Maharaja Education Trust (R), Mysuru

Maharaja Institute of Technology Mysore

Belawadi, Sriranga Pattana Taluk, Mandya – 571 477

Approved by AICTE, New Delhi,
Affiliated to VTU, Belagavi & Recognized by Government of Karnataka

Lecture Notes on

DATA COMMUNICATION (18CS46)

Prepared by

Department of Computer Science & Engineering

Vision/ ಆಶಯ

"To be recognized as a premier technical and management institution promoting extensive education fostering research, innovation and entrepreneurial attitude"

ಸಂಶೋಧನೆ, ಆವಿಷ್ಕಾರ ಹಾಗೂ ಉದ್ಯಮಶೀಲತೆಯನ್ನು ಉತ್ತೇಜಿಸುವ ಅಗ್ರಮಾನ್ಯ ತಾಂತ್ರಿಕ ಮತ್ತು ಆಡಳಿತ ವಿಜ್ಞಾನ ಶಿಕ್ಷಣ ಕೇಂದ್ರವಾಗಿ ಗುರುತಿಸಿಕೊಳ್ಳುವುದು.

Mission/ ಧ್ಯೇಯ

- To empower students with indispensable knowledge through dedicated teaching and collaborative learning.

  ಸಮರ್ಪಣಾ ಮನೋಭಾವದ ಬೋಧನೆ ಹಾಗೂ ಸಹಭಾಗಿತ್ವದ ಕಲಿಕಾಕ್ರಮಗಳಿಂದ ವಿದ್ಯಾರ್ಥಿಗಳನ್ನು ಅತ್ಯತ್ಕೃಷ್ಟ ಜ್ಞಾನಸಂಪನ್ನರಾಗಿಸುವುದು.

- To advance extensive research in science, engineering and management disciplines.

  ವೈಜ್ಞಾನಿಕ, ತಾಂತ್ರಿಕ ಹಾಗೂ ಆಡಳಿತ ವಿಜ್ಞಾನ ವಿಭಾಗಗಳಲ್ಲಿ ವಿಸ್ತೃತ ಸಂಶೋಧನೆಗಳೊಡನೆ ಬೆಳವಣಿಗೆ ಹೊಂದುವುದು.

- To facilitate entrepreneurial skills through effective institute - industry collaboration and interaction with alumni.

  ಉದ್ಯಮ ಕ್ಷೇತಗಳೊಡನೆ ಸಹಯೋಗ, ಸಂಸ್ಥೆಯ ಹಿರಿಯ ವಿದ್ಯಾರ್ಥಿಗಳೊಂದಿಗೆ ನಿರಂತರ ಸಂವಹನಗಳಿಂದ ವಿದ್ಯಾರ್ಥಿಗಳಿಗೆ ಉದ್ಯಮಶೀಲತೆಯ ಕೌಶಲ್ಯ ಪಡೆಯಲು ನೆರವಾಗುವುದು.

- To instill the need to uphold ethics in every aspect.

  ಜೀವನದಲ್ಲಿ ನೈತಿಕ ಮೌಲ್ಯಗಳನ್ನು ಅಳವಡಿಸಿಕೊಳ್ಳುವುದರ ಮಹತ್ವದ ಕುರಿತು ಅರಿವು ಮೂಡಿಸುವುದು.

- To mold holistic individuals capable of contributing to the advancement of the society.

  ಸಮಾಜದ ಬೆಳವಣಿಗೆಗೆ ಗಣನೀಯ ಕೊಡುಗೆ ನೀಡಬಲ್ಲ ಪರಿಪೂರ್ಣ ವ್ಯಕ್ತಿತ್ವವುಳ್ಳ ಸಮರ್ಥ ನಾಗರೀಕರನ್ನು ರೂಪಿಸುವುದು.

VISION/ ಆಶಯ

"To be a leading academic department offering computer science and engineering education, fulfilling industrial and societal needs effectively."

"ಕೈಗಾರಿಕಾ ಮತ್ತು ಸಾಮಾಜಿಕ ಅಗತ್ಯಗಳನ್ನು ಪರಿಣಾಮಕಾರಿಯಾಗಿ ಪೂರೈಸುವ ಮೂಲಕ ಕಂಪ್ಯೂಟರ್ ವಿಜ್ಞಾನ ಮತ್ತು ಎಂಜಿನಿಯರಿಂಗ್ ಶಿಕ್ಷಣವನ್ನು ನೀಡುವ ಪ್ರಮುಖ ಶೈಕ್ಷಣಿಕ ವಿಭಾಗವಾಗುವುದು."

MISSION/ ಧ್ಯೇಯ

- To enrich the technical knowledge of students in diversified areas of Computer science and engineering by adopting outcome based approaches.

  ಫಲಿತಾಂಶ ಆಧಾರಿತ ವಿಧಾನಗಳನ್ನು ಅಳವಡಿಸಿಕೊಳ್ಳುವ ಮೂಲಕ ಕಂಪ್ಯೂಟರ್ ವಿಜ್ಞಾನ ಮತ್ತು ಎಂಜಿನಿಯರಿಂಗ್‌ನ ವೈವಿಧ್ಯಮಯ ಕ್ಷೇತ್ರಗಳಲ್ಲಿನ ವಿದ್ಯಾರ್ಥಿಗಳ ತಾಂತ್ರಿಕ ಜ್ಞಾನವನ್ನು ಅಭಿವೃದ್ಧಿ ಪಡಿಸುವುದು.

