DrillBit

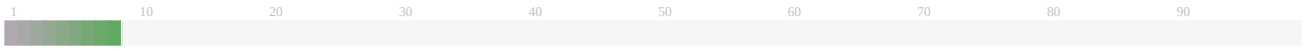
The Report is Generated by DrillBit Plagiarism Detection Software

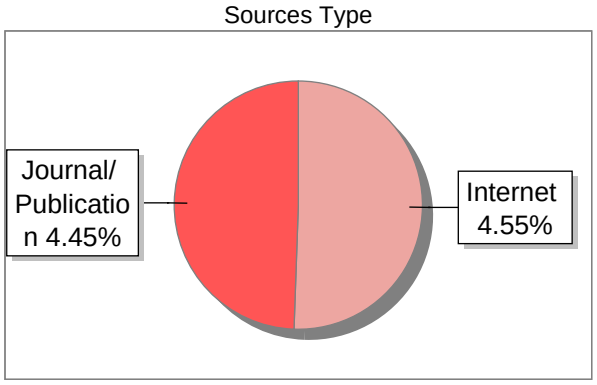
Submission Information

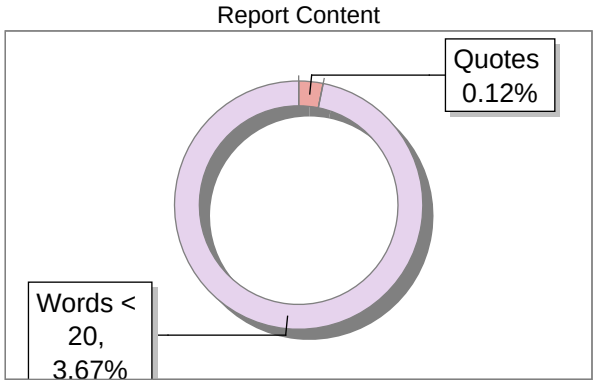
Author Name	AMAL CHAKRAVORTY
Title	DATA COMMUNICATION AND COMPUTER NETWRKS
Paper/Submission ID	3005486
Submitted by	librarian.adbu@gmail.com
Submission Date	2025-01-22 15:20:10
Total Pages, Total Words	124, 20403
Document type	Others

Result Information

Similarity **9 %**







Exclude Information

Quotes	Excluded
References/Bibliography	Excluded
Source: Excluded < 20 Words	Excluded
Excluded Source	**0 %**
Excluded Phrases	Not Excluded

Database Selection

Language	English
Student Papers	Yes
Journals & publishers	Yes
Internet or Web	Yes
Institution Repository	Yes

A Unique QR Code use to View/Download/Share Pdf File

DrillBit

DrillBit Similarity Report

9	15	A
SIMILARITY %	MATCHED SOURCES	GRADE

A-Satisfactory (0-10%)
B-Upgrade (11-40%)
C-Poor (41-60%)
D-Unacceptable (61-100%)

LOCATION	MATCHED DOMAIN	%	SOURCE TYPE
1	docplayer.net	3	Internet Data
2	itsmeebin.files.wordpress.com	3	Publication
3	vaibhav2501.files.wordpress.com	<1	Publication
4	dochero.tips	<1	Internet Data
5	dpvipracollege.in	<1	Publication
6	moam.info	<1	Internet Data
7	Cryptography for the Internet by Zimmermann-1998	<1	Publication
8	jatinderjyoti.in	<1	Publication
9	IEEE 2012 8th International Wireless Communications and Mobile Compu by	<1	Publication
10	network-insight.net	<1	Internet Data
11	ACM Press the 17th ACM conference- Chicago, Illinois, USA (2010.10, by Degabriele, Jean Pa- 2010	<1	Publication
12	moam.info	<1	Internet Data
13	pdfcookie.com	<1	Internet Data

14	digi-lib.stekom.ac.id	<1	Publication
15	index-of.es	<1	Publication

MODULE I- Digital Communications (Unit-1- Signal)

1.0.Introduction: The main function of the physical layer is to transmit data as electromagnetic signals over a communication medium. Typically, the data that is useful to a user or application is not directly suitable for network transmission. To send it, the data must first be transformed into electromagnetic signals. Both the data in its original form and the corresponding signals can be either analog or digital.

ANALOG AND DIGITAL: Data can be classified into two types: analog and digital. Analog data are continuous, meaning they can represent any value within a given range. In contrast, digital data consists of distinct states and takes on specific, individual values.

Analog and Digital Signal: Similar to the data they carry, signals can also be either analog or digital. Analog signals have the ability to take on an infinite number of values within a specified range, whereas digital signals are restricted to a fixed set of values.

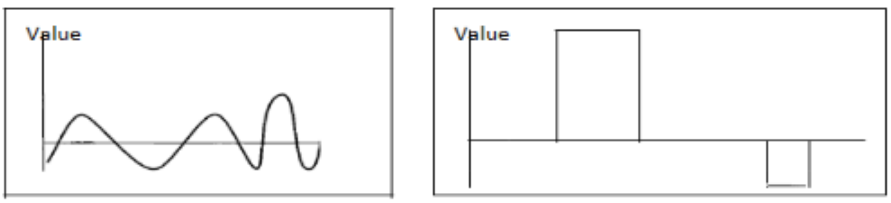


Figure 1.1 Comparison of analog and digital signals

1.2. Periodic and Non-periodic SignalsAnalog and digital signals can each be categorized as either periodic or non-

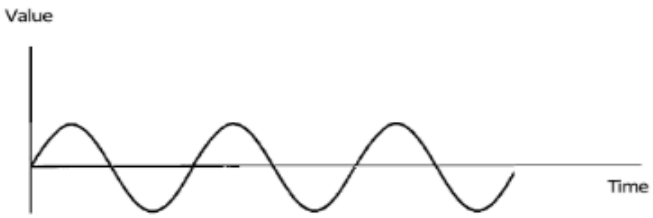


Figure 1.2: A sine wave

periodic. In data communication, periodic analog signals and non-periodic digital signals are commonly used. Periodic analog signals can be further divided into simple or composite types. A simple periodic analog signal, like a sine wave, cannot be broken down into smaller signals. In contrast, a composite periodic analog signal is made up of several sine waves. The sine wave is considered the basic form of a periodic analog signal.

A sine wave can be defined by three key parameters: peak amplitude, frequency, and phase. These parameters completely describe the sine wave.

Peak Amplitude: The peak amplitude is the maximum intensity of the signal, and it reflects the energy the signal carries. In electrical signals, it is usually measured in volts.

Period and Frequency: The period is the time, in seconds, required for a signal to complete one full cycle. Frequency, on the other hand, refers to how many cycles occur in one

second. It's important to note that period and frequency are two different ways to describe the same concept: the period is the inverse of the frequency, and the frequency is the inverse of period.

Phase: Phase refers to the position of the wave relative to time 0. When the wave is considered as being able to shift backward or forward along the time axis, the phase describes how much the wave has shifted. It shows the starting point of the first cycle and is measured in degrees or radians.

Wavelength: Wavelength is another characteristic of a signal as it travels through a transmission medium. It links the period or frequency of the sine wave to the speed at which the signal propagates through the medium.

Bandwidth: The bandwidth of a composite signal is defined as the range between its highest and lowest frequencies.

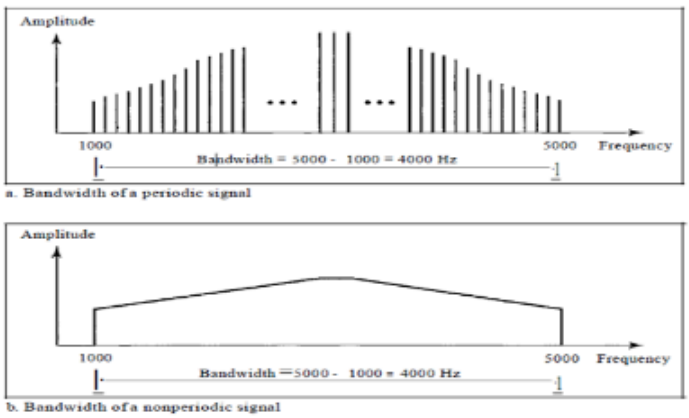


Fig 1.3: The bandwidth of periodic and non-periodic composite signals

Digital Signal: Information can be conveyed using both analog and digital signals. For instance, a 1 could be represented by a positive voltage, while a 0 could be represented by zero voltage. A digital signal may also have more than two levels, allowing multiple bits to be transmitted with each level. Figure 1.4 illustrates two signals: one with two levels and another with four.

Bit Rate: Many digital signals lack a consistent periodic pattern, which makes the notions of period and frequency unsuitable for their analysis. Instead, the bit rate is often employed to characterize digital signals. Bit rate refers to the amount of data transmitted per second and is generally measured in bits per second (bps).

1.3. Summary:

i) Analog Data: Continuous values, can take any value within a range.
ii) Digital Data: Discrete states have distinct values.
iii) Analog Signals: Infinite values within a range.
iv) Digital Signals: Limited number of values, typically binary (0 or 1).
v) Periodic Signals: Repeat at regular intervals.
vi) Non-periodic Signals: Do not repeat, random in nature.

1.4. Check your progress:

1. Which of the following is true about analog data?

a) Analog data is discrete.

b) Analog data takes continuous values.

c) Analog data cannot be transmitted over a network.

d) Analog data has only two possible values.

2. What is the term used to describe the number of periods a signal completes in one second?

a) Period

b) Amplitude

c) Frequency

d) Phase

4. Explain the difference between analog and digital data.
5. Describe the relationship between frequency and period.
6. What is the role of phase in an analog signal?
7. Discuss the concept of bit rate in digital communication.

MODULE I- Digital Communications
(Unit-2-Transmission Impairments)

2.0. Introduction: Transmission media are not perfect; leading to signal degradation, meaning the signal at the end of the medium differs from the original one. This results in the received signal not being identical to the transmitted one. Three major factors contributing to signal impairment are attenuation, distortion, and noise.

Attenuation: Attenuation refers to the loss of signal strength. As a signal passes through a transmission medium, some of its energy is lost because of the resistance of the medium. This is why a wire carrying electrical signals may become warm or even hot, as electrical energy is converted into heat. To counteract this loss, amplifiers are used to boost the signal. Engineers often use decibels (dB) to measure changes in signal strength. A negative dB value indicates attenuation, while a positive value shows

$$dB = 10 \log_{10}\left(\frac{P_2}{P_1}\right)$$

amplification.

The two variables P_1 and P_2 are the power of the signal at point P_1 and P_2 respectively.

Distortion: Distortion happens when the signal's original form is changed during transmission.

Noise: Noise is another element that degrades signal quality. Various types of noise, including thermal noise, induced noise, crosstalk, and impulse noise, can disrupt the signal. Thermal noise is caused by the random motion of electrons in a wire, which generates additional signals. Induced noise originates from external sources, such as motors or electrical appliances, which act as transmitters while the transmission medium acts as a receiver. Crosstalk occurs when signals from one wire interfere with those in a nearby wire.

2.1. Signal-to-Noise Ratio : It is the ratio of signal power to noise power, as shown below:

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

Since SNR is the ratio of two powers, it is commonly expressed in decibel units and defined as:

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

2.2. Nyquist Bit Rate: The Nyquist bit rate formula establishes the theoretical upper limit for the bit rate in a noiseless channel.

$$\text{Bit Rate} = 2 \times \text{bandwidth} \times \log_2 L$$

In the Nyquist bit rate formula, bandwidth refers to the bandwidth of the channel, L represents the number of signal levels used to encode data, and bit rate is the bit rate in bits per second.

2.3. Shannon Capacity: In reality, a completely noiseless channel does not exist, as all channels are affected by noise. In 1944, Claude Shannon presented a formula called the Shannon capacity, which calculates the theoretical maximum data rate for a channel with noise.

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

2.4. Summary:

i) $$SNR = \frac{average\ signal\ power}{average\ noise\ power}$$

ii) $\text{SNR}_{db} = 10 \log_{10} \text{SNR}$

iii) Nyquist Bit Rate = 2 X bandwidth x $\log_2 L$

iv) Shannon Capacity = bandwidth X $\log_2(1 + \text{SNR})$

2.5 Check your progress:

1. What happens to a signal when it undergoes attenuation?

a) It gains energy.

b) It loses energy.

c) Its form changes.

d) It becomes distorted.

2. Which of the following types of noise is caused by random motion of electrons in a wire?

a) Thermal noise

b) Induced noise

c) Crosstalk

d) Impulse noise

3. Fill in the blank:

The Shannon Capacity formula determines the theoretical highest data rate for a _____ channel.

4. How is signal-to-noise ratio (SNR) used in evaluating signal quality?

5. What are the differences between attenuation and distortion in signal transmission?

6. Discuss how noise affects the performance of a communication channel and how it is accounted for in the Shannon Capacity formula

MODULE I- Digital Communications
(Unit-3- Analog to Analog Conversion and Bandwidth Utilization)

3.0. Introduction: The International Organization for Standardization (ISO) introduced a model called the Open Systems Interconnection (OSI) model, which facilitates communication between two systems. The term 'open' refers to being independent and autonomous. The OSI model comprises seven distinct yet interconnected layers, each outlining a specific part of the process for transmitting data across a network. It is important to note that the OSI model is a conceptual framework, not a protocol. The seven layers are depicted in the figure below (Figure 3.0).

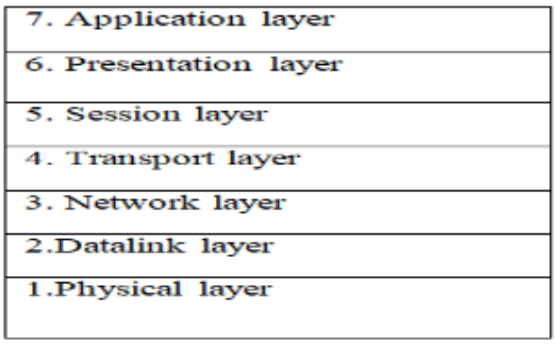


Figure 3.0 : OSI Model

The duties of each layer will be discussed in the subsequent units. In this unit, we will discuss about

one prime duty of Application layer, which is signal transmission.

The analog signal is a continuous signal in which the time-varying characteristic of the signal represents another time-varying quantity, meaning it is analogous to another time-varying signal. Analog-to-analog conversion, or modulation, involves representing analog information using an analog signal. It is a process by virtue of which a characteristic of carrier wave is varied in response to the instantaneous amplitude of the modulating signal. This modulation is generally needed when a bandpass channel is required. Bandpass is a range of frequencies which are transmitted through a bandpass filter which is a filter allowing specific frequencies to pass preventing signals at unwanted frequencies.

Analog to Analog conversion is done in three ways:

1. Amplitude Modulation
2. Frequency Modulation
3. Phase Modulation

3.1. Amplitude Modulation: Amplitude modulation is a method in which the carrier wave's amplitude is adjusted based on the instantaneous amplitude of the modulating signal, while its frequency and phase stay constant. The following diagram illustrates how amplitude modulation works:

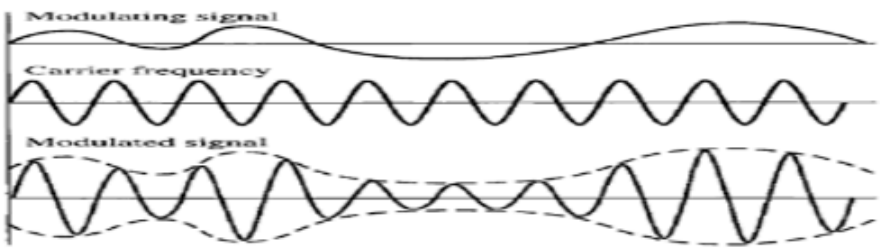


Figure 3.1: Amplitude Modulation

3.2. Frequency Modulation: In this type of modulation the frequency of the carrier wave is altered based on the instantaneous frequency of the modulating signal, while keeping the phase and amplitude constant. The figure 3.2 illustrates the concept of frequency modulation:

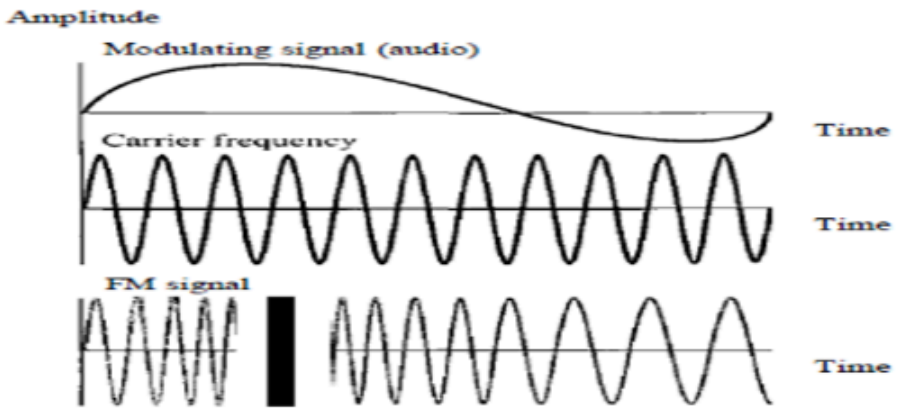


Figure 3.2: Frequency modulation

3.3. Phase Modulation: In phase modulation (PM), the phase of the carrier signal is altered to match the changing voltage levels (amplitude) of the modulating signal. Although the carrier's peak amplitude and frequency stay the same, its phase shifts according to the changes in the

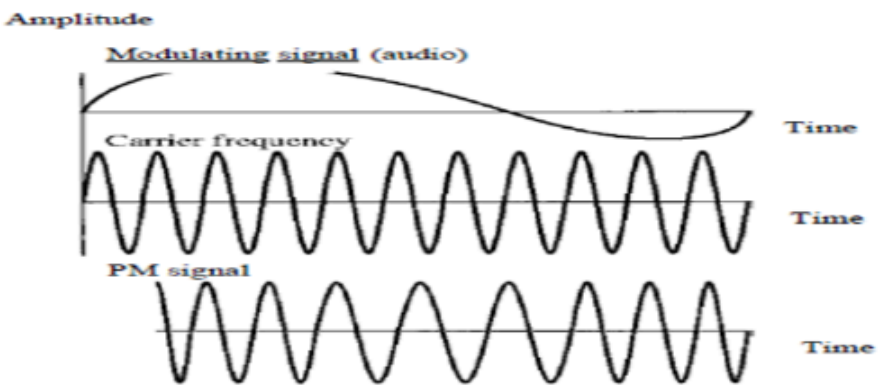


Figure 3.3: Phase Modulation

amplitude of the information signal.

