocol (ISATAP) addresses comprise a valid 64-bit unicast address prefix and the interface identifier ::0:5EFE:w.x.y.z (where w.x.y.z is a unicast IPv4 address assigned to an interface). An example of a link-local ISATAP address is FE80::5EFE:131.107.4.92. ISATAP addresses represent hosts that use the automatic tunneling mechanism

- **Teredo Addresses**

Teredo addresses use the prefix 3FFE:831F::/32. An example of a Teredo address is 3FFE:831F:CE49:7601:8000:EFFF:62C3:FFFE. Beyond the first 32 bits, Teredo addresses encode the IPv4 address of a Teredo server, flags, and the encoded version of the external address and port of a Teredo client. Teredo addresses represent hosts that use the automatic tunnellingmechanism .

15.2.5 IPv6 - Auto Configuration vs DHCPv6

Introduction

A growing number of IPv6 experts are apprehensive about the adoption of the auto-configuration feature offered by IPv6 in contrast to the services offered by the existing DHCPv6 protocol in

the task of configuration of connected devices over an IP network. There are concerns over the potential disadvantages of auto-configuration in IPv6 such as its focus on configuration of IP address while overlooking the configuration of other parameters such as the DNS domain, DNS server, time servers, legacy WINS servers etc.

Using DHCP to supply this information and using IPv6 auto-configuration in its present form only for IP addressing does not make sense. The enterprises could as well use the DHCPv6 to configure the IP addresses too. Apart from the IP addresses, the additional information supplied by DHCPv6 offers the audit, tracking and management capabilities as required by the business enterprises. Despite its present shortcomings, IPv6 offers the most comprehensive long-term solution for the future networking requirements of the business enterprises. Every network administration policy maker across different business enterprises faces the dilemma of using IPv6 auto-configuration versus DHCPv6.

15.2.5IPv6Auto-Configuration

An important feature of IPv6 is that it allows plug and play option to the network devices by allowing them to configure themselves independently. It is possible to plug a node into an IPv6 network without requiring any human intervention. This feature was critical to allow network connectivity to an increasing number of mobile devices.

The proliferation of network enabled mobile devices has introduced the requirements of a mobile device to arbitrarily change locations on an IPv6 network while still maintaining its existing connections. To offer this functionality, a mobile device is assigned a home address where it remains always reachable. When the mobile device is at home, it connects to the home link and makes use of its home address. When the mobile device is away from home, a home agent (router) acts as a conduit and relays messages between the mobile device and other devices on the network to maintain the connection.

Types of auto-configuration:

- Stateful auto configuration and
- Stateless auto configuration.

IPv6 offers two types of auto-configuration: Stateful auto configuration and stateless auto configuration.

Stateful auto-configuration: This configuration requires some human intervention as it makes use of the Dynamic Host Configuration Protocol for IPv6 (DHCPv6) for installation and administration of nodes over a network. The DHCPv6 server maintains a list of nodes and the information about their state to know the availability of each IP address from the range specified by the network administrator.

Stateless auto-configuration: This type of configuration is suitable for small organizations and individuals. It allows each host to determine its address from the contents of received user advertisements

15.2.6 DHCPv6

The Dynamic Host Configuration Protocol (DHCP) facilitates the addition of new machines in a network. Around October 1993, DHCP began to take shape as a standard network protocol. The protocol allows the network devices to obtain the different parameters that are required by the clients to operate in an Internet Protocol (IP) network. The DHCP protocol significantly reduces the system administration workload as the network devices can be added to the network with little or no change in the device configuration.

DHCP also allows network parameter assignment at a single DHCP server or a group of such server located across the network. The dynamic host configuration is made possible with the automatic assignment of IP addresses, default gateway, subnet masks and other IP parameters. On connecting to a network, a DHCP configured node sends a broadcast query to the DHCP server requesting for necessary information. Upon receipt of a valid request, the DHCP server assigns an IP address from its pool of IP addresses and other TCP/IP configuration parameters such as the default gateway and subnet mask. The broadcast query is initiated just after booting and must be completed before the client initiates IP-based communication with other devices over the network.

DHCP allocates IP addresses to the network devices in three different modes: dynamic mode, automatic mode and manual mode. In the dynamic mode, the client is allotted an IP address for a specific period of time ranging from a few hours to a few months. At any time before the expiry of the lease, a DHCP client can request a renewal of the current IP address. Expiry of the lease

during a session leads to a dynamic renegotiation with the server for the original or a new IP address. In the automatic (also called as DHCP Reservation) mode, an IP address is chosen from the range defined by the network administrator and permanently assigned to the client. In the manual mode, the client manually selects the IP address and uses the DHCP protocol messages to inform the server of the choice of the IP address.