- To empower students to be competent professionals maintaining ethicality.

  ನೈತಿಕತೆಯನ್ನು ಕಾಪಾಡುವ ಸಮರ್ಥ ವೃತ್ತಿಪರರಾಗಿ ವಿದ್ಯಾರ್ಥಿಗಳನ್ನು ಸಶಕ್ತಗೊಳಿಸುವುದು.

- To facilitate the development of academia-industry collaboration.

  ಶೈಕ್ಷಣಿಕ-ಉದ್ಯಮ ಸಹಯೋಗದ ಅಭಿವೃದ್ಧಿಗೆ ಅನುಕೂಲವಾಗುವಂತೆ.

- To create awareness of entrepreneurship opportunities.

  ಉದ್ಯಮಶೀಲತೆ ಅವಕಾಶಗಳ ಬಗ್ಗೆ ಜಾಗೃತಿ ಮೂಡಿಸುವುದು.

Program Outcomes

1. **Engineering knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
2. **Problem analysis**: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
3. **Design/development of solutions**: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
4. **Conduct investigations of complex problems**: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
5. **Modern tool usage**: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
6. **The engineer and society**: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
7. **Environment and sustainability**: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
8. **Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
9. **Individual and team work**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
10. **Communication**: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
11. **Project management and finance**: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
12. **Life-long learning**: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Maharaja Institute of Technology Mysore

Department of Computer Science & Engineering

Course Overview

Subject: Data Communication **Subject Code : 18CS46**

Data communication and networking have changed the way we do business and the way we live. Business decisions have to be made ever more quickly and the decision makers require immediate access to accurate information. Data communication and networking have also found much application in political and social issues. The word data refers to the information presented in whatever form is agreed upon by the parties creating and using the data. Data communication refers to the exchange of data between a source and a receiver via form of transmission media such as a wire cable. The main emphasis of this course is on the organization and management of local area networks (LANs). The course objectives include learning about computer network organization and implementation, obtaining a theoretical understanding of data communication and computer networks. The course introduces computer communication network design and its operations. Students are able to understand the basics of data communication and define their components and the types of data exchanged. Students are also able to apply appropriate methods and protocol suites to address the different data communication issues which help them to analyze and evaluate the significance of data network components and the functionalities of various layer protocol and network devices.

Course Objectives

The objectives of this course is to make students to learn-

- Comprehend the transmission technique of digital data between two or more computers and a computer network that allows computers to exchange data.
- Explain with the basics of data communication and various types of computer networks;
- Demonstrate Medium Access Control protocols for reliable and noisy channels.
- Expose wireless and wired LANs.

Course Outcomes

COs	Description
C246.1	**Illustrate** the various components of data communication and the layers of OSI and TCP/IP model
C246.2	**Discuss** the fundamentals of digital communication and switching.
C246.3	**Analyze** different data link layer protocols.
C246.4	**Summarize** wired/Wireless LANs and other Wireless Networks
C246.5	**Demonstrate** IPV4 addressing and sub-netting mechanisms.

Faculty	**Faculty**	**Course Coordinator**

Facilitator	**NBA Coordinator**	**HOD**

Maharaja Institute of Technology Mysore

Department of Computer Science & Engineering

Syllabus

Subject: DATA COMMUNICATION **Subject Code:18CS46**

Module 1

Introduction: Data Communications, Networks, Network Types, Internet History, Standards and Administration, Networks Models: Protocol Layering, TCP/IP Protocol suite, The OSI model, Introduction to Physical Layer-1: Data and Signals, Digital Signals, Transmission Impairment, Data Rate limits, Performance.

Module 2

Digital Transmission: Digital to digital conversion (Only Line coding: Polar, Bipolar and Manchester coding). Physical Layer-2: Analog to digital conversion (only PCM), Transmission Modes, Analog Transmission: Digital to analog conversion.

Module 3

Bandwidth Utilization: Multiplexing and Spread Spectrum, Switching: Introduction, Circuit Switched Networks and Packet switching. Error Detection and Correction: Introduction, Block coding, Cyclic codes, Checksum

Module 4

Data link control: DLC services, Data link layer protocols, Point to Point protocol (Framing, Transition phases only). Media Access control: Random Access, Controlled Access and Channelization, Introduction to Data-Link Layer: Introduction, Link-Layer Addressing, ARP IPv4 Addressing and subnetting: Classful and CIDR addressing, DHCP, NAT

Module 5

Wired LANs Ethernet: Ethernet Protocol, Standard Ethernet, Fast Ethernet, Gigabit Ethernet and 10 Gigabit Ethernet, Wireless LANs: Introduction, IEEE 802.11 Project and Bluetooth. Other wireless Networks: Cellular Telephony

Textbooks:

1. Behrouz A. Forouzan, Data Communications and Networking 5E, 5 th Edition, Tata McGraw-Hill, 2013.

Reference Books:

1. Alberto Leon-Garcia and Indra Widjaja: Communication Networks - Fundamental Concepts and Key architectures, 2nd Edition Tata McGraw-Hill, 2004.
2. William Stallings: Data and Computer Communication, 8th Edition, Pearson Education, 2007.
3. Larry L. Peterson and Bruce S. Davie: Computer Networks – A Systems Approach, 4th Edition, Elsevier, 2007.
4. Nader F. Mir: Computer and Communication Networks, Pearson Education, 2007.