3.4. **Bandwidth Utilization**: In practical scenarios, communication links are constrained by limited bandwidths. Efficiently using these bandwidths has always been a key challenge in electronic communications and will continue to be so. However, the definition of efficient use can vary based on the specific application. In some cases, it may be necessary to combine multiple low-bandwidth channels to form a single higher-bandwidth channel. In other instances, expanding a channel's bandwidth is essential for objectives such as ensuring privacy or increasing resistance to jamming. One of the main strategies to optimize bandwidth utilization is multiplexing.

Multiplexing: Three types of multiplexing techniques ARE normally applied : frequency-division multiplexing, wavelength-division multiplexing, and time division multiplexing. The first two are used for analog signals, while the third is designed for digital signals.

Frequency-Division Multiplexing: Frequency-division multiplexing (FDM) is an analog method used when a communication link's bandwidth (measured in hertz) is larger than the combined bandwidth of the signals that need to be sent. In FDM, each transmitting device creates signals that modulate different carrier frequencies.These modulated signals are combined into a single composite signal for transmission over the link. The carrier frequencies are spaced apart enough to allow room for each modulated signal. Guard bands, which are sections of unused bandwidth, are inserted between the channels to avoid signal overlap. Moreover, the carrier frequencies must not

interfere with the data signals themselves. Figure 3.4 provides a conceptual view of FDM.

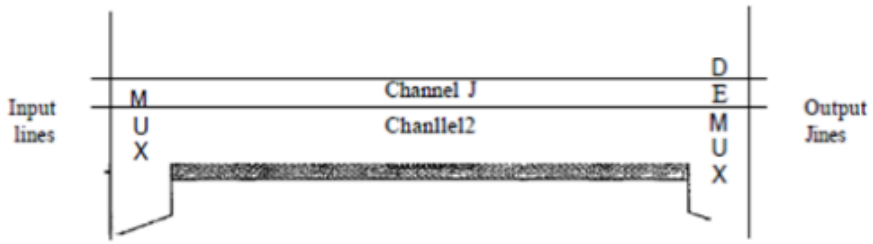


Figure 3.4: Frequency-division multiplexing

Wavelength-Division Multiplexing: Wavelength-division multiplexing (WDM) is a technique aimed at leveraging the high data transmission rates offered by fiber-optic cables. Since optical fiber supports higher data rates compared to metal transmission lines, using a fiber-optic cable for only a single data stream leads to inefficient use of its bandwidth.

Multiplexing allows multiple data streams to be combined, maximizing the full capacity of the fiber-optic cable.

Synchronous Time-Division Multiplexing: **Time-Division Multiplexing (TDM)** is a digital method that enables multiple connections to share the bandwidth of a single communication link. Unlike Frequency-Division Multiplexing (FDM), which divides the bandwidth, TDM divides the time. Each connection is assigned a specific time slot to transmit its data.

Figure 3.5 illustrates the concept of TDM. Although the same link is used as in FDM, TDM operates by allocating the link based on time rather than frequency. In the figure,

segments of signals 1, 2, 3, and 4 occupy the link one after another in a time-sequenced manner.

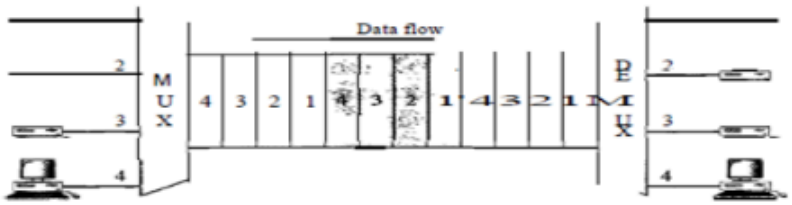


Figure 3.5: Time Division Multiplexing

Note that in Figure 3.5, we are focused solely on multiplexing, not switching. This means that all the data in a message from source 1 always go to a specific destination, whether it is 1, 2, 3, or 4. The delivery is fixed and unchanging, unlike in switching. It's also important to remember that TDM is fundamentally a digital multiplexing technique. Digital data from different sources are combined into one time-shared link. However, this does not imply that the sources must only produce digital data; analog data can be sampled, converted to digital form, and then multiplexed using TDM.

3.5. Summary:

) TDM is a digital multiplexing method that merges multiple low-speed channels into a single high-speed channel.

ii) WDM is an analog multiplexing approach for combining optical-signals.

iii) FDM is an analog technique that combines multiple analog signals.

iv) In amplitude modulation, the phase and frequency stay unchanged.

v) In frequency modulation, the amplitude and phase remain constant.
vi) In phase modulation, the frequency and amplitude do not change.

3.6 Check your progress:

1. Fill in the blank:

   Amplitude Modulation varies the _____ of the carrier wave according to the modulating signal.

2. Fill in the blank:

   _____ is an analog technique used to combine multiple signals for transmission over a single communication link.

3. Explain the process of Amplitude Modulation (AM) and its effect on the carrier wave.
4. What is the role of guard bands in Frequency-Division Multiplexing (FDM), and why are they important?
5. Compare the advantages and disadvantages of Frequency-Division Multiplexing (FDM) and Time-Division Multiplexing (TDM).
6. Explain how Phase Modulation (PM) works and its typical application.

4.0. Introduction: In telecommunications, transmission media are categorized into two primary types: guided and unguided.Guided media includes twisted-pair cables, coaxial cables, and fiber-optic cables. On the other hand, unguided media refers to transmission through free space.

4.1. Twisted pair-cable: A twisted pair-cable consists of two conductors, typically made of copper, each encased in its own layer of plastic insulation, and twisted together, as shown in Figure 4.1.

Figure 4.1: Twisted pair-cable

In a twisted pair, one wire carries the signal to the receiver, while the other serves as a ground reference. The receiver determines the signal by comparing the difference between the two wires. Along with the intended signal, both wires can be influenced by interference (noise) and crosstalk, which can introduce unwanted signals. If the wires are arranged parallel to each other, the impact of noise or crosstalk is not uniform because their positions relative to the interference sources differ—one might be closer while the other is farther. This results in a distortion at the receiver. Twisting the wires together helps preserve balance by alternating the relative positions of the wires. For instance, in one twist, one wire could be closer to the noise source, and in the next twist, the positions are reversed. This design increases the

likelihood that both wires experience the external interference in a similar way, thus reducing unwanted signals. As a result, the receiver, which measures the voltage difference between the two wires, receives minimal noise, as much of it cancels out. Therefore, the quality of the cable is influenced by the number of twists per unit length (e.g., inch).

4.2. Coaxial Cable: Coaxial cable (commonly referred to as coax) is designed to transmit signals at higher frequencies than twisted pair cables, largely due to its unique construction. Unlike twisted pair cables, coaxial cables feature a central core conductor made of either solid or stranded wire, typically copper. Surrounding this core is an insulating layer, followed by an outer conductor made of metal foil, braid, or a combination of both. The outer metallic layer serves a dual purpose: it acts as a shield to protect the signal from noise and also functions as the second conductor that completes the circuit.

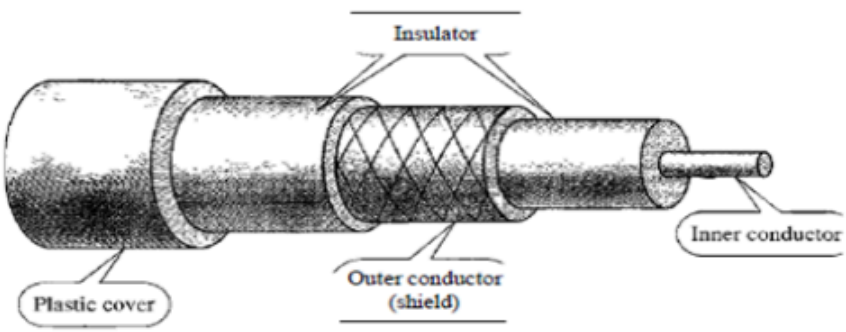


Figure 4.2: Twisted Pair Cable

The outer conductor is additionally enclosed in an insulating layer, and the entire cable is safeguarded by a plastic outer cover (as shown in Figure 4.2).

4.3. Fiber-Optic Cable: A fiber-optic cable is constructed from glass or plastic and transmits data using light signals. To grasp how optical fiber works, it's important to first understand some properties of light. Light travels in a straight line as long as it remains within a uniform medium.

When a light ray moves from one substance to another with a different density, it changes direction. If the angle at which the light hits the interface (called the angle of incidence) is less than the critical angle, the light bends towards the surface. If the angle of incidence equals the critical angle, the light travels along the interface. If the angle exceeds the critical angle, the light reflects and stays within the denser material. The critical angle is a characteristic of the substance and varies depending on the materials involved.

Optical fibers utilize this principle of reflection to guide light through the fiber. The central core, made from glass or plastic, is surrounded by a layer of cladding that is made from a less dense material, also typically glass or plastic. The difference in density between the core and the cladding causes light traveling within the core to reflect off the cladding instead of passing into it.

4.4 UNGUIDED MEDIA: WIRELESS: Unguided media transmit electromagnetic waves without requiring a physical conductor. This type of transmission is often referred to as wireless communication. The signals are generally broadcast through free space, which allows any device with the proper receiver to access them. Figure 4.3 illustrates the part of the electromagnetic spectrum, from 3 kHz to 900 THz, that is used for wireless communication.

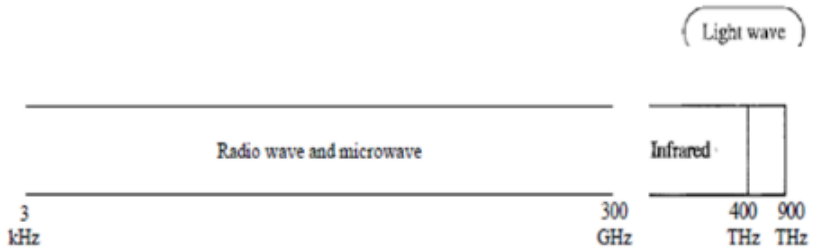


Fig.4.3: Electromagnetic spectrum for wireless communication

Unguided signals can reach their destination using several methods: ground propagation, sky propagation, and line-of-sight propagation.

In ground propagation, radio waves move through the lower atmosphere, staying close to the Earth's surface. These low-frequency signals radiate in all directions from the transmitting antenna and follow the curvature of the Earth. The distance they can travel depends on the signal's strength—higher power allows for greater distances.

In sky propagation, higher-frequency radio waves are directed upward into the ionosphere, a layer of the atmosphere where particles are ionized, and then reflected back to Earth. This method enables signals to travel longer distances with lower transmission power.

In line-of-sight propagation, high-frequency signals move directly in a straight path from one antenna to another, without any deviation.

4.5. Summary:

ii) A guided medium provides a physical pathway for signals to travel between devices. The most common types of guided media include twisted pair cables, coaxial cables, and optical fibers.

ii) Twisted-pair cable is made of two copper wires, each insulated, and twisted together. This type of cable is used for both voice and data transmission.

iii) Coaxial cable features a central conductor and a surrounding shield. It is capable of carrying higher-frequency signals compared to twisted-pair cables. Coaxial cables are commonly used in cable television networks and traditional Ethernet local area networks (LANs).

iv) Fiber-optic cables are made up of a central core of glass or plastic, surrounded by cladding, and covered by an outer protective jacket. These cables transmit data in the form of light signals.

4.6 Check your progress:

1. What role does the second wire in a twisted pair cable serve?
2. What are the primary components of a coaxial cable?
3. How does the outer metallic conductor in a coaxial cable protect against noise?
4. How does light travel through fiber-optic cables?
5. What is the difference between guided media and unguided media in telecommunications?
6. What are the three types of propagation used in unguided media?

MODULE I- Digital Communications
(Unit-5- Ethernet, Fast Ethernet, and Introduction to Gigabit Ethernet)

5.0 Introduction:

Ethernet is a popular technology for Local Area Networks (LANs) that follows the IEEE 802.3 standard. It enables devices to communicate by transmitting data as frames over a physical medium.

5.1 Ethernet Standards

Ethernet has evolved over time to increase speed and improve performance. Here are the main standards:

1. 10BASE-T (Original Ethernet)
    - Speed: 10 Mbps.

    - Cable: Unshielded twisted pair (UTP) cables.
    - Distance: Up to 100 meters.
2. 100BASE-TX (Fast Ethernet)
    - Speed: 100 Mbps.
    - Cable: Cat 5 or higher.
    - Distance: Up to 100 meters.
3. 1000BASE-T (Gigabit Ethernet)
    - Speed: 1 Gbps (1000 Mbps).
    - Cable: Cat 5e or higher.
    - Distance: Up to 100 meters.

5.2 Fast Ethernet (100BASE-TX)

Fast Ethernet is an upgrade over the original 10BASE-T standard, providing 10 times the speed (100 Mbps) with improved capabilities.

Key Features of Fast Ethernet:

- Speed: 100 Mbps, significantly faster than the original Ethernet.
- Transmission Medium: Uses Cat 5 or higher twisted-pair cables.
- Duplex Modes: Supports half-duplex (send or receive data at a time) and full-duplex (send and receive data simultaneously).
- Backward Compatibility: Fully compatible with 10BASE-T, making it easy to upgrade existing networks.

Fast Ethernet Variants:

- 100BASE-TX: Most common, used with twisted-pair cables.
- 100BASE-FX: Uses fiber-optic cables for higher distance and better noise immunity.

5.3 Introduction to Gigabit Ethernet (1000BASE-T)

Gigabit Ethernet brings Ethernet into the era of high-speed data transfer, offering speeds up to 1 Gbps (1000 Mbps), which is 10 times faster than Fast Ethernet.

Key Features of Gigabit Ethernet:

- Speed: 1 Gbps (1000 Mbps), ideal for modern high-bandwidth applications.
- Transmission Medium: Can run over Cat 5e or Cat 6 twisted-pair cables, or fiber optics.
- Duplex Mode: Full-duplex is standard, supporting simultaneous sending and receiving of data.
- Distance: Can maintain 1 Gbps speed for up to 100 meters.
- Compatibility: Backward compatible with 10BASE-T and 100BASE-TX, ensuring smooth upgrades.

Gigabit Ethernet Variants:

- 1000BASE-T: Most common, uses Cat 5e or higher cables.
- 1000BASE-LX: Uses single-mode fiber-optic cables for longer distance.
- 1000BASE-SX: Uses multi-mode fiber-optic cables for short-range applications.

Advantages of Gigabit Ethernet:

- Higher Speed: A significant performance boost, ideal for demanding applications.
- Scalability: Handles increasing network demands effectively.
- Reduced Latency: Faster speeds result in quicker data transfer, beneficial for real-time applications like video conferencing and VoIP.

5.4 Comparison of Ethernet, Fast Ethernet, and Gigabit Ethernet

Feature	Ethernet (10BASE-T)	Fast Ethernet (100BASE-TX)	Gigabit Ethernet (1000BASE-T)
Speed	10 Mbps	100 Mbps	1000 Mbps (1 Gbps)
Cable Type	Cat 5 or higher	Cat 5 or higher	Cat 5e, Cat 6, or higher
Max Distance	100 meters	100 meters	100 meters (Cat 5e, Cat 6)
Duplex Mode	Half-duplex	Half or Full-duplex	Full-duplex
Compatibility	Limited to 10BASE-T	Backward-compatible with 10BASE-T	Backward-compatible with 10BASE-T and 100BASE-TX
Applications	Basic networking needs	Small to medium-sized networks	High-performance networks, data centers, multimedia

---

5.5 Future Trends in Ethernet Technology

- 10 Gigabit Ethernet (10GbE): Provides speeds of 10 Gbps, generally used in data centers and high-demand networks.
- 40/100 Gigabit Ethernet: For extremely high-speed networks, especially in cloud computing and data centers.
- Power over Ethernet (PoE): Delivers both power and data over the same Ethernet cable, useful for devices like IP cameras, phones, and wireless access points.
- Higher Speeds over Copper: Future developments aim to push Ethernet speeds over copper cables even further (e.g., 10GBASE-T over Cat 6a cables).

5.6.Summary:

1. Ethernet is the most widely used LAN technology based on the IEEE 802.3 standard, facilitating frame-based communication.
2. Ethernet Standards have evolved:
    - 10BASE-T: 10 Mbps speed using Cat 5 cables.
    - 100BASE-TX (Fast Ethernet): 100 Mbps, backward-compatible with 10BASE-T.
    - 1000BASE-T (Gigabit Ethernet): 1 Gbps, uses Cat 5e or Cat 6 cables.
3. Fast Ethernet (100BASE-TX) offers speeds of 100 Mbps and supports both half-duplex and full-duplex transmission modes.
4. Gigabit Ethernet (1000BASE-T) provides 1 Gbps speed, ideal for high-performance applications, supporting full-duplex communication.

5. Comparing Ethernet technologies: Ethernet (10BASE-T) offers 10 Mbps, Fast Ethernet (100BASE-TX) offers 100 Mbps, and Gigabit Ethernet (1000BASE-T) offers 1 Gbps.
6. Future Trends include faster versions of Ethernet (10GbE, 40/100 GbE), PoE (Power over Ethernet), and improvements in Ethernet speed over copper cables.

5.7 Check your progress:

1. What is the IEEE standard for Ethernet in Local Area Networks (LANs)?

1. What is the maximum speed of the original Ethernet (10BASE-T) standard?
2. Which type of cable is used in the 10BASE-T Ethernet standard?
3. What is the maximum distance that Ethernet signals can travel using 10BASE-T?
4. How does Fast Ethernet (100BASE-TX) differ from the original Ethernet (10BASE-T) in terms of speed?
5. What type of cable is required for Fast Ethernet (100BASE-TX)?
6. What is the maximum distance supported by 100BASE-TX Ethernet?

MODULE I- Digital Communications
(Unit-6- Repeater, Hubs, Bridges, Switches, Router and Gateway.)

6.0 Introduction:

In a network, several devices are used to facilitate communication, manage traffic, and extend the reach of the network. These devices operate at different layers of the

OSI model to ensure data transmission is efficient, secure, and reliable.

Key Network Devices:

- Repeater
- Hub
- Bridge
- Switch
- Router
- Gateway

Each device has a specific function, ranging from signal amplification to traffic management and routing between different networks.

6.1 Repeater

A repeater is a basic network device used to extend the range of signals across long distances, particularly in wired or wireless networks.