15.2.7 Summary of Benefits of IPv6 in a nutshell:

1) Increased address space

2) More efficient routing

3) Reduced management requirement

4) Improved methods to change ISP

5) Better mobility support

6) Multi-homing
7) Security
8) Scoped address: link-local, site-local and global-address space

15.3 REVIEW QUESTION

1) Explain IP6 auto configuration

2) Explain DHCP using IP6

15.4 SUMMARY

The chapter includes the following pieces of key information:

- IPv6 address configuration and verification commands
- The neighbor discovery or solicitation phase.
- Stateless autoconfiguration.
- The processes used to connect IPv6 devices on:
    - Broadcast multiaccess connections
    - Point-to-point connections
    - point-to-multipoint connections.

- A configured tunnel requires manual configuration of the tunnel endpoints.

- An automatic tunnel is a tunnel that does not require manual configuration. Tunnel endpoints are determined from routes and tunneling interfaces.

- ISATAP is an address assignment and host-to-host, host-to-router, and router-to-host automatic tunneling technology that provides unicast IPv6 connectivity between IPv6 hosts across an IPv4 intranet.

- ISATAP addresses are composed of a valid 64-bit unicast address prefix and the interface identifier::0:5EFE:*w.x.y.z* (*w.x.y.z* is a unicast IPv4 address assigned to an interface).
- Teredo is an address assignment and host-to-host or host-to-router automatic tunneling technology that provides unicast IPv6 connectivity across the IPv4 Internet when IPv6/IPv4 hosts are located behind one or multiple IPv4 NATs.

15.5 REFERENCES:

- IPv6 Main Page
www.cisco.com/go/ipv6

- IPv6 Headers At-a-Glance
http://www.cisco.com/en/US/technologies/tk648/tk872/technologies_white_paper0900aecd80260042.pdf

- The Cisco IOS Software Releases 12.4 Mainline Command References, available at:
http://www.cisco.com/en/US/products/ps6350/prod_command_reference_list.html

- The Cisco IOS IPv6 Command Reference, available at:
http://www.cisco.com/en/US/docs/ios/ipv6/command/reference/ipv6_book.html

- The Cisco IOS IPv6 Configuration Guide, Release 12.4, available at:
http://www.cisco.com/en/US/docs/ios/ipv6/configuration/guide/12_4/ipv6_12_4_book.html

- Cisco IOS IPv6 Multicast Introduction, available at:
http://www.cisco.com/en/US/tech/tk828/technologies_white_paper09186a0080203e90.shtml

F.Y.B.Sc. (IT) (Semester II)

Data Communication & Network Standards

Unit - I **Introduction to data communications and networking**

Introduction, Fundamental concepts, Data communications, Protocol, standards, standard organizations, signal propagation, analog and digital signals, bandwidth of signal and a medium, Fourier analysis and the concept of bandwidth of a signal. The data transmission rate and bandwidth.

Unit - II **Network Models**

Layered Tasks, The OSI reference model, Layers in the OSI reference model.

TCP/IP protocol suite, Addressing IPv4

Unit - III Information Encoding, Errors Detection and Correction

Introduction, representing different symbols, Minimizing errors, Multimedia, Multimedia and Data compression. Error classification, types of errors, redundancy, detection versus correction, hamming distance, cyclic redundancy check.

Unit - IV Media and Transmission modes

Data and signals, Periodic analog signals, Digital signals, Transmission impairment, Data rate limits, Performance, Digital to digital, Analog to digital conversion, Transmission modes, Digital to analog conversion, Analog to analog conversion, Guided media and Unguided media

Unit - V Network topologies, Switching and routing algorithms

Mesh, star, tree, ring, bus, hybrid, switching basics, circuit switching, packet switching and Message switching, routing algorithms

Unit - VI IP version 6

Overview, Terminology, IPv6 addresses, Special addresses, IPv6 header formats, IPv6 extension headers, IPv6 auto configuration via DHCP v6, IPv6 transition

Books :

Behrouz A Forouzan, “Data communications and Networking”, Fourth Edition, Mc-Graw Hill Achyut Godbole, “Data communications and Networks, TMH

Dr. Sidnie Feit, “TCP/IP”, Second Edition, TMH

Reference

W. Stallings, “Data and Computer Communications”, Eight Edition, Pearson Education.

Term Work and Tutorial

Should contain minimum 10 assignments and two class tests. (One case study in lieassignments)

Practical

None