Index

Subject:Data Communication **Subject Code:18CS46**

Module 1

1.1 Introduction to Data Communication

Man has learnt to convey information to others from ancient ages through many ways. He has used symbols, different languages to convey data or information to others. His body langauge covey a lot to others. A traffic police gives signal without speaking. He just say who should stop and who should move through his hands. Communicating with others has been found an essential task for a man from ages. Sharing information is what is known as Communication. Data flows from those who have it to those who dont have it. Like a waterflow. When two buddhas meet each other, they dont communicate. They dont have to speak because whatever one has, even second has that. Symbolic communication is found in wars between soldiers to exchange certain information. A language is widely used to communicate in most of the cases to share knowledge, information, data etc. Sharing can be local or remote. Between individuals, local communication usually occurs face to face, while remote communication takes place over distance. Figure 1.1 gives the componenets essential for any communication system.

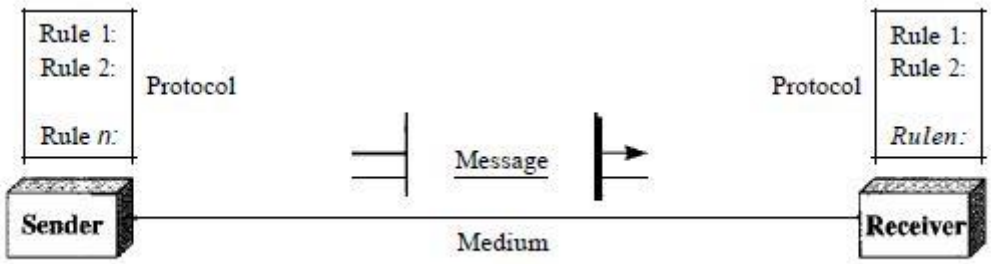


Fig 1.1 : Communication System and its Components

i. **Sender** : is a person or station or telephone handset which has data to share.

ii. **Receiver** : is a person or station or telephone handset that aqctually needs and receive data.

iii. **Medium** : The transmission medium is the physical path by which a message travels from sender to receiver. Some examples of transmission media include twisted-pair wire, coaxial cable, fiber-optic cable, and radio waves through which sender sends data. When we speak to someone next to us, air is the channel. Water is not carrying our data here. When we speaks in phone, its cable or channels like optical fibre or space or accoustic channels under ocean that carries our data in the

form of a signal.

iv. **Transmitter** : If only sender with knowledge is present but he or she is dumb, then we will not call this as a good communication system. Along with source of information, sender should have mouth to speak. A station must possess a transmitter through wich data goes into medium.

v. **Recieving antenna** : is very much essential to catch the data from medium. If a person is speaking infront of deaf, its not a communication as recieving antenna or ears here are missing.

vi. **Data** : is the information or knowledge that is to be sent from sender to receiver. Popular forms of information include text, numbers, pictures, audio, and video.

vii. **Protocol** : Set of rules to be followed in a communication system. If reciever understands only english and if sender is speaking in chinese, its not a communication. Rules to be followed by everyone and thats what is called as Protocol.Protocol represents an agreement between the communicating devices. Without a protocol, two devices may be connected but not communicating, just as a person speaking english cannot be understood by a person who speaks only chinese.

1.2 Data Representation

Information today comes in different forms such as text, numbers, images, audio, and video.

i. Text:

In data communications, text is represented as a bit pattern, a sequence of bits (O or 1). Different sets of bit patterns have been designed to represent text symbols. Each set is called a code, and the process of representing symbols is called coding. Today, the prevalent coding system is called Unicode, which uses 32 bits to represent a symbol or character used in any language in the world. The American Standard Code for Information Interchange (ASCII), developed some decades ago in the United States, now constitutes the first 127 characters in Unicode and is also referred to as Basic Latin.

ii. Numbers:

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

iii. Images:

Images are also represented by bit patterns. In its simplest form, an image is composed of a matrix of pixels (picture elements), where each pixel is a small dot. The size of the pixel depends on the resolution. For example, an image can be divided into 1000 pixels or 10,000 pixels. In the second case, there is a better representation of the image (better resolution), but more memory is needed to store the image. After an image is divided into pixels, each pixel is assigned a bit pattern. The size and the value of the pattern depend on the image. For an image made of only blackand- white dots (e.g., a chessboard), a 1-bit pattern is enough to represent a pixel. If an image is not made of pure white and pure black pixels, you can increase the size of the bit pattern to include gray scale. For example, to show four levels of gray scale, you can use 2-bit patterns. A black pixel can be represented by 00, a dark gray pixel by 01, a light gray pixel by 10, and a white pixel by 11. There are several methods to represent color images. One method is called RGB, so called because each color is made of a combination of three primary colors: red, green and blue. The intensity of each color is measured, and a bit pattern is assigned to it. Another method is called YCM, in which a color is made of a combination of three other primary colors: yellow, cyan, and magenta.

iv. Audio:

Audio refers to the recording or broadcasting of sound or music. Audio is by nature different from text, numbers, or images. It is continuous, not discrete. Even when we use a microphone to change voice or music to an electric signal, we create a continuous signal.

v. Video:

Video refers to the recording or broadcasting of a picture or movie. Video can either be produced as a continuous entity (e.g., by a TV camera), or it can be a combination of images, each a discrete entity, arranged to convey the idea of motion. Again we can change video to a digital or an analog signal.