Functions of a Repeater:

- Signal Regeneration: Amplifies and regenerates weak or distorted signals to restore them for further transmission.
- Extend Distance: Helps overcome distance limitations of transmission media like Ethernet or Wi-Fi, by retransmitting the signal over long distances.
- OSI Layer: Operates at the Physical Layer (Layer 1).

- Limitations: A repeater does not filter data, nor does it segment traffic; it simply amplifies and repeats all signals.

Applications:

- Used in large networks, extending the range of signals over long cables or wireless links.
- Common in long-distance communications, like telephone or fiber-optic networks.

6.2 Hub

A hub is a basic network device used to connect multiple devices in a network. Hub operates at the Physical Layer of the OSI model.

Functions of a Hub:

- Data Broadcasting: A hub receives data from a device and broadcasts it to all other connected devices.
- Shared Bandwidth: All devices connected to a hub share the available bandwidth.
- OSI Layer: Operates at the Physical Layer (Layer 1).
- Limitations: Does not filter traffic, does not provide security, and does not reduce network congestion efficiently.

Types of Hubs:

- Passive Hub: Simply connects devices without amplifying the signal.
- Active Hub: Amplifies and retransmits signals to extend the range of transmission.
- Intelligent Hub: Offers basic management features like monitoring traffic.

Applications:

- Hubs are mostly used in small, simple networks with limited devices where high-level performance and security are not required.

6.3 Bridge

A bridge is a network device that provides connectivity inbetween two or more network segments and filters traffic based on MAC addresses. It helps segment traffic to improve network performance.

Functions of a Bridge:

- Traffic Filtering: Bridges filter data packets based on MAC addresses, forwarding only relevant packets to other segments.
- Collision Domain Segmentation: Reduces the collision domain by dividing large networks into smaller segments.
- OSI Layer: Operates at the Data Link Layer (Layer 2).

- Learning Function: Bridges maintain a MAC address table to learn which devices are on which segments.

Applications:

- Used to connect segments of a LAN, improving traffic management and reducing network congestion.
- Used in older networks but being largely replaced by switches in modern environments.

6.4 Switch

A switch is a more advanced network device that operates at the Data Link Layer (Layer 2) and sometimes at the Network Layer (Layer 3).

Functions of a Switch:

- Traffic Management: Switches forward data based on MAC addresses, ensuring data reaches the correct device on the network.
- Collision Domain Segmentation: Unlike hubs, switches create a separate collision domain for each port, allowing for more efficient communication.
- Full-Duplex Communication: Supports simultaneous sending and receiving of data, unlike hubs.

- Learning and Forwarding: Switches dynamically learn the MAC addresses of devices and use this information to forward data only to the relevant port.

Types of Switches:

- Unmanaged Switch: Basic switches with no configuration options, plug-and-play.
- Managed Switch: Offers advanced features like VLANs, QoS (Quality of Service), network monitoring, and remote management.

Applications:

- Common in modern LANs for efficient data transfer, supporting multiple devices while maintaining network performance.

6.5 Router

A router is a network device that connects different networks and routes data between them based on IP addresses.

Functions of a Router:

- Routing Data: Routers examine the destination IP address of packets and route them to the appropriate network or sub-network.

- Inter-Network Communication: Connects different types of networks (e.g., LAN to WAN or multiple LANs).
- Network Address Translation (NAT): Translates private IP addresses to public ones for internet communication and vice versa.
- OSI Layer: Operates at the Network Layer (Layer 3).

Types of Routers:

- Static Router: Routes data based on manually configured routes.
- Dynamic Router: Uses routing protocols like RIP, OSPF, or BGP to determine the best route for data.

Applications:

- Used to connect local networks to the internet or connect multiple network segments.
- Routers are critical in wide-area networks (WANs) and the internet.

6.6 Gateway

A gateway is a device that connects two different networks and translates data between different protocols or architectures.

Functions of a Gateway:

- Protocol Translation: Acts as a translator between different network protocols, ensuring that communication is possible between incompatible networks.

- Network Security: Often includes features like firewall capabilities, traffic filtering, and monitoring.
- OSI Layer: Operates at the Application Layer (Layer 7), but can function at lower layers depending on the type of gateway.

Types of Gateways:

- Protocol Gateway: Converts one protocol to another, like HTTP to FTP.
- Network Gateway: Used for interconnecting different types of networks, such as connecting a LAN to the internet.

Applications:

- Common in enterprises where different types of networks (e.g., LAN and WAN) need to communicate, or where different protocol systems need to interact.

6.7 Summary Points:

1. Repeater:
    I. Amplifies and regenerates signals to extend transmission range.
    II. Operates at the Physical Layer (Layer 1).
    III. Does not filter or segment traffic.
2. Hub:
    I. Connects multiple devices, broadcasts data to all ports.
    II. Operates at the Physical Layer.

III. Does not segment traffic or provide advanced features like security.

3. Bridge:
   I. Connects network segments and filters traffic using MAC addresses.
   II. Operates at the Data Link Layer (Layer 2).
   III. Reduces network collisions by segmenting traffic.
4. Switch:
   I. More advanced than hubs, forwards data based on MAC addresses.
   II. Operates at the Data Link Layer (Layer 2) and sometimes Network Layer (Layer 3).
   III. Segments collision domains and supports full-duplex communication.
5. Router:
   I. Routes data between different networks based on IP addresses.
   II. Operates at the Network Layer (Layer 3).
   III. Connects different networks and manages IP traffic.
6. Gateway:
   I. Acts as a protocol translator between different network architectures and protocols.
   II. Operates primarily at the Application Layer (Layer 7).
   III. Connects incompatible networks and provides network security.

6.8 Check your progress:

1. What is the primary function of a repeater in a network?

2. At which layer of the OSI model does a repeater operate?

3. What is a key limitation of a repeater in a network?
4. How does a hub differ from a repeater in terms of data transmission?
5. What is the difference between a passive hub, an active hub, and an intelligent hub?
6. How does a bridge help improve network performance?

Module II: Media Access Control and Data Link Layer
(Unit-7- Data Link Layer Fundamentals.)

7.0 Introduction: Data link layer converts the physical layer, which is a basic transmission medium, into a link responsible for node-to-node (hop-to-hop) communication. Its specific responsibilities include framing, addressing, flow control, error control, and media access control.

The data link layer breaks down the bit stream it received from the network layer into manageable data units known as frames.

7.1 Types of Errors: Whenever bits travel from one point to another, they can be affected by unpredictable changes due to interference, which can alter the shape of the signal.

In a single-bit error, a 0 is changed to a 1 or a 1 to a 0.

In a burst error, multiple bits are altered. For example, a burst of impulse noise, such as 11100, on a transmission with a data rate of 1200 bps could change all or some of the 12 bits of information.

Single-Bit Error: The term single-bit error refers to a situation where only one bit of a given data unit (such as a byte, character, or packet) is altered, changing from 1 to 0 or from 0 to 1.

Burst Error: The term burst error refers to a situation where two or more bits in a data unit are changed from 1 to 0 or from 0 to 1.

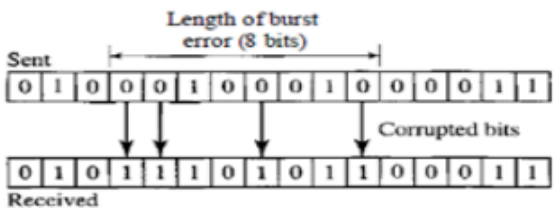


Figure 7.1: Burst error of length 8

Redundancy: The main concept in error detection or correction is redundancy. To identify or correct errors, additional bits are transmitted alongside the data. These extra bits are added by the sender and removed by the receiver. Their inclusion allows the receiver to detect or correct any corrupted bits.

Detection Versus Correction: The procedure followed for error correction is more complicated than error detection.

In error detection, the goal is simply to determine whether an error has occurred, with the answer being either yes or no. The number of errors is not relevant; a single-bit error is treated the same as a burst error.

In error correction, however, it is essential to identify the number of bits corrupted and their specific positions in the message. Both the number of errors and the message size are important factors. For example, to correct one error in an 8-bit data unit, we would need to check eight potential error locations; to correct two errors in the same 8-bit unit, we need to examine 28 possibilities. This illustrates the challenge of detecting and correcting multiple errors, especially in larger data units, such as locating 10 errors in a 1000-bit message.

Coding: Redundancy is incorporated through various coding techniques. The sender introduces extra bits by creating a relationship between these redundant bits and the actual data bits. The receiver then uses these relationships to detect or correct any errors. The ratio of redundant bits to data bits, along with the efficiency of the process, are key elements of any coding method.

Coding techniques are generally classified into two main types: block coding and convolution coding. This discussion will focus on block coding, as convolution coding is more complex and not the focus here.

7.2 BLOCK CODING**:** In block coding, the message is split into blocks, each consisting of k bits, known as datawords. We then add r redundant bits to each block, making the total length of the block n = k + r. The resulting n-bit blocks are called codewords. The method used to determine or calculate the extra r bits will be covered later. For now, it's important to note that we have a set of datawords, each with k bits, and a set of codewords, each with n bits. With k bits, we can generate 2^k possible datawords, and with n bits, we can generate 2^n possible codewords.

Since n > k, the number of codewords is greater than the number of datawords. The block coding process is a one-to-one mapping; each dataword is always encoded as the same codeword. As a result, there are 2^n - 2^k codewords that are not used, and these are referred to as invalid or illegal codewords. Figure 7.2 illustrates this concept.

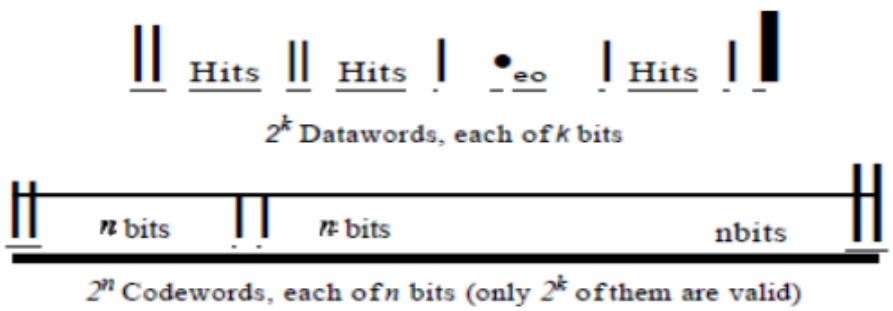


Figure 7.2: Datawords and codewords in block coding

Error Detection: Errors can be detected using block coding if the following two conditions are satisfied:

1. The receiver has access to (or can generate) a list of valid codewords.
2. The original codeword has been altered into an invalid codeword.

Figure 7.3 illustrates how block coding aids in error detection.

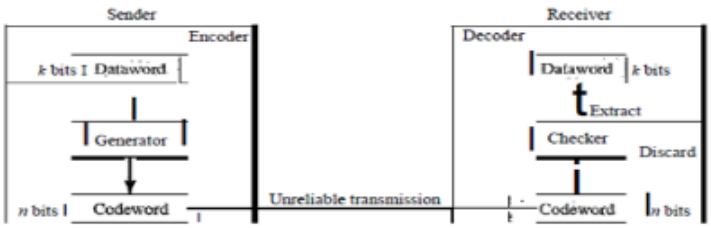


Figure 7.3 Process of error detection in block coding

Error Correction: As previously noted, error correction is significantly more complex than error detection. In error detection, the receiver simply needs to recognize that the received codeword is invalid. In contrast, error correction requires the receiver to determine (or infer) the original codeword that was transmitted. As a result, error correction demands more redundant bits than error detection. Figure 7.4 illustrates the role of block coding in error correction. The concept is similar to error detection, but the checker functions are much more complex.

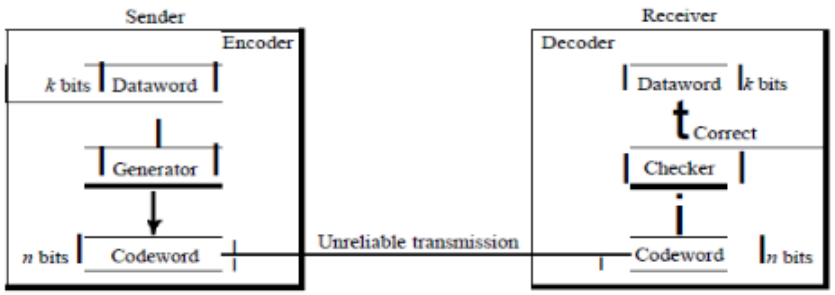


Figure 7.4 Structure of encoder and decoder in error correction

Hamming Distance: A fundamental concept in error control coding is the Hamming distance. The Hamming distance between two words (of the same length) is the number of bit positions in which they differ. This is represented as d(x, y) for two words x and y.

To calculate the Hamming distance, the XOR operation is applied to the two words, and the number of 1s in the result is counted. It's important to remember that the Hamming distance is always greater than zero.

The minimum Hamming distance refers to the smallest Hamming distance between any two words in a set.

Before proceeding, it's essential to highlight that every coding scheme must include at least three parameters: the codeword size n, the dataword size k, and the minimum Hamming distance dmin. A coding scheme is denoted as C(n, k) with a separate notation for dmin. For example, one coding scheme can be labeled as C(3, 2) with dmin = 2, and another as C(5, 2) with dmin = 3.

Hamming Distance and Error: To understand error detection and correction, it's important to first explore the concept of Hamming distance and how it relates to transmission errors. When a transmitted codeword is altered, the Hamming distance between the original and received codewords indicates how many bits have been changed due to the error. For example, if the codeword 00000 is sent but 01101 is received, three bits have been corrupted, and the Hamming distance between the two codewords is d(00000, 01101) = 3.

To ensure that up to **s** errors can be detected, the minimum Hamming distance for a block code must be d_min = s + 1.

On the other hand, to ensure that up to **t** errors can be corrected, the minimum Hamming distance for a block code must be d_min = 2t + 1.

7.3 Summary:

i) For detecting up to *s* errors in all cases, the minimum Hamming distance in a block code must be dmin = s+1.

ii) For correcting up to *t* errors in all cases, the minimum Hamming distance in a block code must be dmin == *2t* + 1.

7.4 Check your progress:

1. What is the primary function of the data link layer in networking?
2. How does a single-bit error differ from a burst error?

3. Define redundancy in the context of error detection and correction.
4. Why is error correction more complex than error detection?
5. What is the purpose of adding redundant bits in error detection and correction?
6. State the minimum Hamming distance required to detect up to 's' errors in all cases?
7. How can the minimum Hamming distance ensure the correction of up to 't' errors in a block code?

Module II: Media Access Control and Data Link Layer
(Unit-8- Media Access Protocols.)

8.0.Introduction:

Data Link layer provides the control mechanism for media access when multiple stations try to access a common media. The protocols that are commonly used in data link layer for media access control are broadly classified into three types: random access protocol, controlled access protocol and channelization protocol. The different protocols which falls under these catagories are shown in figure 8.0.

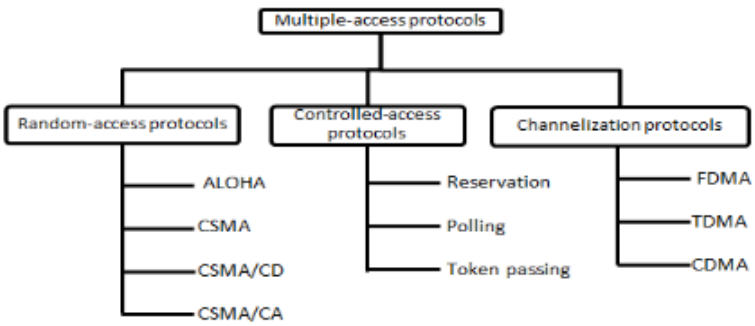


Figure: 8.0

8.1. ALOHA:

The original ALOHA protocol, referred to as pure ALOHA, is a simple but efficient communication method. In this system, each station sends a frame as soon as it has data to transmit. However, since all stations share the same channel, there is a possibility of frame collisions if two or more stations attempt to send their frames at the same time.

A collision happens when two or more stations transmit at the same time. If these stations try to

resend their frames after a time-out, another collision is probable. In pure ALOHA, each station waits for a random duration before retransmitting its frame following the time-out period.

This random delay helps to reduce the possibility of further collisions. This delay is referred to as the back-off time. The procedure for ALOHA is shown in figure 8.1.

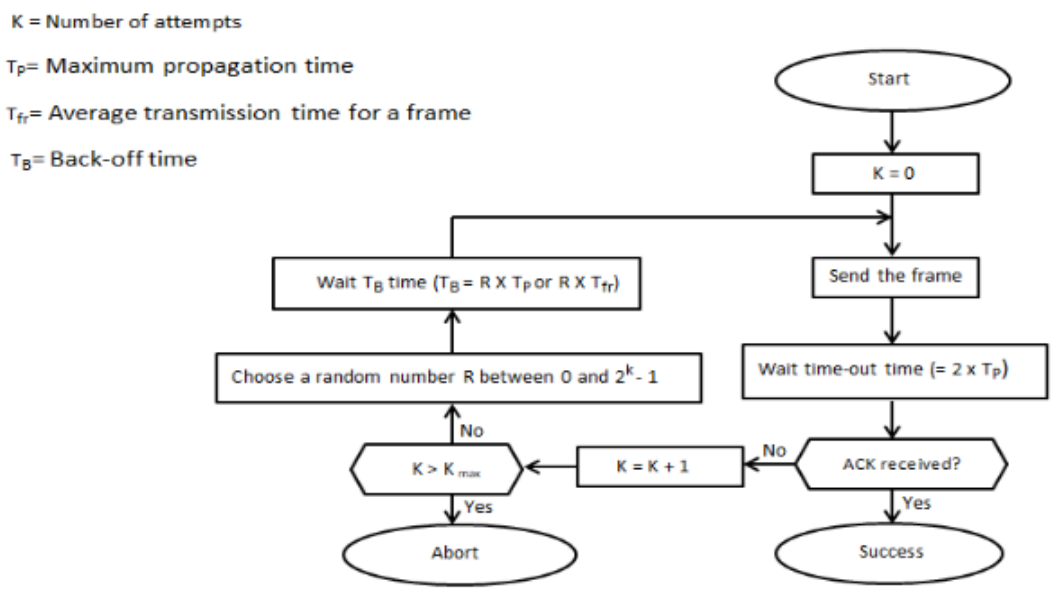


Figure 8.1: Procedure of ALOHA

The value of K_{max} is usually choosen as 15.

Vulnerable time: Let's calculate the duration of the vulnerable time, which is the period during which a collision can happen. We assume that stations transmit fixed-length frames, with each frame taking T_{fr} seconds to

send. Figure 8.2 illustrates the vulnerable time for station A.

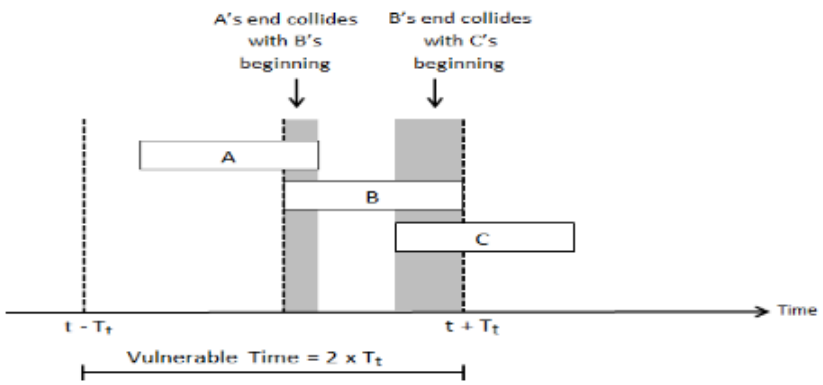


Figure 8.2: – Vulnerable time

Station A sends a frame at time t. Now, consider that station B has already sent a frame between

t – T_{fr} and t. This would cause a collision between the frames from station A and station B, with the end of B's frame overlapping the beginning of A's frame. On the other hand, if station C sends a frame between t and t + T_{fr}, a collision occurs between the frames from station A and station C, where the start of C's frame overlaps with the end of A's frame. From Figure 8.2, we can conclude that the vulnerable time during which a collision could occur in pure ALOHA is twice the frame transmission time. Thus, the vulnerable time for pure ALOHA = 2 x T_{fr}.

Throughput: Let G represent the average number of frames produced by the system in one frame transmission time. It

can be shown that the average number of successful transmissions in pure ALOHA is given by S = G x e^(-2G). The maximum throughput, S_{max}, is 0.184 when G=1/2.

8.2.Slotted Aloha: Pure ALOHA has a vulnerable time of 2 x Tfr because there are no specific rules regarding when a station can transmit. This means a station might send its frame immediately after one begins or just before another ends. To improve the efficiency of pure ALOHA, slotted ALOHA was developed. In slotted ALOHA, time is divided into intervals of Tfr seconds, and stations are required to transmit their frames only at the start of these intervals.Figure 8.3 illustrates an example of frame collisions in slotted ALOHA.

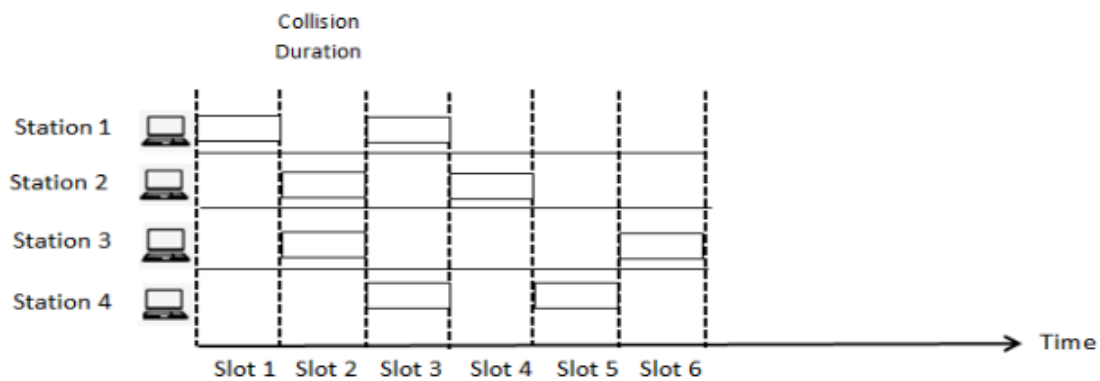


Figure 8.3: Allocation of time slots

Slotted ALOHA vulnerable time = $T_{fr.}$

Throughput: It can be demonstrated that the average number of successful transmissions in slotted ALOHA is S

= G x e^(-G). The maximum throughput, Smax, occurs at 0.368 when G is equal to 1.

8.3. Carrier Sense multiple Access (CSMA)

To reduce the likelihood of collisions and improve performance, the CSMA method was created. The chance of collision is minimized when a station checks the medium before attempting to use it. Carrier Sense Multiple Access (CSMA) requires each station to listen to the medium (or assess its status) before transmitting. In essence, CSMA follows the principle of "sense before transmit" or "listen before talking."

Persistence Methods:What should a station do when the channel is occupied? What action should it take if the channel is idle? Three persistence methods have been developed to address these situations: the 1-persistent method, the non-persistent method, and the p-persistent method.

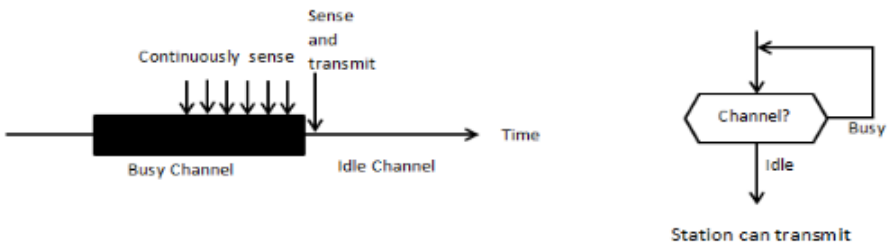


Figure 8.5: 1-persistent CSMA

Figure 8.5 illustrates how 1-persistence method behaves when a station encounters a busy channel.

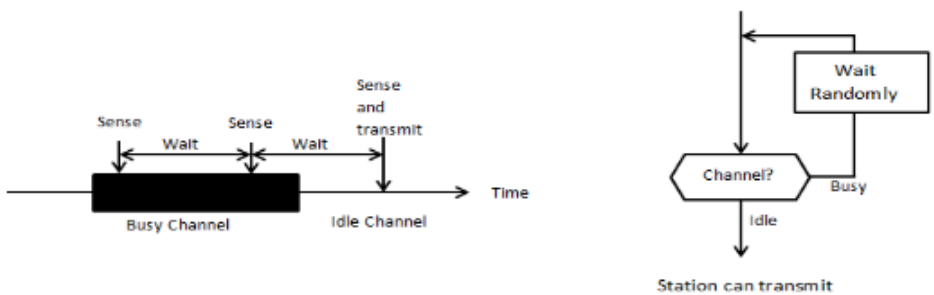


Figure 8.6: non-persistent CSMA

Figure 8.6 illustrates how non-persistence method behaves when a station encounters a busy or idle channel.

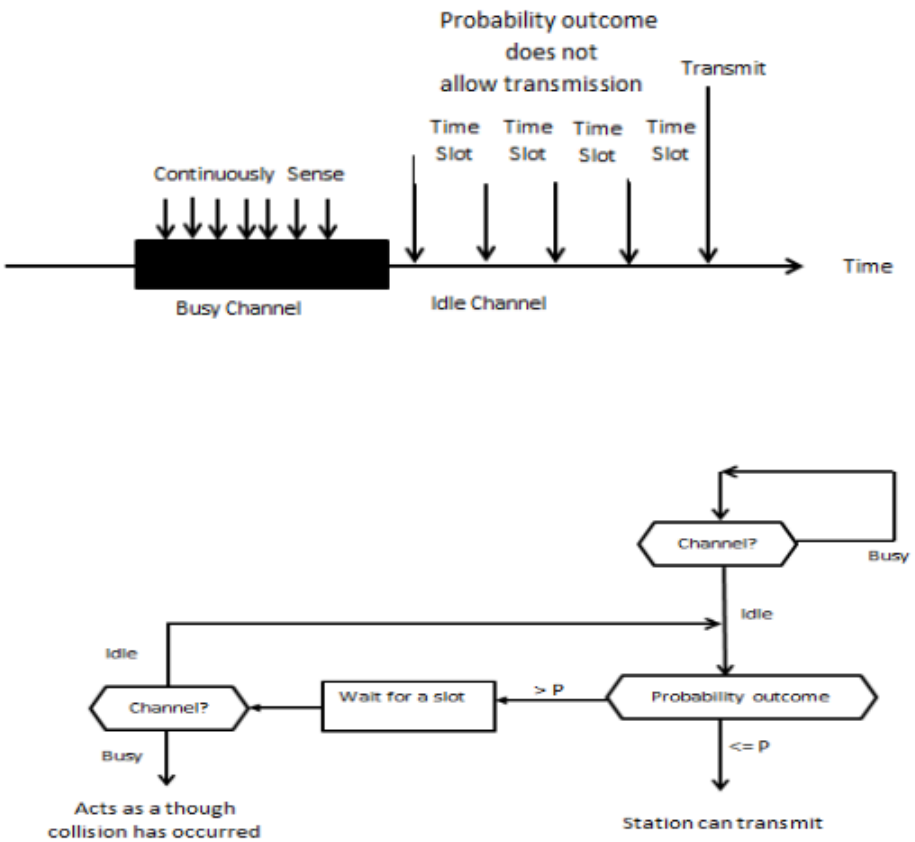


Figure 8.7: p - persistent approach

Figure 8.7 illustrates the behaviour of p-persistent method.

8.4.CSMA/CD: The CSMA method does not outline the steps to take after a collision occurs. Carrier Sense Multiple Access with Collision Detection (CSMA/CD) enhances the algorithm to address and manage collisions. The procedure followed in CSMA/CD is shown in figure 8.8.

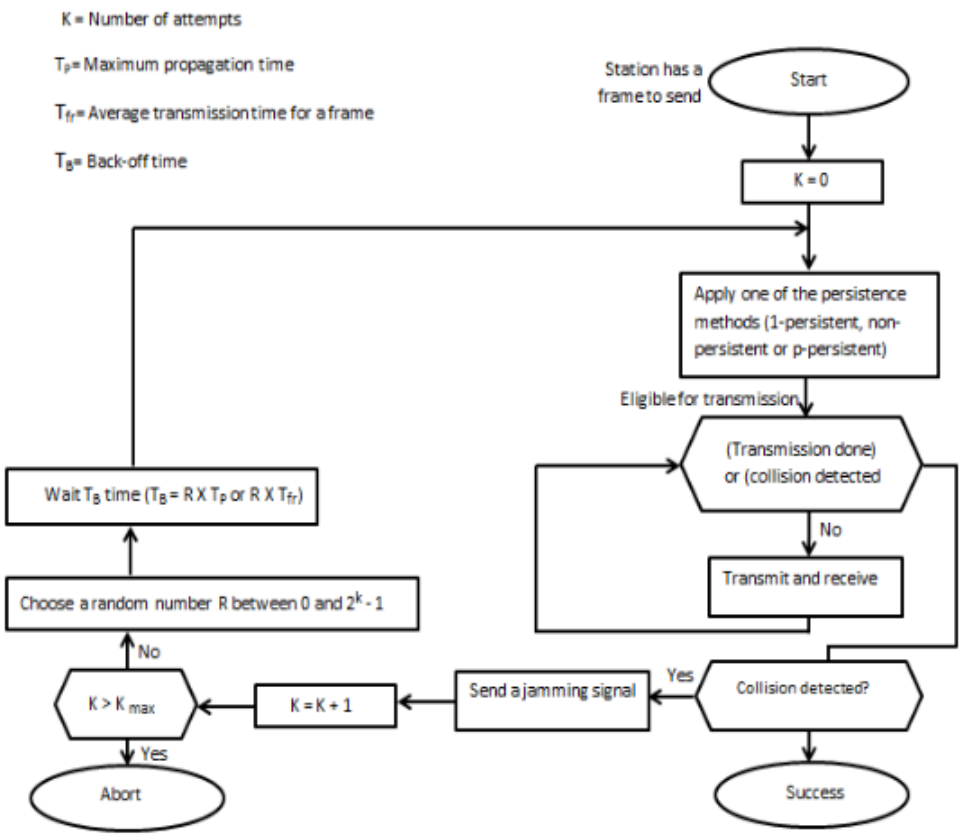


Figure 8.8: Procedure of CSMA/CD

8.5.CSMA/CA: The fundamental concept of CSMA/CD is that a station must be able to listen while transmitting in order to detect collisions. If no collision occurs, the station only receives its own signal. In the event of a collision, the station receives both its own signal and the signal from another station. To differentiate between these two scenarios, the signals must be noticeably different in strength. In other words, the second station's signal must significantly amplify the first station's signal.

In a wired network, the received signal typically matches the transmitted signal in strength, either due to short cable lengths or repeaters that boost the signal. In such cases, a collision almost doubles the energy detected. However, in wireless networks, much of the transmitted energy is lost during transmission, leaving the received signal with very low energy. As a result, a collision only contributes an additional 5 to 10 percent of energy, which is insufficient for reliable collision detection.

To address this issue, collisions must be avoided in wireless networks, as they can't be easily detected. Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) was developed to prevent collisions in such environments. CSMA/CA uses three techniques: interframe space, contention window, and acknowledgments -to avoid collisions.

Interframe Space (IFS): Collisions are prevented by delaying transmission even if the channel is idle. When a station detects an idle channel, it doesn't transmit right away. Instead, it waits for a period known as the interframe space (IFS). Although the channel may seem idle, a station further away may have already started transmitting, and its signal hasn't yet reached the local station. The IFS period gives enough time for the distant station's signal to arrive. If the channel is still idle after the IFS period, the station can then transmit, but it must first wait for a period corresponding to the contention time.

Contention Window: The contention window is a period of time divided into slots. When a station is ready to transmit, it selects a random number of slots as its waiting time. The number of slots in the window changes based on the binary exponential back-off method. Initially, the station waits for one slot, and each time it fails to detect an idle channel

after the IFS period, the waiting time doubles. This process is similar to the p-persistent method, but instead of a fixed time, the number of slots is determined randomly. A key feature of the contention window is that the station must check the channel after each slot. If the channel is found to be busy, the station does not restart the timer. It simply pauses the timer and resumes it once the channel becomes idle, thus giving priority to the station that has been waiting the longest.

Acknowledgment:. To ensure that the receiver has successfully received the frame, a positive acknowledgment along with a time-out timer are used.

The procedure followed in CSMA/CA is shown in the figure 8.9.

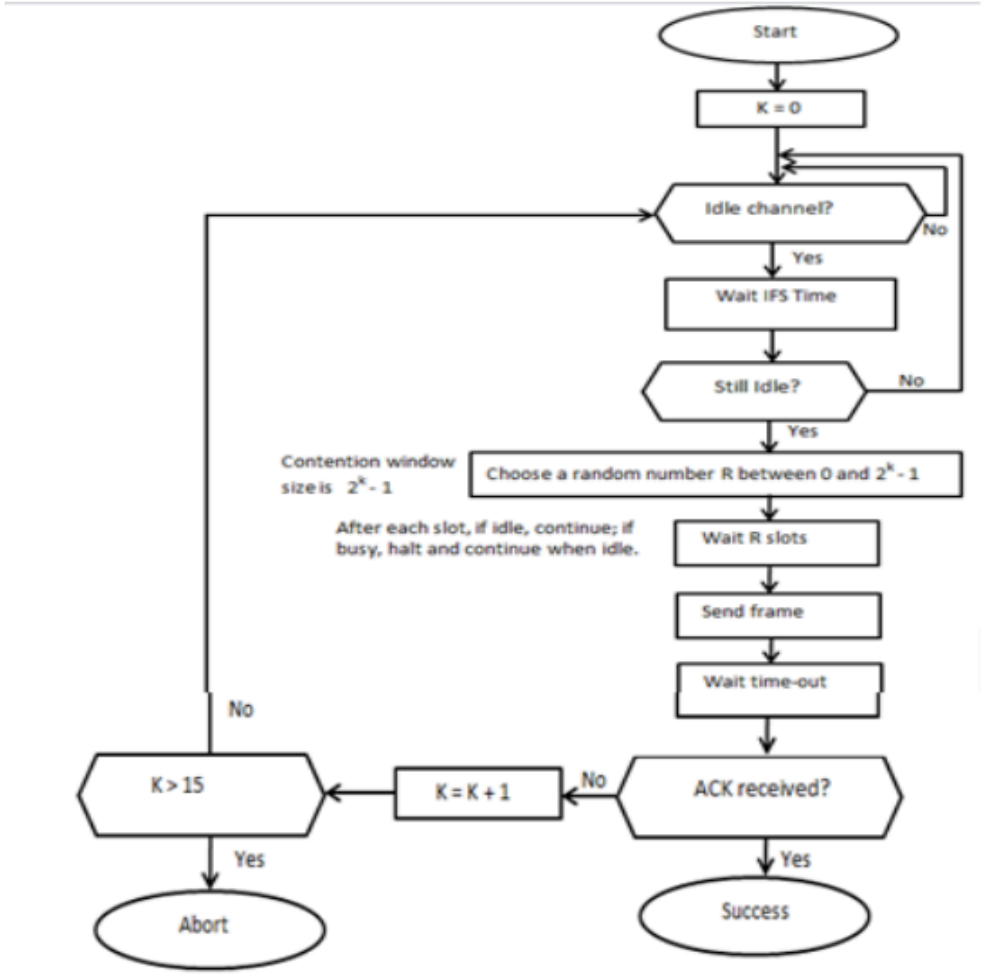


Figure 8.9: Procedure of CSMA/CA

8.6.Summary:

i) Various formal protocols have been created to manage access to a shared link. These protocols are divided into three categories: random access protocols, controlled access protocols, and channelization protocols.

ii) ALOHA enables multiple access (MA) to the shared medium, but this setup can lead to collisions. When one

station transmits data, another station may try to transmit simultaneously, causing the two signals to collide and become corrupted.

iii) Carrier Sense Multiple Access with Collision Detection (CSMA/CD) enhances the CSMA algorithm to manage collisions. In this approach, a station listens to the medium after transmitting a frame to verify if the transmission was successful. If the transmission is successful, the station completes the process. If a collision occurs, the frame is retransmitted.

iv) To prevent collisions in wireless networks, Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) was developed. Collisions are avoided using three strategies: the interframe space, the contention window, and acknowledgments.

8.7 Check your progress:

1. What is the main idea behind the original ALOHA protocol?
2. What is the back-off time in ALOHA, and why is it important?
3. How is the vulnerable time in pure ALOHA calculated, and what does it represent?
4. What is the throughput formula for Slotted ALOHA, and what is its maximum throughput?
5. Explain the concept of Carrier Sense Multiple Access (CSMA) and its role in reducing collisions.
6. What are the three persistence methods used in CSMA, and how do they behave when the channel is idle or busy?
7. How does CSMA/CD work to detect collisions after a transmission, and what happens if a collision is detected?