1.3 Data Flow

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

i. Simplex:

In simplex mode, the communication is unidirectional, as on a one-way street. Only one of the two devices on a link can transmit; the other can only receive (see Figure 1.2a). Keyboards and traditional monitors are examples of simplex devices. The keyboard can only introduce input; the monitor can only accept output. The simplex mode can use the entire capacity of the channel to send data in one direction.

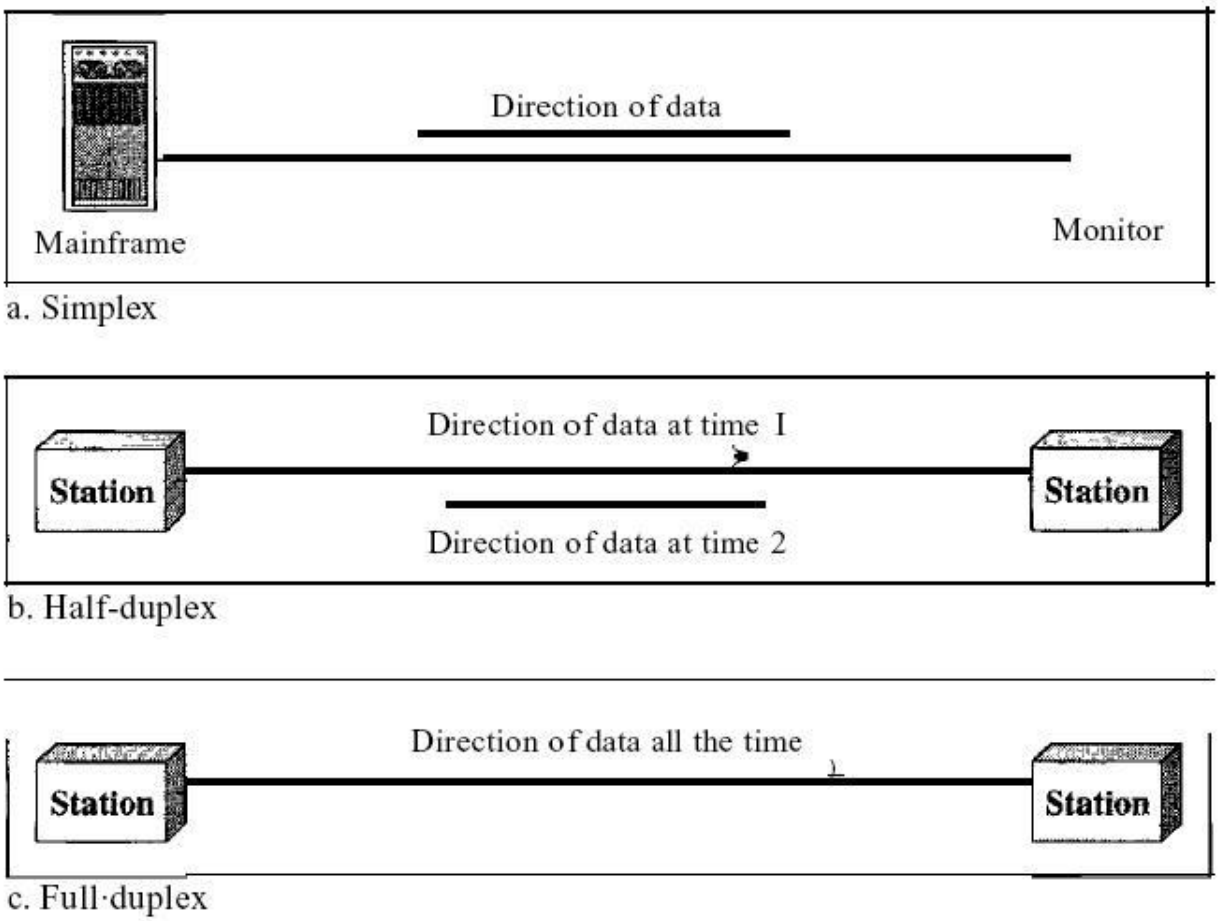


Figure 1.2 : Types of Data Flow

ii. Half-Duplex:

In half-duplex mode, each station can both transmit and receive, but not at the same time (Figure 1.2b). When one device is sending, the other can only receive, and vice versa. The half-duplex mode is like a one-lane road with traffic allowed in both directions. When cars are traveling in one direction, cars going the other way must wait. In a half-duplex transmission, the entire capacity of a channel is taken over by whichever of the two devices is transmitting at the time. Walkie-talkies and CB (citizens band) radios are both half-duplex systems. The half-duplex mode is used in cases where there is no need for communication in both directions at the same time; the entire capacity of the channel can be utilized for each direction.