8. Why is CSMA/CD not effective in wireless networks, and what solution was developed for such environments?

Module II: Media Access Control and Data Link Layer
(Unit-9- Flow Control.)

9.0.Introduction: In the data link layer of the OSI model, flow control mechanisms are essential to ensure reliable communication between devices. The primary goal is to manage the rate of data transmission so that the receiver is not overwhelmed by the incoming data. Four key protocols used for flow control are Stop-and-Wait ARQ, Go-Back-N ARQ, Selective Repeat ARQ, and Piggybacking. These protocols ensure that frames are successfully transmitted, acknowledged, and retransmitted if necessary.

9.1. Stop-and-Wait ARQ (Automatic Repeat reQuest): The Stop-and-Wait ARQ protocol is one of the simplest and earliest methods of ensuring reliable communication in data link layer flow control. In this protocol, the sender sends one frame at a time and waits for an acknowledgment (ACK) from the receiver before sending the next frame. The process works as follows:

- The sender transmits a frame and stops to wait for the receiver's acknowledgment.
- Once the acknowledgment for the frame is received, the sender sends the next frame.
- If the sender does not receive an acknowledgment within a specified time (due to packet loss or other issues), it retransmits the frame.
- This process ensures reliable communication, but it can be inefficient, especially for high-latency networks, because the sender remains idle while waiting for the acknowledgment.

The primary advantage of Stop-and-Wait ARQ is its simplicity. However, its main limitation is that the sender can only transmit one frame at a time, leading to low link utilization and poor performance over long distances or networks with high round-trip delays.

9.2. Go-Back-N ARQ: Go-Back-N ARQ improves upon the Stop-and-Wait ARQ by allowing the sender to transmit multiple frames before needing to wait for an acknowledgment for each individual frame. The main feature of Go-Back-N is that it allows the sender to transmit a window of N frames without waiting for acknowledgments for each one. The process works as follows:

- The sender maintains a "window" of N frames, which can be sent consecutively without waiting for an acknowledgment after each frame.
- Each frame is assigned a sequence number. The receiver is expected to acknowledge frames in sequence.
- If a frame is lost or corrupted, the receiver discards it and sends a negative acknowledgment (NACK) for that specific frame.
- Upon receiving a NACK, the sender will retransmit the frame and all subsequent frames in the sequence, starting from the lost or corrupted frame.
- This method is more efficient than Stop-and-Wait because it increases the throughput by allowing the sender to keep transmitting without waiting for individual acknowledgments. However, it can lead to the retransmission of multiple frames if one frame in the sequence is lost or corrupted.

The disadvantage of Go-Back-N ARQ is that, in the event of a lost frame, all subsequent frames must be

retransmitted, leading to potentially unnecessary retransmissions.

3. Selective Repeat ARQ: The Selective Repeat ARQ protocol is an enhancement over Go-Back-N ARQ. It also allows the sender to transmit multiple frames at once, but unlike Go-Back-N, it only retransmits the specific frames that were lost or corrupted, rather than all frames following the lost one. The process works as follows:

- Similar to Go-Back-N, the sender can send multiple frames before receiving an acknowledgment for each.
- Each frame is given a sequence number, and the receiver is expected to acknowledge frames individually.
- If a frame is lost or corrupted, the receiver will send a NACK for that specific frame, requesting it to be retransmitted.
- The sender only retransmits the frame that was negatively acknowledged, not all subsequent frames.
- This selective acknowledgment helps reduce unnecessary retransmissions, making the protocol more efficient than Go-Back-N ARQ.

Selective Repeat ARQ is more efficient than Go-Back-N because it reduces the number of retransmitted frames. However, it requires the receiver to have a buffer to store out-of-order frames, as frames may arrive at the receiver out of order if retransmissions occur.

9.4. Piggybacking: Piggybacking is a technique used in both Go-Back-N and Selective Repeat ARQ to improve efficiency by combining data and acknowledgment frames. Instead of sending separate acknowledgment frames, the

receiver can include its acknowledgment within the data frame it sends back to the sender. This process works as follows:

- In a typical ARQ system, the receiver sends an acknowledgment frame to the sender to confirm the successful receipt of a data frame.
- In piggybacking, the receiver waits until it has data to send back to the sender. When the receiver sends its own data frame, it "piggybacks" the acknowledgment for the previous frame in the header of this new data frame.
- By combining the acknowledgment and data frame, piggybacking reduces the overhead and increases the efficiency of the communication, especially in bidirectional communication systems.

Piggybacking is most effective in full-duplex systems where both the sender and the receiver can transmit and receive data simultaneously. It helps reduce the number of frames on the network, optimizing bandwidth and reducing delays, making it particularly useful in situations with a lot of back-and-forth data exchange.

9.5.**Summary:**

i) Each of these protocols—Stop-and-Wait ARQ, Go-Back-N ARQ, Selective Repeat ARQ, and Piggybacking—plays a crucial role in the management of flow control at the data link layer.

ii)Stop-and-Wait ARQ is simple but inefficient for high-latency networks, while Go-Back-N and Selective Repeat ARQ improve throughput by allowing multiple frames to be sent before waiting for acknowledgments.

iii) Selective Repeat ARQ is more efficient than Go-Back-N, as it minimizes unnecessary retransmissions.

iv) Piggybacking optimizes communication by combining data and acknowledgment frames, reducing overhead and improving efficiency in bidirectional systems.

9.6 Check your progress:

1. What is the main limitation of the Stop-and-Wait ARQ protocol?
2. In the Go-Back-N ARQ protocol, how does the sender improve efficiency compared to Stop-and-Wait ARQ?
3. What is the purpose of the "window" in Go-Back-N ARQ?
4. How does the Selective Repeat ARQ protocol differ from Go-Back-N ARQ?
5. What happens when a frame is lost or corrupted in Selective Repeat ARQ?
6. In which type of communication systems is piggybacking most effective?

Module II: Network Layer
(Unit-10: IPV_4 address)

10.0.Introduction: The communication at the network layer occurs between hosts (computer-to-computer); a computer in one location communicates with another computer in a different location. Typically, these computers connect via the Internet. The data packet sent by the source computer might traverse multiple LANs or WANs before reaching its destination. For this type of communication, a global addressing system is required, which is referred to as logical addressing. Today, the term IP address is used to

describe the logical address at the network layer of the TCP/IP protocol suite.

Internet addresses are 32 bits long, providing a total of 2^32 possible addresses. These are known as IPv4 (IP version 4) addresses, or simply IP addresses when the context is clear. The growing demand for more addresses, along with other issues related to the IP layer, led to the development of a new version, IPv6 (IP version 6). This version uses 128-bit addresses, allowing for greater flexibility in address allocation. These are known as IPv6 addresses. In this module-III, we will first cover IPv4 addresses, which are still in use today, and then explore IPv6 addresses, which may become the dominant standard in the future.

10.1. IPV4 address: An IPv4 address is a 32-bit identifier that uniquely and globally represents a device's connection to the Internet, such as a computer or router. It is 32 bits in length and ensures that each device connected to the Internet has a distinct address. No two devices can share the same address at the same time. However, through specific techniques, an address can be temporarily assigned to one device and later reassigned to another. Additionally, a device with multiple connections to the Internet, such as a router, would require multiple addresses. The IPv4 addressing system is universal, meaning that all hosts that wish to connect to the Internet must adhere to it. IPv4 addresses are both unique and universal. The address space of a protocol like IPv4 refers to the total number of possible addresses it can use. For IPv4, which employs 32-bit addresses, the address space consists of 2^32 addresses, or 4,294,967,296, meaning that over 4 billion devices could theoretically be connected to the Internet. However, due to certain limitations, the actual number of usable addresses is

less than this. There are two common ways to represent an IPv4 address: binary notation and dotted-decimal notation.

In binary notation, an IPv4 address is expressed as 32 bits, with each octet (or byte) typically displayed separately. For example, an IPv4 address in binary might look like this:

01110101 10010101 00011101 00000010

In dotted-decimal notation, the address is written in decimal format, with each byte separated by a dot for easier readability. The same address in dotted decimal notation would appear as: 117.149.29.2

10.2.Classful addressing: When IPv4 addressing was first introduced, it used a system based on address classes, known as classful addressing. Although this method is now largely outdated, it is briefly explained here to provide context for the transition to classless addressing. In classful addressing, the address space is divided into five distinct classes: A, B, C, D, and E. Each class covers a specific portion of the address space.

To determine the class of an address, we can examine the address in either binary or dotted-decimal notation. In binary format, the initial bits of the address reveal its class. In dotted-decimal notation, the first byte of the address identifies its class. The ranges of the five classes along with their usage are shown in table 10.1.

Class	Address Range (Decimal)	Address Range (Binary)	Default Subnet Mask	Usage
A	0.0.0.0	00000000.0000000	255.0.0	Used for

	to 127.255.255.255	0.00000000.00000000 to 01111111.11111111.11111111.11111111	.0 (8 bits)	large networks, supporting up to 16 million hosts per network.
B	128.0.0.0 to 191.255.255.255	10000000.00000000.00000000.00000000 to 10111111.11111111.11111111.11111111	255.255.0.0 (16 bits)	Used for medium-sized networks, supporting up to 65,000 hosts per network.
C	192.0.0.0 to 223.255.255.255	11000000.00000000.00000000.00000000 to 11011111.11111111.11111111.11111111	255.255.255.0 (24 bits)	Used for small networks, supporting up to 254 hosts per network.
D	224.0.0.0 to	11100000.00000000.00000000.00000	N/A	Reserved for

	239.255.255.255	000 to 11101111.11111111.11111111.11111111		multicast addressing, used to send data to multiple receivers simultaneously.
E	240.0.0.0 to 255.255.255.255	11110000.00000000.00000000.00000000 to 11111111.11111111.11111111.11111111	N/A	Reserved for experimental purposes and future use.

Table 10.1: Ranges of different classes

10.3. Subnetting**:** Subnetting was introduced during the classful addressing period. When an organization was allocated a large address block in Class A or B, it could divide the addresses into several smaller, contiguous groups. Each group could then be assigned to individual networks, known as subnets, or, in some instances, the addresses could be shared with neighboring networks. Subnetting effectively increases the number of bits in the subnet mask.

10.3. Address Depletion**:** The limitations of the classful addressing system, coupled with the rapid expansion of the Internet, resulted in the near exhaustion of available addresses. Although the number of devices on the Internet is still far below the 2^32 address space, we have exhausted the class A and B address ranges, and class C blocks are too small for many medium-sized organizations. One solution that has helped address this issue is the inception of classless addressing.

10.4.Summary:

i) Classful addressing is an early method of dividing the IPv4 address space into distinct address classes (A, B, C, D, E)
ii) Subnetting allowed for more efficient use of address space by allocating smaller network portions.
iii) The rigid class boundaries led to inefficient use of the available address space, as larger networks with fewer hosts could not use smaller address ranges effectively.

10.5 Check your progress:

1. What is the main purpose of logical addressing at the network layer?
2. How long is an IPv4 address, and how many possible addresses does it provide?
3. What is the difference between IPv4 and IPv6 ?
4. What is the primary method for representing an IPv4 address?
5. How can the class of an IPv4 address be determined?

6. What was the impact of classless addressing on the efficient use of the IPv4 address space?

Module II: Network Layer
(Unit-11: Classless address)

11.0.Introduction: To address the issue of address depletion and provide more organizations with Internet access, classless addressing was developed and put into practice. In this system, there are no predefined address classes, but addresses are still allocated in blocks.

11.1.Classless addressing: In classless addressing, when an entity, regardless of size, requires an Internet connection, it is assigned a block (range) of addresses. The block's size, or the number of addresses, depends on the entity's needs and scale. For instance, a household might receive just two addresses, while a large organization could be allocated thousands. An Internet Service Provider (ISP), based on the number of customers it serves, might receive tens of thousands or even hundreds of thousands of addresses.

To streamline address management, Internet authorities enforce three rules on classless address blocks:

1. The addresses within a block must be consecutive.
2. The total number of addresses in a block must be a power of 2 (1, 2, 4, 8, etc.).
3. The first address in the block must be divisible evenly by the total number of addresses.

11.2.Mask**:** A more efficient way to define a range of addresses is to choose any address within the block and its corresponding mask. As previously mentioned, a mask is a 32-bit value where the leftmost n bits are set to 1, and the remaining 32 - n bits are set to 0. However, in classless addressing, the mask for a block can vary from 0 to 32. It is more convenient to simply specify the value of n, preceded by a slash (known as CIDR notation).

In IPv4 addressing, a block of addresses can be represented as
x.y.z.t/n, where x.y.z.t is an address in the block and n represents the mask.

The address and the n value together fully describe the block, including the first address, the last address, and the total number of addresses in the block.

The first address in the block can be determined by setting the rightmost 32 - n bits to 0 in the binary representation of the address.The number of addresses in the block is the difference between the last and first address. It can easily be found using the formula 2^{32-n}.

11.3.IPv6: IPv6 (Internet Protocol version 6) is the latest version of the Internet Protocol (IP) designed to address the limitations of IPv4, particularly the exhaustion of available IP addresses. IPv6 offers a much larger address space and introduces several improvements in terms of security, routing, and address configuration.

IPv6 Address Format: An IPv6 address is a 128-bit address, written as eight groups of four hexadecimal digits, separated by colons. Each group represents 16 bits (2 bytes), and the entire address is expressed as follows:

xxxx:xxxx:xxxx:xxxx:xxxx:xxxx:xxxx:xxxx

Each group uses hexadecimal notation (0-9, a-f), and each hexadecimal digit represents 4 bits, meaning two hexadecimal digits represent one byte (8 bits).The 8 groups are separated by colons.

11.4 IPV4 header: Here is a table summarizing the IPv4 header format and an explanation of each field:

Field	Length (in bits)	Description
Version	4	Indicates the version of the IP protocol. For IPv4, this value is set to 4.
IHL (Internet Header Length)	4	Specifies the length of the IP header in 32-bit words. The minimum value is 5 (indicating 20 bytes), and it can go higher for headers with options.
Type of Service (ToS)	8	Defines how the datagram should be handled. It includes parameters for priority, delay, throughput, and reliability.
Total Length	16	Specifies the entire length of the IP packet, including both the header and the data, in bytes.
Identification	16	Used to uniquely identify fragments of a datagram, enabling the reassembly of fragments into the original packet.
Flags	3	Controls fragmentation. It includes a reserved bit, a "Don't Fragment" bit (DF), and a "More Fragments" bit (MF).
Fragment Offset	13	Indicates the position of the fragment's data in the original datagram (used during fragmentation).
Time to Live (TTL)	8	Specifies the maximum number of hops a packet can make before being discarded.

		It is decreased by one at each hop.
Protocol	8	Indicates the protocol used in the data portion of the IP packet (e.g., TCP, UDP, ICMP).
Header Checksum	16	A checksum used for error-checking the header. It helps detect errors in the IP header during transmission.
Source Address	32	Contains the 32-bit IPv4 address of the sender (source).
Destination Address	32	Contains the 32-bit IPv4 address of the recipient (destination).
Options (Optional)	Variable	Provides additional control information for special handling or routing. This field is optional and can vary in length.
Padding	Variable	Ensures that the header is a multiple of 32 bits. It is used if the Options field is present and doesn't fill up the entire header length.

Table: 11.1

Field Explanations:

- Version: The version field defines which IP version the packet is using. For IPv4, this is always 4.
- IHL (Internet Header Length): This field specifies the size of the IP header in 32-bit words. It indicates how long the header is, which helps in identifying the start of the data section.

- Type of Service (ToS): The ToS field helps prioritize packets based on delay, throughput, and reliability requirements. In modern use, it's often replaced by Differentiated Services Code Point (DSCP).
- Total Length: The total length field is crucial to know the entire size of the packet, including the header and the data. This helps in determining where the packet ends.
- Identification: This field helps uniquely identify fragments of a larger packet. It allows the receiver to correctly reassemble the fragments back into the original packet.
- Flags: These three bits control packet fragmentation. The "Don't Fragment" (DF) bit prevents fragmentation, while the "More Fragments" (MF) bit indicates that more fragments follow.
- Fragment Offset: When a packet is fragmented, this field tells the receiver where the data of this fragment belongs in the original packet.
- Time to Live (TTL): This field limits the lifespan of the packet in the network by decrementing by one at each hop. If TTL reaches zero, the packet is discarded to avoid endless looping.
- Protocol: The protocol field defines which transport layer protocol is being used (e.g., TCP is 6, UDP is 17, ICMP is 1).
- Header Checksum: A 16-bit checksum that is used for error-checking the integrity of the header. The checksum is recalculated and verified at each hop to ensure no corruption occurred in transit.
- Source Address: Contains the IP address of the machine sending the packet, allowing the recipient to know where the packet came from.

- Destination Address: Contains the IP address of the destination machine, determining where the packet should go.
- Options: The options field can contain additional settings such as time-stamping, security, or routing information. This field is optional and is rarely used in practice.
- Padding: Padding is used when the length of the IP header isn't a multiple of 32 bits, ensuring the header aligns correctly.

The total length of the IPv4 header can range from 20 to 60 bytes, depending on the presence of options.

11.5.Summary:

i) An IPv4 address is 32 bits in length and provides a unique, global identifier for a host or router on the Internet.

ii) In classful addressing, the part of the IP address that represents the network is called the netid.

iii)Subnetting is the process of splitting a large network into smaller subnetworks, introducing an additional layer of hierarchy in IP addressing.

11.6 Check your progress:

1. List the three rules enforced by Internet authorities on classless address blocks.

2. What is CIDR notation, and how does it simplify the representation of address blocks in IPv4?

3. Explain how the first address in a classless block is determined.

4. What are the main differences between IPv4 and IPv6 addressing?

5. What is the purpose of the IPv4 header, and how does it facilitate packet delivery across networks?

6. In the IPv4 header, what is the role of the "Time to Live" (TTL) field?

7. What does the "Protocol" field in an IPv4 header indicate?

.

8. What is subnetting, and how does it enhance IP address management in large networks?

Module III: Network Layer
(Unit-12: Transition from IPV4 to IPV6)

12.0.Introduction: Due to the vast number of systems on the Internet, the shift from IPv4 to IPv6 cannot occur abruptly. It will take a significant amount of time before all systems on the Internet can transition from IPv4 to IPv6. The changeover must be gradual to avoid issues between IPv4 and IPv6 systems. The IETF has developed three strategies to facilitate this transition: dual stack, tunneling and header translation method.

12.1 Dual Stack Method: The dual stack method allows systems to run both IPv4 and IPv6 protocols simultaneously. This approach enables devices, routers, and networks to support both IPv4 and IPv6 addresses, ensuring that they can communicate with both IPv4-only and IPv6-only devices.

Key Points of Dual Stack:

i) Simultaneous Operation: Devices use both IPv4 and IPv6 addresses. They can communicate over either protocol, depending on the destination address.
ii) Compatibility: It ensures compatibility between IPv4 and IPv6 networks by enabling devices to operate using IPv4 or IPv6, depending on the requirements of the communication.
iii) Implementation: A device that supports dual stack must have both an IPv4 address and an IPv6 address. Routers and other network devices must also be dual-stack capable to forward packets for both IP versions.
iv) Advantages:

----No need for major modifications in existing systems or networks.

----IPv4 and IPv6 systems can coexist, and there's no immediate need for the entire network to switch to IPv6.

----Provides a gradual transition mechanism.

v) Disadvantages:

----More complex network management due to maintaining both IPv4 and IPv6 configurations.

----Increased resource consumption (memory, processing) because both protocol stacks are running.

12.2.Tunneling Method: Tunneling involves encapsulating IPv6 packets inside IPv4 packets (or vice versa) to traverse IPv4 infrastructure. It allows IPv6 traffic to be sent over an IPv4 network by encapsulating the IPv6 data within an IPv4 header.

Key Points of Tunneling:

i) Encapsulation of IPv6 in IPv4: In this method, IPv6 packets are encapsulated in IPv4 packets for transmission across IPv4-only networks. The encapsulated IPv6 packets are then decapsulated at the destination, allowing IPv6 communication.

ii) Tunnel Types:

------6to4 Tunnel: IPv6 packets are transmitted over IPv4 networks by automatically assigning IPv6 addresses based on IPv4 addresses.

-----ISATAP (Intra-Site Automatic Tunnel Addressing Protocol): A mechanism for automatically assigning IPv6 addresses using IPv4 within a private network.

-------Teredo: A tunneling protocol that provides IPv6 connectivity across NAT (Network Address Translation)-enabled IPv4 networks.

iii) Advantages:

-------Allows IPv6 packets to traverse IPv4-only networks, making it possible for organizations with IPv4-only infrastructure to begin using IPv6.

-------Facilitates a gradual transition as it allows IPv6 traffic over an IPv4 backbone.

iv) Disadvantages:

------Increased overhead due to the encapsulation and decapsulation process.

-------Complexity in network configuration and management.

-------Can introduce latency and potential reliability issues.

12.3.Header Translation Method (NAT64 and DNS64):
The header translation method involves translating IPv6 packets into IPv4 packets (or vice versa) at the network boundary. This is typically used in networks where IPv4

and IPv6 systems need to communicate directly, and there's no direct path for communication between them.

Key Points of Header Translation:

i) NAT64 (Network Address Translation 64):

----NAT64 enables IPv6-only clients to communicate with IPv4-only servers. It translates IPv6 packets into IPv4 packets at the network boundary, allowing IPv6-only devices to access IPv4 services.

------Translation Process: The IPv6 address is translated into an IPv4 address, and the IPv6 header is rewritten to conform to IPv4 standards.

ii) DNS64:

-------DNS64 works in conjunction with NAT64 to enable IPv6-only clients to resolve domain names of IPv4-only services. It synthesizes IPv6 addresses from IPv4 addresses, allowing IPv6 clients to reach IPv4 servers.

---------When an IPv6-only device queries a DNS server for a website, DNS64 responds with an IPv6 address corresponding to the IPv4 address of the destination, allowing the device to communicate via NAT64.

iii) Advantages:

-------Enables communication between IPv6-only and IPv4-only devices,

facilitating the coexistence of both protocols.

-------Helps maintain IPv4-only services accessible to IPv6-only clients.

iv) Disadvantages:

------Requires additional configuration and devices to support NAT64 and DNS64.

------Can cause issues with certain applications (e.g., protocols that embed IP addresses in the payload, such as FTP).

-------May add some overhead and complexity in managing the translation process.

12.4.Summary:

Comparison of Methods:

Method	Description	Advantages	Disadvantages
Dual Stack	Devices and networks run both IPv4 and IPv6.	Simple to implement, supports both IPv4 and IPv6.	Complex management, higher resource consumption.
Tunneling	Encapsulating IPv6 in IPv4 (or vice versa) to traverse non-IPv6	Allows IPv6 to pass over IPv4 networks.	Increased overhead, potential latency, and reliability issues.

	networks.		
Header Translation (NAT64, DNS64)	Translates IPv6 packets to IPv4 (and vice versa) at network boundaries.	Enables communication between IPv6-only and IPv4-only networks.	Requires configuration, potential issues with certain protocols.

In conclusion, all three transition methods—dual stack, tunneling, and header translation—offer different approaches for enabling IPv4 and IPv6 communication. The choice of method depends on network requirements, infrastructure, and the need for a gradual transition during the IPv4 to IPv6 shift.

12.5 Check your progress:

1. What are the three strategies developed by the IETF to facilitate the transition from IPv4 to IPv6?

2. Explain the concept of the dual stack method in IPv6 transition. How does it work?

3. What are the disadvantages associated with implementing the dual stack method?

4. How does the tunneling method enable IPv6 traffic to traverse

5. What are the main advantages and disadvantages of using the tunneling method for IPv6 transition?

6. What is the header translation method (NAT64 and DNS64), and how does it facilitate communication between IPv6-only and IPv4-only systems?

Module III: Network Layer
(Unit-13: Address mapping,delivery, forwarding and routing protocols)

13.0.Introduction: The Internet Protocol (IP) is the primary protocol at the network layer. While it was designed for best-effort delivery, it does not include features like flow control or error correction. IP functions as a host-to-host protocol using logical addressing. However, to address the requirements of contemporary networking, other protocols are necessary in addition to IP.

For example, protocols are needed to translate physical addresses into logical ones. While IP packets use logical (host-to-host) addresses, they need to be encapsulated within a frame that relies on physical (node-to-node) addresses. The Address Resolution Protocol (ARP) was created to perform this translation. Additionally, in certain situations, reverse mapping is required, such as when starting a diskless system or assigning an IP address to a host. This is accomplished through protocols like RARP, BOOTP, and DHCP.

The absence of flow and error control in IP led to the creation of the ICMP protocol, which is used to inform about network congestion and specific errors at the destination host. Initially, IP was designed for unicast delivery, where data is sent from one source to a single destination. However, as the Internet grew, there was a substantial rise in the need for multicast delivery, where a single source transmits data to multiple destinations.

13.1.Address mapping: The internet is made up of various physical networks connected by internetworking devices such as routers. A packet sent from a source host may traverse several distinct physical networks before reaching its destination host. At the network layer, hosts and routers are

recognized by their logical (IP) addresses, while devices are identified by their physical addresses as packets travel through physical networks.

A physical address is a local identifier within a network, required to be unique within that specific network but not necessarily globally. It is called a physical address because it is typically, though not always, implemented in hardware. A common example is the 48-bit MAC address used in Ethernet, which is assigned to the NIC in a host or router.

Physical and logical addresses are separate identifiers, both of which are necessary because a single physical network, such as Ethernet, can support multiple network-layer protocols, like IP and IPX (Novell), at the same time. Likewise, a packet using the IP network layer may pass through different physical networks, such as Ethernet and LocalTalk (Apple). Therefore, delivering a packet to a host or router requires both types of addresses. Mapping between logical and physical addresses is crucial and can be done through static or dynamic methods.

Static mapping involves creating a table that links logical addresses to physical addresses, stored on each device in the network. If a device knows the IP address of another but not its physical address, it can consult the table to find the information. However, static mapping has its drawbacks since physical addresses may change for various reasons.

1. A machine may change its NIC, altering its physical address.

2. In some LANs, such as LocalTalk, physical addresses change every time the computer restarts.
3. A mobile device moving between networks can have its physical address change.

To handle these changes, the static mapping table needs to be updated frequently, which can introduce overhead and affect network performance. In dynamic mapping, when a device knows one of the addresses (either logical or physical), it can use a protocol to find the corresponding address.

13.2 Mapping Logical to Physical Address:

When a host or router needs to send an IP datagram to another device, it knows the recipient's logical (IP) address, which it can obtain from the DNS (if the sender is a host) or a routing table (if the sender is a router). However, to send the IP datagram over the physical network, it must be wrapped in a frame, which requires the destination device's physical address. To find this, the sender sends an ARP query packet across the network, which contains both the sender's physical and IP addresses, as well as the recipient's IP address.

Every device on the network receives and processes the ARP query, but only the device with the corresponding IP address replies with an ARP response. This reply includes both the recipient's IP and physical addresses and is sent directly to the requester using the physical address specified in the query.

13.3 Mapping Physical to Logical Address: RARP, BOOTP, and DHCP:

There may be situation where a host knows its physical address but requires its logical address. This can occur in two scenarios:

1. A diskless station that has just started up can identify its physical address through its interface but does not yet have an assigned IP address.
2. An organization with not enough IP addresses for each station may need to assign IP addresses dynamically. In this case, the station can send its physical address and request a temporary IP assignment.

RARP: The Reverse Address Resolution Protocol (RARP) is used to find the logical address of a device that only knows its physical address. Every host or router is assigned one or more unique logical (IP) addresses, separate from the device's physical (hardware) address. To create an IP datagram, the device must know its IP address, which is usually stored in a configuration file on the disk.

However, a diskless device that boots from ROM, which contains only basic boot information often provided by the manufacturer, does not have an IP address. Network administrators assign IP addresses. The machine can identify its unique physical address (e.g., from its NIC) and use the RARP protocol to request its IP address. The RARP request is broadcast on the local network, and a machine with the IP address responds with a RARP reply. The requesting machine must run a RARP client, while the responding machine must have a RARP server.

A major drawback of RARP is that its broadcasts occur at the data link layer. The physical broadcast address, such as the "all" address used in Ethernet, is limited to a single network and doesn't reach across subnets. Therefore, separate RARP

servers are needed for each subnet in networks with multiple subnets. Due to this limitation, RARP has been largely replaced by BOOTP and DHCP.

BOOTP: The Bootstrap Protocol (BOOTP) is a client-server protocol designed to assign IP addresses, but it does not support dynamic configuration. When a client requests an IP address, the BOOTP server refers to a table that maps the client's physical address to an IP address. This mapping between the physical and IP addresses is static and predefined.

BOOTP is not suitable for situations where a host moves between networks or needs a temporary IP address, as the physical-IP address mapping remains fixed until an administrator makes changes. Thus, BOOTP is a static configuration protocol.

DHCP: The Dynamic Host Configuration Protocol (DHCP) has the ability to handle both static and dynamic address allocation, allowing for either manual or automatic assignment of IP addresses.

i) **Static Address Allocation:** In this mode, DHCP operates similarly to BOOTP. It is compatible with BOOTP, it allows a host with a BOOTP client to request a static IP address from a DHCP server. The DHCP server maintains a database that permanently associates physical addresses with IP addresses.

ii) **Dynamic Address Allocation:** DHCP also supports dynamic address allocation. It has a separate database containing a pool of available IP addresses. Whenever a DHCP client requests a temporary IP address, the server assigns an available address from the pool for a specified duration.

When a client sends a request to a DHCP server, the server first checks its static database. If a matching entry for the physical address is found, it provides the client's permanent IP address. If no match is found, the server picks an available IP address from the dynamic pool, assigns it to the client, and updates the dynamic database.

This dynamic capability of DHCP is particularly useful when a host moves between networks or frequently connects and disconnects, such as in the case of service provider subscribers. DHCP assigns temporary IP addresses with lease times for a certain period. Once the lease expires, the client must either stop using the address or request a renewal, which the server may accept or decline. If the renewal is denied, the client must stop using the IP address.

13.4. Delivery, forwarding and routing: Delivery refers to how a packet is managed by the networks beneath the control of the network layer. Forwarding describes how a packet is sent to the next station. Routing involves the creation of routing tables that assist in forwarding. Routing protocols are used to regularly update these tables, which are then used for both forwarding and routing decisions.

The network layer manages how packets are handled by the underlying physical networks, which we refer to as packet delivery. A packet reaches its final destination through two delivery methods: direct and indirect.

Direct Delivery: In direct delivery, the packet's final destination is a host that is on the same physical network as the sender. This occurs when both the source and destination are on the same network or when the delivery happens between the last router and the destination host.

The sender can easily determine if the delivery is direct by comparing the network address of the destination (using the mask) with the addresses of networks to which it is connected. If a match is found, the delivery is direct.

Indirect Delivery: When the destination host is not on the same network as the sender, the packet is delivered indirectly. In this case, the packet travels through routers until it reaches the router connected to the same network as the destination. Each delivery involves at least one direct delivery, but there may be multiple indirect deliveries, with the final one always being direct.

Forwarding: Forwarding refers to the process of routing a packet to its destination. It requires a host or router to have a routing table. When a host wants to send a packet or a router needs to forward a packet, it checks the routing table to find the path to the destination. However, in large internetworks like the Internet, using a simple routing table approach becomes inefficient due to the large number of entries required for lookups.

Forwarding Techniques: Several methods help manage the size of the routing table and address issues like security. Here are a couple of key techniques:

Next-Hop Method vs Route Method: One way to reduce the size of the routing table is the next-hop method. Instead of storing the entire route, this method only records the address of the next hop. This simplifies the table compared to the route method, which stores the full route information.

Network-Specific Method v/s Host-Specific Method: Another technique is the network-specific method, which reduces the routing table by grouping all hosts on the same physical network under a single entry, instead of having an

entry for each individual host. For example, instead of having 1000 separate entries for 1000 hosts on the same network, there would only be one entry for the network itself. This streamlines the searching process.

Routing: Unicast routing protocols are essential for managing how data packets are routed between a source and a specific destination in a network. These protocols focus on providing one-to-one communication, ensuring that packets are directed to a unique destination address. They form the backbone of IP-based networks, enabling efficient data transmission across various devices and segments of the internet or local networks.

Unicast routing protocols can generally be divided into two categories: Interior Gateway Protocols (IGPs) and Exterior Gateway Protocols (EGPs). IGPs are used within a single autonomous system (AS), while EGPs are designed for routing between different autonomous systems. The most commonly used IGPs are Routing Information Protocol (RIP), Open Shortest Path First (OSPF), and Enhanced Interior Gateway Routing Protocol (EIGRP), while the most widely adopted EGP is Border Gateway Protocol (BGP).

Routing Information Protocol (RIP) is a distance-vector protocol that uses hop count as its metric to determine the best path to a destination. RIP is one of the oldest routing protocols, and its simplicity makes it easy to configure, but it has limitations. The maximum hop count allowed by RIP is 15, which restricts its scalability. Moreover, RIP is known for slow convergence times, meaning that when network changes occur, it takes time for all routers to update their routing tables. Although RIP v2, an updated version of RIP, supports classless inter-domain routing

(CIDR) and provides better security, it still has limitations in large networks.

In contrast, Open Shortest Path First (OSPF) is a link-state protocol that uses Dijkstra's Shortest Path First (SPF) algorithm to compute the shortest path to each destination in a network. OSPF is designed to handle larger, more complex networks and scales better than RIP. It uses a cost metric, which is based on the bandwidth of links, and supports features like variable-length subnet masking (VLSM), which enhances IP address allocation efficiency. OSPF is also faster than RIP in terms of convergence time, meaning it can quickly adapt to network changes. One of OSPF's strengths is its hierarchical structure, where networks are divided into areas to reduce the size of routing tables and enhance performance.

Enhanced Interior Gateway Routing Protocol (EIGRP) is a hybrid protocol that combines the best features of both distance-vector and link-state protocols. It uses a complex metric that considers bandwidth, delay, load, and reliability of links, allowing it to provide more granular control over routing decisions compared to RIP. EIGRP supports rapid convergence due to its Diffusing Update Algorithm (DUAL), which ensures loop-free routing paths. EIGRP is a Cisco proprietary protocol, but it is widely used in Cisco-based networks due to its robustness, scalability, and efficient use of resources.

Border Gateway Protocol (BGP) is an inter-domain routing protocol that operates as a path-vector protocol. BGP is used to exchange routing information between different autonomous systems, making it the primary protocol for routing data across the internet. Unlike IGPs, BGP uses a plenty of attributes, such as AS Path, Next Hop, and Prefix Length, to determine the best path to

a destination. BGP is highly scalable, allowing it to handle the vast number of routes necessary for global internet routing. However, BGP has a slow convergence time compared to IGPs, which means it can take longer to adapt to network changes. BGP also supports policy-based routing, which enables administrators to control routing decisions based on various factors.

The key concepts in unicast routing revolve around the routing table, which each router maintains to determine the best path to reach a destination. These tables are updated dynamically as routers exchange routing information. For example, RIP routers send periodic updates, while OSPF and EIGRP routers exchange more complex link-state or hybrid updates. A crucial aspect of routing is convergence, the process by which all routers in a network agree on the best routes. Faster convergence minimizes the risk of routing loops and inconsistencies. Additionally, routing loops are undesirable because they lead to packets circulating indefinitely through the network. Various mechanisms, such as split horizon and poison reverse in RIP, and SPF and DUAL in OSPF and EIGRP respectively, are used to prevent these loops.

In terms of scalability, RIP's hop count limit makes it unsuitable for larger networks, while OSPF and EIGRP are designed to handle more complex topologies. OSPF's hierarchical design, for instance, helps optimize large networks by dividing them into areas. BGP, used for internet-scale routing, is the most scalable and is capable of handling the massive number of routes on the internet.

Each of these protocols has its strengths and weaknesses. RIP, though easy to configure, is limited in scalability and convergence speed. OSPF and EIGRP offer better scalability and faster convergence, with OSPF providing a

more structured approach suitable for larger and more complex networks. BGP is the preferred protocol for routing between autonomous systems, especially on the internet, due to its policy-based routing and scalability.

To summarize, unicast routing protocols are vital for the proper functioning of networks, and the choice of protocol depends on different factors such as network size, complexity, and performance requirements. While RIP is useful in small networks, protocols like OSPF and EIGRP are preferred in larger, more dynamic environments, with BGP being essential for inter-domain routing and managing the global internet.

13.5 Summary:

1. IP is a fundamental protocol operating at the network layer, designed to deliver packets across different networks. However, it lacks features like error correction and flow control.