iii. Full-Duplex:

In full-duplex, both stations can transmit and receive simultaneously (see Figure 1.2c). The full-duplex mode is like a two-way street with traffic flowing in both directions at the same time. In full-duplex mode, signals going in one direction share the capacity of the link: with signals going in the other direction. This sharing can occur in two ways: Either the link must contain two physically separate tranmsmission paths, one for sending and the other for receiving; or the capacity of the channel is divided between signals traveling in both directions. One common example of full-duplex communication is the telephone network. When two people are communicating by a telephone line, both can talk and listen at the same time. The full-duplex mode is used when communication in both directions is required all the time. The capacity of the channel, however, must be divided between the two directions.

1.4 Introduction to Networks

Generally, network means an association. A network of all advocates in your city. It is formed to share certain resources. A Computer network is a set of devices (often referred to as nodes) connected by communication links. A node must be an autonomous computer that is capable of sending and/or receiving data generated by other nodes on the network. A system with one control unit and many slaves is not a network, nor is a large computer with remote printers and terminals. So, an interconnected collection of autonomous computers is called a computer network.

1.4.1 Distributed Processing

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

1.4.2 Network Criteria

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

i. Performance:

Performance can be measured in many ways, including transit time and response time.Transit time is the amount of time required for a message to travel from one device to another. Response time is the elapsed time between an inquiry and a response. The performance of a network depends on a number of factors, including the number of users, the type of transmission medium, the capabilities of the connected hardware, and the efficiency of the software. Performance is often evaluated by two networking metrics: throughput and delay. We often need more throughput and less delay. However, these two criteria are often contradictory. If we try to send more data to the network, we may increase throughput but we increase the delay because of traffic congestion in the network.

ii. Reliability:

In addition to accuracy of delivery, network reliability is measured by the frequency of failure, the time a link takes to recover from a failure, and the network's robustness in a catastrophe.

iii. Security:

Network security issues include protecting data from unauthorized access, protecting data from damage and development, and implementing policies and procedures for recovery from breaches and data losses.

1.4.3 Applications of Networks: We have entered into era where we see computer networks everywhere. Though there are thousands of applications of Computer Networks, we list some here.

- Accessing Remote databases

- Accessing Remote programs
- Value added communication facility
- Marketing and sales
- Financial services
- Manufacturing
- Electronic message
- Directory services
- Information services
- Teleconferencing
- Cellular telephone
- Cable television

1.4.4 Types of Networks: Based on the coverage area a network covers, networks can be classified as Local Area Networks (LAN), Metropolitan Area Networks(MAN) & Wide area networks(WAN). Speed is high in LAN compared to bigger networks.

1.4.4.1 Local Area Network(LAN):

Local area networks, generally called LANs, are privately-owned networks within a single building or campus of up to a few kilometres in size. They are widely used to connect personal computers and workstations in company offices and factories to share resources (e.g., printers) and exchange information. LANs are distinguished from other kinds of networks by three characteristics:

(i) Their size,

(ii) Their transmission technology, and

(iii) Their topology.

LANs are restricted in size, which means that the worst-case transmission time is bounded and known in advance. Knowing this bound makes it possible to use certain kinds of designs that would not otherwise be possible. It also simplifies network management. LANs may use a transmission technology consisting of a cable to which all the machines are attached, like the telephone company party lines once used in rural areas. Traditional LANs run at speeds of 10 Mbps to 100 Mbps, have low delay (microseconds or nanoseconds), and make very few errors.

Newer LANs operate at up to 10 Gbps Various topologies are possible for broadcast LANs. Figure1.3 shows two of them.

In a bus (i.e., a linear cable) network, at any instant at most one machine is the master and is allowed to transmit. All other machines are required to refrain from sending. An arbitration mechanism is needed to resolve conflicts when two or more machines want to transmit simultaneously. The arbitration mechanism may be centralized or distributed. IEEE 802.3, popularly called Ethernet, for example, is a bus-based broadcast network with decentralized control, usually operating at 10 Mbps to 10 Gbps. Computers on an Ethernet can transmit whenever they want to; if two or more packets collide, each computer just waits a random time and tries again later.

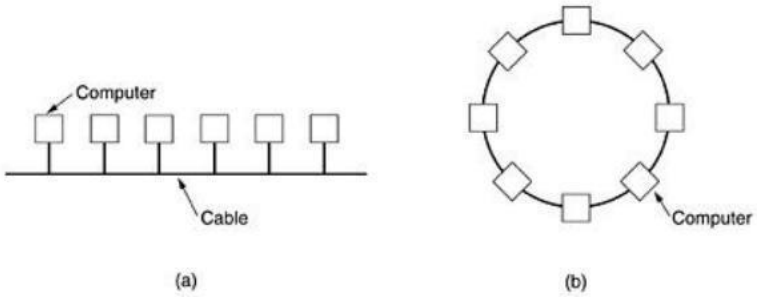


Figure 1.3 : Examples for LANs

Figure 1.3b shows second type of broadcast system known as the ring. In a ring, each bit propagates around on its own, not waiting for the rest of the packet to which it belongs. Typically, each bit circumnavigates the entire ring in the time it takes to transmit a few bits, often before the complete packet has even been transmitted. As with all other broadcast systems, some rule is needed for arbitrating simultaneous accesses to the ring. Various methods, such as having the machines take turns, are in use. IEEE 802.5 (the IBM token ring), is a ring-based LAN operating at 4 and 16 Mbps. FDDI is another example of a ring network.

1.4.4.2 Metropolitan Area Network (MAN):

A metropolitan area network, or MAN, covers a city. The best-known example of a MAN is the cable television network available in many cities. This system grew from earlier community antenna systems used in areas with poor over-the-air television reception. In these early systems, a large antenna was placed on top of a nearby hill and signal was then piped to the subscribers' houses. At first, these were locally-designed, ad hoc systems. Then companies began jumping into the business, getting contracts from city governments to wire up an entire city. The next step was television programming and even entire channels designed for cable only. Often these channels

were highly specialized, such as all news, all sports, all cooking, all gardening, and so on. But from their inception until the late 1990s, they were intended for television reception only.

To a first approximation, a MAN might look something like the system shown in Figure 1.4. In this figure both television signals and Internet are fed into the centralized head end for subsequent distribution to people's homes. Cable television is not the only MAN. Recent developments in high-speed wireless Internet access resulted in another MAN, which has been standardized as IEEE 802.16.

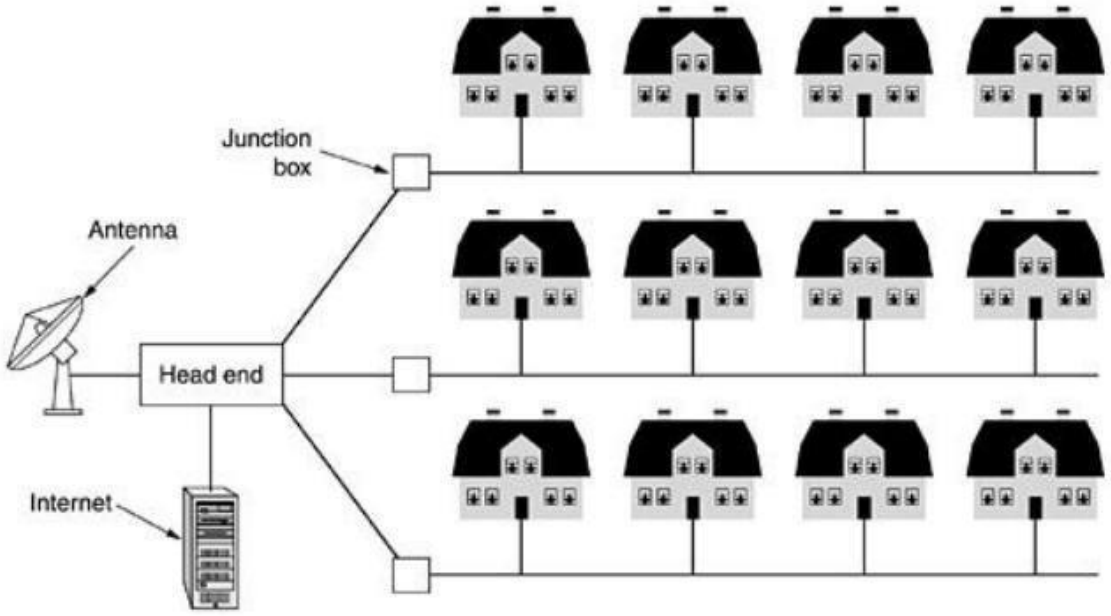


Figure 1.4: Metropolitan area network based on cable TV.

A MAN is implemented by a standard called DQDB (Distributed Queue Dual Bus) or IEEE 802.16. DQDB has two unidirectional buses (or cables) to which all the computers are attached.

1.4.4.3 Wide Area Network (WAN):

A wide area network, or WAN, spans a large geographical area, often a country or continent. It contains a collection of machines intended for running user (i.e., application) programs. These machines are called as hosts. The hosts are connected by a communication subnet, or just subnet for short. The hosts are owned by the customers (e.g., people's personal computers), whereas the communication subnet is typically owned and operated by a telephone company or Internet service provider. The job of the subnet is to carry messages from host to host, just as the telephone system carries words from speaker to listener. Separation of the pure communication aspects of the network

(the subnet) from the application aspects (the hosts), greatly simplifies the complete network design. In most wide area networks, the subnet consists of two distinct components: transmission lines and switching elements. Transmission lines move bits between machines. They can be made of copper wire, optical fiber, or even radio links. In most WANs, the network contains numerous transmission lines, each one connecting a pair of routers. If two routers that do not share a transmission line wish to communicate, they must do this indirectly, via other routers. When a packet is sent from one router to another via one or more intermediate routers, the packet is received at each intermediate router in its entirety, stored there until the required output line is free, and then forwarded. A subnet organized according to this principle is called a store-and-forward or packet-switched subnet. Nearly all wide area networks (except those using satellites) have store-and-forward subnets. When the packets are small and all the same size, they are often called cells. The principle of a packet-switched WAN is so important. Generally, when a process on some host has a message to be sent to a process on some other host, the sending host first cuts the message into packets, each one bearing its number in the sequence. These packets are then injected into the network one at a time in quick succession. The packets are transported individually over the network and deposited at the receiving host, where they are reassembled into the original message and delivered to the receiving process. A stream of packets resulting from some initial message is illustrated in Figure 1.5. In this figure, all the packets follow the route ACE, rather than ABDE or ACDE. In some networks all packets from a given message must follow the same route; in others each packed is routed separately. Of course, if ACE is the best route, all packets may be sent along it, even if each packet is individually routed.