2. Static mapping involves creating a fixed table to associate logical and physical addresses, but it is less flexible due to changing physical addresses.

3. ARP is used when a host or router knows the IP address of a destination but not the physical address (MAC address).

4. DHCP is more flexible than BOOTP, allowing both static (manual) and dynamic (temporary lease) IP address assignments.

13.6 Check your progress:

1. What is the primary function of the Internet Protocol (IP) at the network layer?
2. Why does IP lack features like flow control and error correction?
3. What is the role of the Address Resolution Protocol (ARP) in IP networking?
4. What are some scenarios where reverse mapping of physical addresses to logical addresses is necessary?
5. What are the differences between direct and indirect packet delivery in IP networking?

6. What are some techniques used in forwarding to optimize routing table sizes and improve network performance?

Module IV: Transport Layer and Application Layer
(Unit-14:Process to process delivery,connectionless and connection oriented service,TCP and UDP)

14.0. Introduction: The data link layer is responsible for sending data frames between two directly connected nodes over a communication link. This is called node-to-node delivery. The network layer, on the other hand, is in charge of sending datagrams between two hosts, which is known as host-to-host delivery. However, when we talk about actual communication on the Internet, it's not just about exchanging data between nodes or hosts. The real communication happens between processes, which are the programs running on the hosts. To achieve this, we need process-to-process delivery.

At any moment, numerous processes may be running on both the source and destination hosts. To make sure the data reaches the correct process on the destination host, a system is needed to identify and deliver the data from a specific process on the source host to the corresponding process on the destination host. The transport layer is responsible for process-to-process communication—the transmission of a packet, which is a part of a message, from one process to another.

14.1 Client/Server Paradigm: There are various ways to enable communication between processes, but the most common method is the client/server model. In this model, a process on the local machine, known as the client, requests services from a process on a remote machine, known as the server. Both processes typically share the same name. For instance, if you want to fetch the current day and time from a remote system, you would need a Daytime client process on the local machine and a Daytime server process on the remote machine.

Modern operating systems support multiuser and multiprogramming environments, meaning a remote computer can run several server programs simultaneously, just as the local machine can run multiple client programs at the same time. To facilitate this communication, we need to clearly define the following elements:

1. The local host (the machine requesting services).
2. The local process (the program or task on the local host that makes the request).
3. The remote host (the machine providing the service).
4. The remote process (the program or task on the remote machine that fulfills the request).

14.2 Addressing: When delivering something to a specific destination among many, an address is needed to identify the correct recipient. At the data link layer, this address is a MAC address, which helps identify a specific node in a network when the connection is not directly between two points. A frame at this layer requires a destination MAC address to direct it to the correct node, and a source MAC address for the reply to be sent back.

At the network layer, the address needed is an IP address, which identifies a particular host among millions of others. A datagram at this layer uses a destination IP address to reach the correct host, and a source IP address for the reply.

At the transport layer, the address is a port number, which identifies a specific process running on the destination host. The destination port number is needed to send the data to the correct process, while the source port number is used for the reply.

In the Internet model, port numbers are 16-bit integers ranging from 0 to 65,535. The client process typically selects a port number randomly, which is known as an ephemeral port number. However, the server process must use a fixed port number that is well-known, so clients can find and connect to the server. If the server randomly picked a port number, clients wouldn't know which port to contact. Although it's possible to request the port number through extra communication, this would add unnecessary complexity. To avoid this, the Internet uses a system of well-known port numbers for servers, which clients can use to connect directly. In some cases, clients can also be assigned well-known port numbers.

The Internet Assigned Number Authority (IANA) has organized port numbers into three categories: well-known, registered, and dynamic (or private). Here's how they are divided:

i) **Well-known ports**: These port numbers range from 0 to 1023 and are assigned and controlled by IANA. They are used by widely recognized services, like HTTP (port 80) and FTP (port 21), and are

considered "well-known" because they are universally recognized and managed by IANA.

ii) **Registered ports**: These port numbers range from 1024 to 49,151. These ports are not directly controlled by IANA, but they can be registered with IANA to avoid conflicts between services. This registration helps ensure that no two services use the same port in this range.

iii) **Dynamic (ephemeral) ports**: These port numbers range from 49,152 to 65,535. They are not controlled or registered by IANA, and can be used by any process for temporary or short-term communication. These ports are often chosen dynamically by applications for the duration of a session, and once the session ends, the port is available for reuse.

Socket Address: To achieve process-to-process communication, two identifiers are needed at each end of the connection: the IP address and the port number. Together, these two elements form what is called a **socket address**. The client's socket address uniquely identifies the client process, while the server's socket address uniquely identifies the server process.

In order for a transport layer protocol to establish communication, it requires both a client socket address and a server socket address. These two socket addresses allow the protocol to properly route the data to the right processes on both the client and server.

14.3.Connectionless and connection oriented Service: A transport layer protocol can be either connectionless or connection-oriented, and this distinction describes how data is transmitted between devices.

In a **connectionless** service, data (in the form of packets) is sent from the sender to the receiver without establishing a formal connection. There is no need to set up or release a connection before or after sending data. The packets are sent individually, and each packet is independent. This means-the packets may arrive in any order, or some packets may even get lost or delayed;there is no guarantee of successful delivery or acknowledgment from the receiver; the protocol does not keep track of packet numbers or the state of the communication.A good example of a connectionless protocol is **UDP (User Datagram Protocol)**, which is used in situations where speed is more important than reliability, such as real-time applications (e.g., streaming or online gaming).

In a **connection-oriented** service, a formal connection must be established between the sender and the receiver before any data is transferred. After the connection is established, data can be exchanged, and once the transfer is complete, the connection is closed. This type of service ensures-reliable data transfer with proper sequencing of packets; acknowledgment of received data and mechanisms for retransmitting lost packets; the protocol manages the connection's state and ensures data is delivered correctly. Protocols like **TCP (Transmission Control Protocol)** and **SCTP (Stream Control Transmission Protocol)** are connection-oriented, providing a more reliable and orderly transfer of data, typically used for applications where data integrity and order are crucial.

14.4 Reliable v/s Unreliable service**:** The transport layer can offer either reliable or unreliable services. If the application layer requires reliability, a reliable transport layer protocol is used, which involves implementing flow and error control, resulting in slower and more complex service. Conversely, if the application does not require

reliability—because it has its own flow and error control mechanisms, or it prioritizes speed, or the service nature (like real-time applications) does not require flow and error control—then an unreliable protocol is appropriate.

In the Internet, there are three commonly used transport layer protocols: UDP, which is connectionless and unreliable; and TCP and SCTP, which are connection-oriented and reliable. These protocols cater to the needs of different application layer programs.

A common question is whether the reliability provided by the data link layer, with its flow and error control, makes reliability at the transport layer unnecessary. The answer is yes, reliability is still needed at the transport layer. The data link layer only ensures reliability between two nodes, but the transport layer is responsible for ensuring reliability across the entire communication between two ends. This is essential because the network layer in the Internet is unreliable and provides best-effort delivery, necessitating reliability at the transport layer.

14.5 User Datagram Protocol (UDP): It is a transport layer protocol that operates without establishing a connection, which is why it's often described as connectionless. Unlike other protocols such as TCP (Transmission Control Protocol), UDP does not guarantee reliable data delivery or ensure that packets arrive in the correct order. It does not include mechanisms for retransmitting lost data, reordering packets, or even confirming receipt of data, making it unreliable.

However, UDP offers some benefits due to its simplicity:

i) Minimal overhead: It has a small header and doesn't require the extensive handshake or communication setup that more reliable protocols like TCP need.
ii) Faster communication: Because it skips the reliability checks and connection setup, UDP allows for quicker transmission of small messages.

Processes may choose UDP when speed is more important than reliability. For example, in applications like streaming video, online gaming, or voice over IP (VoIP), where it's better to lose a few packets than experience delays caused by retransmission, UDP is ideal.

UDP packets, called user datagrams, have a fixed-size header of 8 bytes. Below is the explanation of the UDP datagram format in figure 15.1.

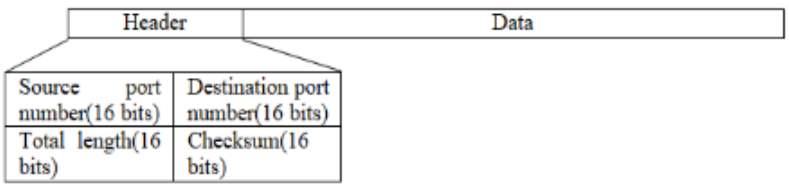


Figure 15.1: UDP datagram format

Source Port: 16-bit value representing the port number of the sender.
Destination Port: 16-bit value representing the port number of the receiver.
Length: 16-bit value indicating the total length of the UDP packet, which includes the header and the data portion.
Checksum: 16-bit value used for error detection. It ensures that the data has not been corrupted during transmission. If no checksum is used, this field is set to zero.

14.6 Transmission Control Protocol (TCP): TCP (Transmission Control Protocol) and UDP (User Datagram Protocol) are both transport layer protocols used to enable communication between programs running on different devices. While both use port numbers to identify specific applications or processes, they differ significantly in their approach to managing communication.

The key difference between TCP and UDP is that TCP is connection-oriented, while UDP is connectionless. This means that before any data is transmitted using TCP, a virtual connection is established between the sender and the receiver. This process is known as the three-way handshake, where both sides confirm the readiness to send and receive data.

TCP provides reliable delivery of data. This means that it ensures all data packets arrive at the destination in the correct order, and any lost packets are retransmitted. It also uses error control mechanisms to detect errors during transmission and requests the retransmission of corrupted data. Additionally, TCP uses flow control to manage the rate at which data is sent, preventing the receiver from being overwhelmed by too much data at once.

A packet in TCP is called a segment. The segment consists of a 20 to 60-byte header, followed by data from the application program. The header is 20 bytes if there are no options and up to 60 bytes if it contains options. Below is the explanation of the TCP segment format in figure 15.2.

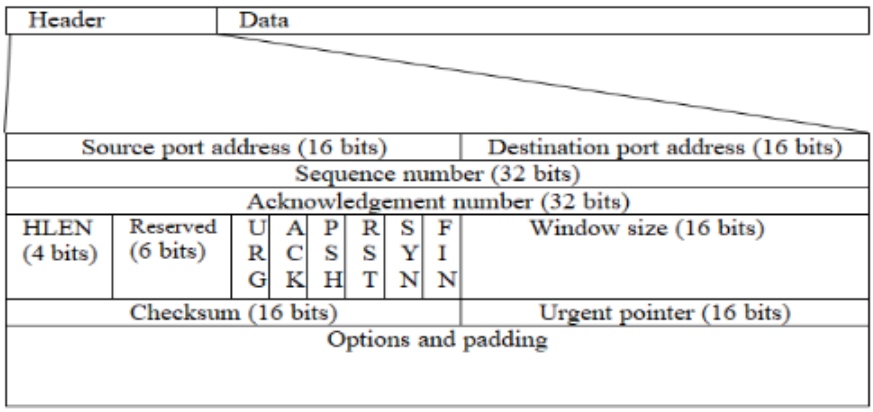


Figure 15.2: TCP segment format

Source Port and **Destination Port**: These fields are used to identify the sending and receiving applications (or processes) based on their respective port numbers.

Sequence Number: TCP uses sequence numbers to keep track of the order of bytes in a stream of data. This ensures that if packets are received out of order, they can be correctly reassembled.

Acknowledgment Number: When a receiver successfully receives data, it sends an acknowledgment (ACK) with the acknowledgment number indicating the next expected byte. This helps ensure reliable delivery.

Data Offset: This tells how many 32-bit words the TCP header is, which is important for finding where the data begins within the segment.

Flags: These are used for various control purposes:

- **SYN**: Initiates a connection.
- **ACK**:Acknowledges receipt of data.
- **FIN**: Indicates the end of a connection.

- **RST**: Resets the connection.
- **PSH**: Pushes data to the application immediately.
- **URG**: Indicates urgent data.

Window Size: Specifies how much data the receiver is willing to accept, which helps in flow control.

Checksum: A mechanism for detecting errors in the data transmission, ensuring that the received data is correct.

Urgent Pointer: When the URG flag is set, this field indicates the location of urgent data, which is processed before other data.

Options: Optional fields that can provide additional features such as the maximum size of segments or timestamps for round-trip time measurement.

Data (Payload): The actual data being transferred, following the header. The length of the data is determined by the remaining space in the segment after the header.

14.7 3-way handshaking: The TCP three-way handshake is the process used to establish a reliable connection between a client and a server before data transfer begins. This process ensures that both parties are ready for communication and that they can synchronize their sequence numbers to maintain the order of the data transmitted.

The process starts when the client, which wants to initiate the connection, sends a **SYN** (synchronize) message to the server. This message includes an initial sequence number, which is randomly chosen by the client. This step indicates the client's desire to establish a

connection and synchronize its sequence number with the server.

In response, the server acknowledges the client's request by sending a **SYN-ACK** (synchronize-acknowledge) message. This response serves two purposes: first, it acknowledges the receipt of the client's SYN message, and second, it includes the server's own initial sequence number, signaling its readiness to establish the connection. The server's sequence number is also chosen randomly.

Finally, the client sends an **ACK** (acknowledge) message back to the server. This message confirms the receipt of the server's SYN-ACK and completes the handshake process. The client includes the next expected sequence number based on the server's response.

At this point, both the client and server have exchanged sequence numbers and confirmed their readiness to send and receive data, establishing a reliable communication channel. This three-way handshake ensures that both sides are synchronized before any actual data is transmitted, enabling the reliable, ordered delivery of packets in TCP communication.

14.8 Summary:

i) Every application program is assigned a port number to differentiate it from other programs running on the same machine at the same time.

ii) The client program is assigned a temporary port number, referred to as an ephemeral port number, while

the server program is assigned a fixed port number known as a well-known port number.

iii) The ICANN has defined specific ranges for different types of port numbers.
iv) A combination of the IP address and port number, known as a socket address, identifies both a process and a host.
v) UDP is a connectionless and unreliable transport layer protocol, lacking built-in flow or error control mechanisms, except for the checksum used for error detection.
vi) TCP provides process-to-process communication, supporting full-duplex and connection-oriented services.

14.9 Check your progress:

1. What is the main responsibility of the transport layer in network communication?
2. What is the difference between node-to-node delivery and host-to-host delivery in the network?
3. What is the purpose of an IP address, and how does it differ from a MAC address at the network and data link layers?
4. Describe the three categories of port numbers defined by IANA.
5. What is a socket address, and how does it help in the communication between processes?
6. Describe the structure of a UDP packet and explain the purpose of each field in the UDP datagram.

7. How does the TCP three-way handshake process work, and what role does it play in establishing a reliable connection?
8. What is the difference between TCP and UDP in terms of reliability, connection establishment, and flow control?

Module IV: Transport Layer and Application Layer
(Unit-15: DNS,SMTP,POP, IMAP, Cryptography)

15.0 Introduction: The Domain Name System (DNS) is a fundamental component of the internet's architecture, enabling users to access websites and services using human-readable domain names instead of numeric IP addresses. DNS acts as the "phonebook" of the internet, translating domain names such as www.example.com into IP addresses, which are the addresses that computers use to identify each other on the network.

When you enter a domain name into your web browser, such as www.example.com, the DNS system translates this name into an IP address, allowing the browser to locate the appropriate web server. The DNS consists of several components that work together to provide this translation.

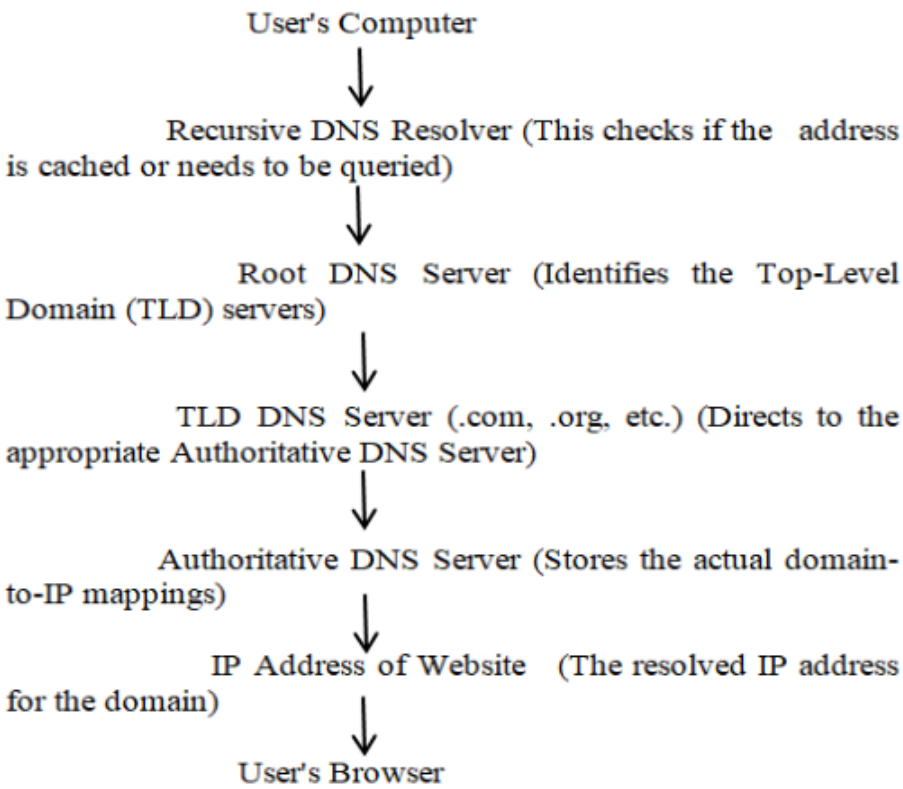
15.1 Components of DNS: The first component of DNS is the Domain Name, which consists of several levels. The highest level is the Top-Level Domain (TLD), such as .com, .org, or .edu. Below the TLD is the Second-Level Domain (SLD), like example in example.com. Sometimes, there may also be a Third-Level Domain, also known as a subdomain (e.g., www in www.example.com). These domains are registered and managed through domain registrars.

When a user requests a website, the request is sent to a Recursive Resolver. The recursive resolver is a DNS server that is responsible for querying other DNS servers to find the correct IP address. The recursive resolver first checks if it has the IP address cached. If it does not, it initiates a process called a DNS Lookup.