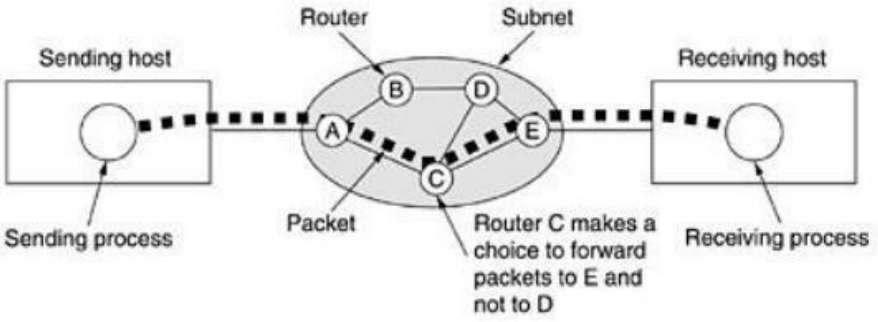


Figure 1.5: Stream of packets from sender to receiver.

Not all WANs are packet switched. A second possibility for a WAN is a satellite system. Each router has an antenna through which it can send and receive. All routers can hear the output from the satellite, and in some cases they can also hear the upward transmissions of their fellow routers to the satellite as well. Sometimes the routers are connected to a substantial point-to-point subnet, with

only some of them having a satellite antenna. Satellite networks are inherently broadcast and are most useful when the broadcast property is important.

1.4.5 Network Topologies : The term topology refers to the way in which a network is laid out physically. One or more devices connect to a link; two or more links form a topology. The topology of a network is the geometric representation of the relationship of all the links and linking devices (usually called nodes) to one another. There are four basic topologies possible: mesh, star, bus, and ring as shown in Figure 1.6.

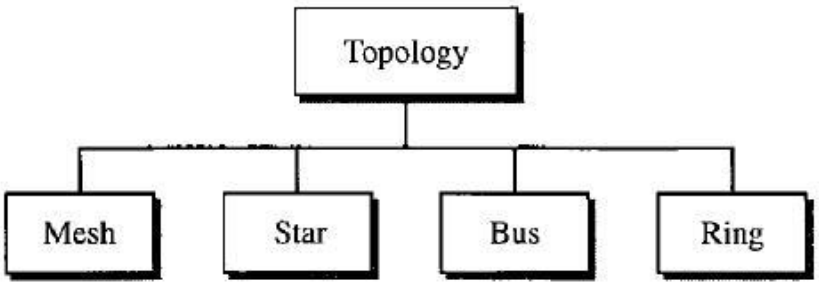


Figure 1.6 : Network Topologies

1.4.5.1 Mesh

In a mesh topology, every device has a dedicated point-to-point link to every other device. The term dedicated means that the link carries traffic only between the two devices it connects. To find the number of physical links in a fully connected mesh network with n nodes, we first consider that each node must be connected to every other node. Node 1 must be connected to n - I nodes, node 2 must be connected to n – 1 nodes, and finally node n must be connected to n - 1 nodes. We need n(n - 1) physical links. However, if each physical link allows communication in both directions (duplex mode), we can divide the number of links by 2. In other words, we can say that in a mesh topology, we need n(n -1) /2 duplex-mode links. To accommodate that many links, every device on the network must have n – 1 input/output (VO) ports to be connected to the other n - 1 stations.

Advantages:

i. The use of dedicated links guarantees that each connection can carry its own data load, thus eliminating the traffic problems that can occur when links must be shared by multiple devices.

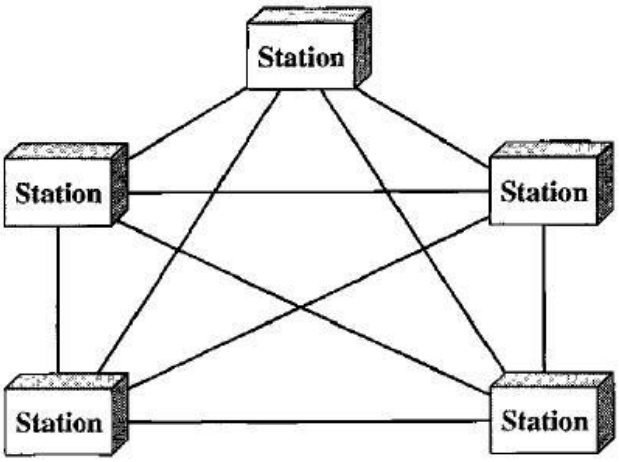


Figure 1.7 : Mesh Topology

ii. A mesh topology is robust. If one link becomes unusable, it does not incapacitate the entire system. There is the advantage of privacy or security. When every message travels along a dedicated line, only the intended recipient sees it. Physical boundaries prevent other users from gaining access to messages. Finally, point-to-point links make fault identification and fault isolation easy. Traffic can be routed to avoid links with suspected problems. This facility enables the network manager to discover the precise location of the fault and aids in finding its cause and solution.