The DNS lookup involves several steps. First, the recursive resolver queries the Root DNS Servers, which know where the authoritative servers for each TLD are located. The root servers provide a referral to the TLD DNS Servers responsible for the .com domain (or any other relevant domain, depending on the TLD). Next, the TLD DNS servers direct the request to the Authoritative DNS Servers, which store the actual mapping of domain names to IP addresses. Once the recursive resolver receives the correct IP address from the authoritative DNS server, it returns this information to the user's browser. The browser can then make the connection to the web server using the IP address, allowing the website to load.

The DNS system is hierarchical and decentralized. The authoritative DNS servers for each domain are managed by different organizations and entities, ensuring redundancy and resilience across the network. The cache mechanism within DNS servers helps reduce latency and improve speed by storing recent query results.

To better understand how the DNS works, we can represent the process visually as given below:



DNS ensures the seamless operation of the internet by converting human-friendly domain names into machine-readable IP addresses. Through a series of interconnected servers and caching mechanisms, it enables fast, reliable, and efficient access to websites across the globe. The DNS system is crucial to the everyday functioning of the internet, making it possible for users to interact with web services using easy-to-remember names rather than numerical IP addresses.

15.2 Electronic mail: Electronic mail, or e-mail, has become one of the most widely used services on the Internet. When the creators of the Internet originally designed it, they likely didn't foresee how essential and popular e-mail would become. E-mail allows people to send and receive messages electronically, making communication faster and more efficient compared to traditional mail systems.

The architecture of e-mail involves several key components that work together to ensure the system functions smoothly. These components include: message transfer agent and message access agent.

A Message Transfer Agent (MTA) is responsible for transferring emails between systems over the Internet. The communication process between MTAs is governed by a protocol called Simple Mail Transfer Protocol (SMTP). SMTP is used twice: once between the sender and the sender's mail server, and again between the two mail servers, making it a PUSH protocol.

SMTP is used in the first two stages of mail delivery. However, SMTP is not involved in the third stage because it is a push protocol, meaning it pushes the message from the client to the server. The third stage uses a message access agent, with two main protocols available for this: Post Office Protocol version 3 (POP3) and Internet Mail Access Protocol version 4 (IMAP4). Figure 16.1 illustrates where these two protocols fit into the typical email delivery process alongside SMTP.

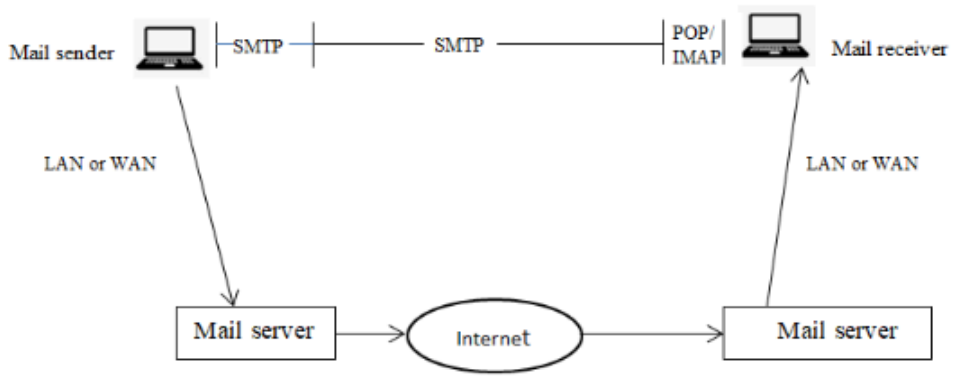


Figure 16.1: E-mail system

15.3: Cryptography: Cryptography involves the study and application of techniques to ensure secure communication, even when there are third parties, known as adversaries, involved. More broadly, cryptography focuses on creating and evaluating protocols that prevent unauthorized individuals or the public from accessing private messages. Key aspects of information security, including data confidentiality, integrity, authentication, and non-repudiation, are fundamental to modern cryptography. Cryptography is applied in areas such as online commerce, chip-based payment systems, digital currencies, computer passwords, and military communications.

The pictorial representation of cryptography is shown in figure 16.2.

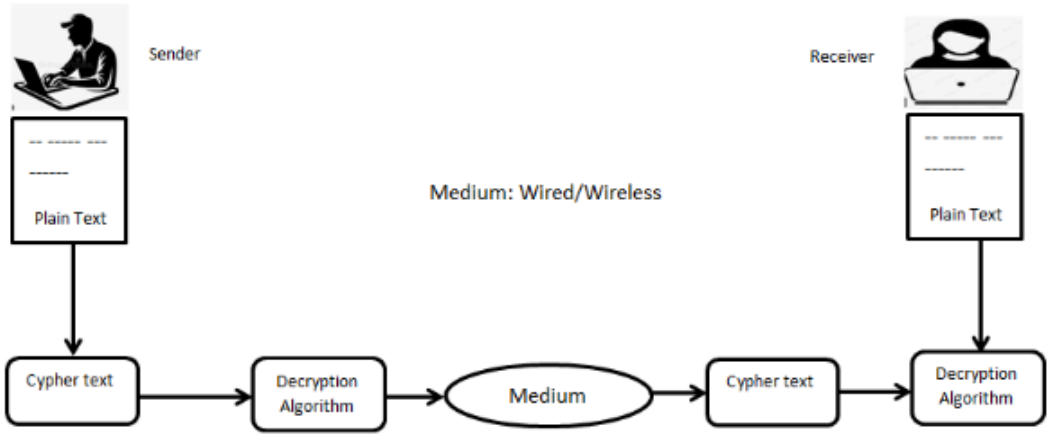


Figure 16.2: Cryptography

Cryptography is of two types: symmetric-key and asymmetric-key cryptography.

15.4: Symmetric-key cryptography: Symmetric-key cryptography involves encryption techniques where both the sender and receiver use the same key, or in some cases, keys that are different but mathematically related in a straightforward manner. This is illustrated in figure 16.3..

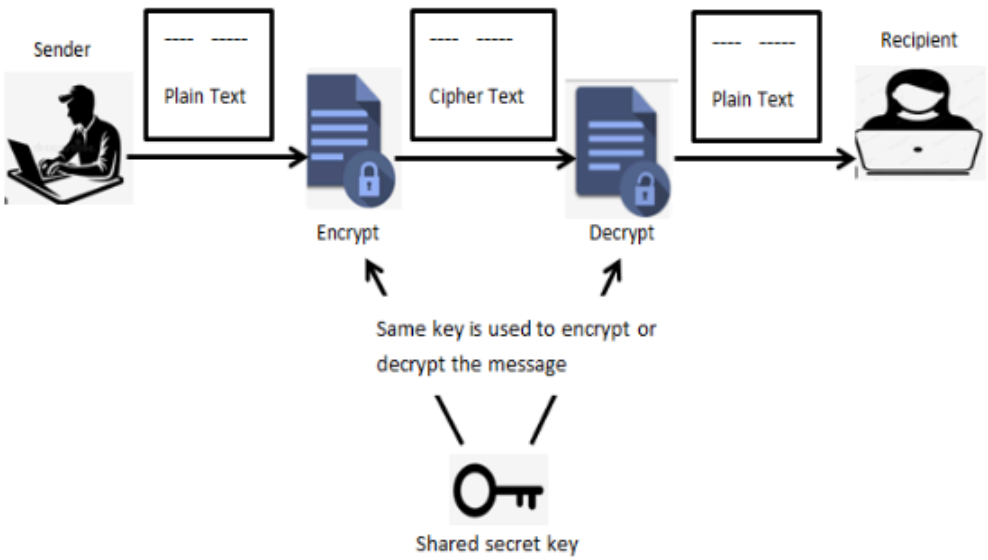


Figure 16.3: Symmetric-key cryptography

The most popular algorithm that was developed using symmetric-key cryptography was Data Encryption Standard (DES). It has two variants: triple-DES-variant and Advanced Encryption Standard(AES). But in the long run, the idea of symmetric–key cryptography was replaced by asymmetric–key crtptography. The basic idea behind the working of DES algorithm is shown in figure 16.4.

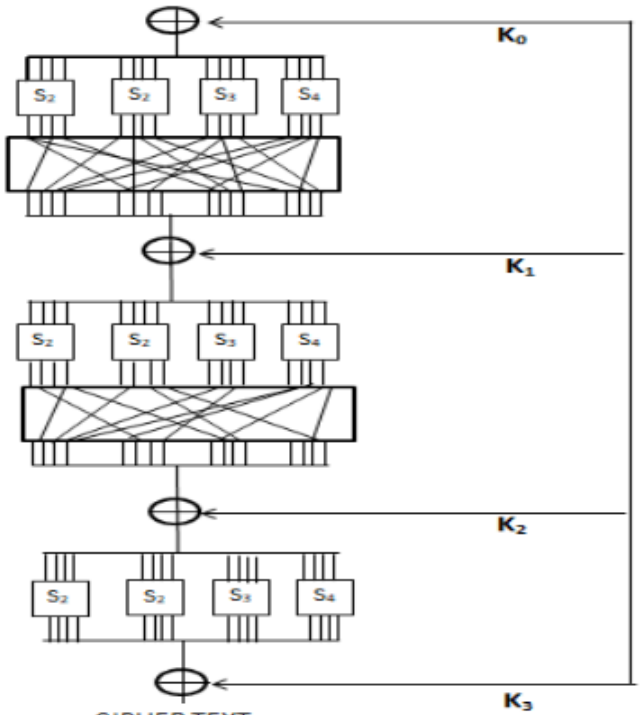


Figure 16.4: Data encryption standard

The DES algorithm is a block cipher that used shared-secret key and some important points on DES are given below:

i) It is basically a mono-alphabetic substitution cipher using a 64 bit character.

ii) Whenever the same 64 bit plaintext block goes in, the same 64 bit cipher-text block comes out.

Working of DES involves the following stages:

i) The first stage is a key independent transposition on the 64 bit plain-text.

ii) The last stage is the exact inverse, before that is an exchange of the leftmost with the rightmost 32 bits.

iii) The remaining 16 stages are functionally identical but are parameterized by different functions of the key.

iv) The left output of an iteration stage is simply a copy of the right input.The right output is the exclusive OR of the left input and a function of the right input and the key for this iteration. All the complexity lies in this function which consists of four sequential steps.

15.5: Asymmetric-key cryptography: Symmetric-key cryptosystems utilize the same key for both encrypting and decrypting a message, although different keys can be used for different messages or groups of messages. A major drawback of symmetric ciphers is the key management required for secure use. Ideally, each pair of communicating parties must share a unique key, and possibly a different key for each cipher text exchanged. The number of keys needed grows as the square of the number of network participants, which rapidly leads to the need for complex key management systems to maintain consistency and secrecy.

In asymmetric-key cryptography, each of sender and receiver has a different pair of keys. On key is

called public key which is openly distributed over the internet and another key is kept as secret.This is shown in figure 16.5.

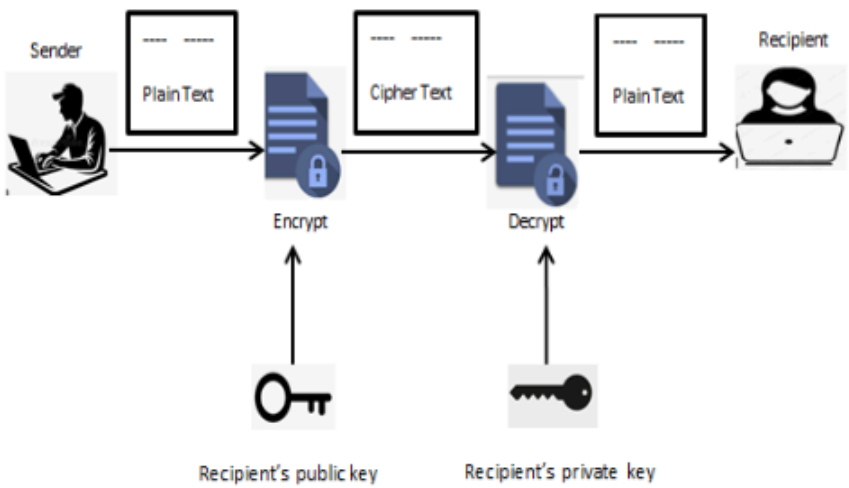


Figure 16.5: Asymmetric-key cryptography

Here, encryption is done with receiver's public key and decrypted by receiver's private key. So, only the receiver can read the message (there is only confidentiality, but no authenticity). The encryption by sender's private key and decryption by sender's public key is also allowed in asymmetric-key cryptography (there is only authenticity, but no confidentiality). This thought process was implemented by three people- Ronald Rivest, Adi Shamir and Len Adlemen in 1978 and the algorithm made was known as RSA algorithm.RSA algorithm is the most popular algorithm till date.

15.6 RSA algorithm: RSA (Rivest-**S**hamir-Adleman) Algorithm is an asymmetric or public-key cryptography algorithm which means it works on two different keys: Public Key and Private Key. The Public Key is used for encryption and is known to everyone, while the Private Key is used for decryption and must be kept secret by the receiver. RSA Algorithm is named after Ron Rivest, Adi **S**hamir and Leonard Adleman, who published the algorithm in 1977.

RSA Algorithm is based on factorization of large number and modular arithmetic for encrypting and decrypting data. It consists of three main stages:

i) Key Generation: Creating Public and Private Keys.
ii) Encryption: Sender encrypts the data using Public Key to get cipher text.
iii) Decryption: Decrypting the cipher text using Private Key to get the original data

Key Generation:

i) Choose two large prime numbers, say p and q. These prime numbers should be kept secret.
ii) Calculate the product of primes, n = p * q. This product is part of the public as well as the private key.
iii) Calculate Euler Totient Function $\Phi(n)$ as $\Phi(n) = \Phi(p * q) = \Phi(p) * \Phi(q) = (p - 1) * (q - 1)$.
iv) Choose encryption exponent e, such that-

a) $1 < e < \Phi(n)$, and
b) gcd $(e, \Phi(n)) = 1$, that is e should be co-prime with $\Phi(n)$.

v) Calculate decryption exponent d, such that
a) $(d * e) \equiv 1 \bmod \Phi(n)$, that is d is modular multiplicative inverse of e mod $\Phi(n)$. Some common methods to calculate multiplicative inverse are: Extended Euclidean Algorithm, Fermat's Little Theorem, etc.
b) We can have multiple values of d satisfying $(d * e) \equiv 1 \bmod \Phi(n)$ but it does not matter which value we choose as all of them are valid keys and will result into same message on decryption.

Finally, the Public Key = (n, e) and the Private Key = (n, d).

Encryption:

To encrypt a message M, it is first converted to numerical representation using ASCII and other encoding schemes. Now, use the public key (n, e) to encrypt the message and get the cipher text using the formula:

$$C = M^e \bmod n,$$

- Where C is the Cipher text and e and n are parts of public key.

Decryption:

To decrypt the cipher text C, use the private key (n, d) and get the original data using the formula:

$$M = C^d \bmod n,$$

-where M is the message and d and n are parts of private key.

15.7: Digital Signature: A digital signature is a cryptographic technique used to verify the authenticity and integrity of digital messages or documents. It is the electronic equivalent of a handwritten signature or a stamped seal, but much more secure. Whenever, the sender encrypts a message with its own private key,it is known as digital signature. A digital signature serves three main purposes:

i) Authentication: Confirms the identity of the sender.
ii) Integrity: Ensures the message or document has not been altered during transmission.
iii) Non-repudiation: Prevents the sender from denying the authenticity of the message.

One important point about digital signature is that it does not provide confidentiality.

15.8: Digital Certificate: A digital certificate is a cryptographic tool used to prove the ownership of a public key. It binds the identity of an entity (like a person, organization, or device) to its public key through a process called public key infrastructure (PKI). Digital certificates are used to facilitate secure communication, establish trust between parties, and authenticate the identity of the certificate holder in online transactions.

A digital certificate typically includes:

i) The public key of the entity.

ii) Information about the entity (e.g., name, email address, company).
iii) The Certificate Authority (CA) that issued the certificate.
iv) A digital signature from the CA to confirm its authenticity.

A certificate authority (CA) is a trusted organization that issues digital certificates to verify the identity of websites, email addresses, companies or individuals.

For example, let us consider a scenario where Mr. Bob needs a digital certificate to prove the authenticity of his own website. For this he has to approach a CA. The CA(which is a trusted organization) will first apply hash function on the collective personal informations of Bob such as Bob's ID, Bob's public key, CA's information and date up to which the certificate will be valid. After the application of hash function, it becomes message digest (MD). The CA will then encrypt the MD with CA's public key to develop the digital signature. Now, the CA will hand over the digital signature to Bob. Bob will send signature along with his informations to anyone whio visits his website. The combination of digital signature and Bob's unencrypted information is called the digital certificate.

Suppose, Alice is another person who is browsing Bob's website. Now, Bob's server will send the digital certificate to Alice. Alice will first extract the information and see the CA's name. She will separate the digital signature and decrypt with the CA's public key (Most of the CA's public keys are usually installed in the popular web-browsers) to get the MD. Now, she will apply the same hash function (previously applied by the CA) to

the information to make a comparison of the two MDs. If both MDs are same, then CA's information is correct and it prove the authenticity of Bob's website.

Now, Alice will take the Bob's public key and send the secret key using asymmetric algorithm (ex. RSA). Bob will get the secret key and now actual data communication between Bob and Alice can happen using the shared secret key, (i.e, using symmetric-key cryptography).

15.9: Summary:

i) DNS acts as the "phonebook" of the internet, translating domain names such as into IP addresses.
ii) SMTP is a push protocol and POP/IMAP are pull protocols.
iii) Cryptography is the practice and study of techniques for secure communication in the presence of third parties called adversaries.
iv) DSA is a symmetric-key algorithm which involves iteration of 16 times.
v) RSA is the most popular asymmetric-key algorithm till date.
vi) In asymmetric key cryptography the public key and private keys can be used interchangeably depending on the requirement (i.e, whether authentication is crucial or confidentiality is crucial).

15.10 Check your progress:

1. Which of the following is the primary function of DNS?

a) To encrypt messages
b) To translate domain names into IP addresses
c) To send e-mails
d) To generate encryption keys

2. What does the Recursive Resolver do in the DNS process?

a) It stores domain name information
b) It queries other DNS servers to find the IP address
c) It encrypts the data
d) It sends emails to recipients

3. Which encryption algorithm is based on symmetric-key cryptography?

a) RSA
b) DES
c) AES
d) Both b and c

4. Fill in the blank :

In the DES algorithm, the encryption method is based on a _____ cipher.

5. What are the benefits and challenges of using symmetric-key cryptography in a large network? How does asymmetric-key cryptography address some of these challenges?

6. What are the main purposes of a digital signature, and how does it ensure message integrity, authenticity, and non-repudiation?