Disadvantages:

i. Disadvantage of a mesh are related to the amount of cabling because every device must be connected to every other device, installation and reconnection are difficult.

ii. The sheer bulk of the wiring can be greater than the available space (in walls, ceilings, or floors) can accommodate. Finally, the hardware required to connect each link (I/O ports and cable) can be prohibitively expensive. For these reasons a mesh topology is usually implemented in a limited fashion, for example, as a backbone connecting the main computers of a hybrid network that can include several other topologies.

1.4.5.2 Star

In a star topology as given in Figure 1.8, each device has a dedicated point-to-point link only to a central controller, usually called a HUB. The devices are not directly linked to one another. Unlike a mesh topology, a star topology does not allow direct traffic between devices. The controller acts as an exchange: If one device wants to send data to another, it sends the data to the central system hub, which then relays the data to the other connected device. A star topology is less expensive than a mesh topology. In a star, each device needs only one link and one I/O port to

connect it to any number of others. This factor also makes it easy to install and reconfigure. Far less cabling needs to be housed, and additions, moves, and deletions involve only one connection: between that device and the hub. Other advantages include robustness. If one link fails, only that link is affected. All other links remain active. This factor also lends itself to easy fault identification and fault isolation. As long as the hub is working, it can be used to monitor link problems and bypass defective links. One big disadvantage of a star topology is the dependency of the whole topology on one single point, the hub. If the hub goes down, the whole system is dead. Although a star requires far less cable than a mesh, each node must be linked to a central hub. For this reason, often more cabling is required in a star than in some other topologies (such as ring or bus).

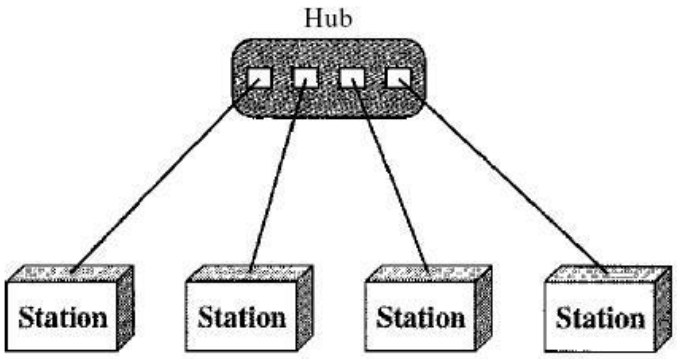


Figure 1.8 : Star Topology

1.4.5.3 Bus

The preceding examples all describe point-to-point connections. A bus topology, on the other hand, is multipoint. As shown in Figure 1.9, one long cable acts as a backbone to link all the devices in a network. Nodes are connected to the bus cable by drop lines and taps. A drop line is a connection running between the device and the main cable. A tap is a connector that either splices into the main cable or punctures the sheathing of a cable to create a contact with the metallic core. As signal travels along the backbone, some of its energy is transformed into heat. Therefore, it becomes weaker and weaker as it travels farther and farther. For this reason there is a limit on the number of taps a bus can support on the distance between those taps.

Advantages of a bus topology include ease of installation. Backbone cable can be laid along the most efficient path, then connected to the nodes by drop lines of various lengths. In this way, a bus uses less cabling than mesh or star topologies. In a star, for example, four network devices in the same room require four lengths of cable reaching all the way to the hub. In a bus, this redundancy is eliminated. Only the backbone cable stretches through the entire facility. Each drop line has to reach only as far as the nearest point on the backbone.

Disadvantages include difficult reconnection and fault isolation. A bus is usually designed to be optimally efficient at installation. It can therefore be difficult to add new devices. Signal reflection at the taps can cause degradation in quality. This degradation can be controlled by limiting the number and spacing of devices connected to a given length of cable. Adding new devices may therefore require modification or replacement of the backbone.

In addition, a fault or break in the bus cable stops all transmission, even between devices on the same side of the problem. The damaged area reflects signals back in the direction of origin, creating noise in both directions. Bus topology was the one of the first topologies used in the design of early local area networks. Ethernet LANs can use a bus topology, but they are less popular.

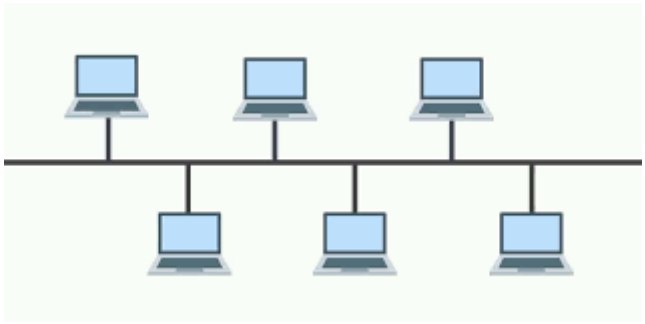
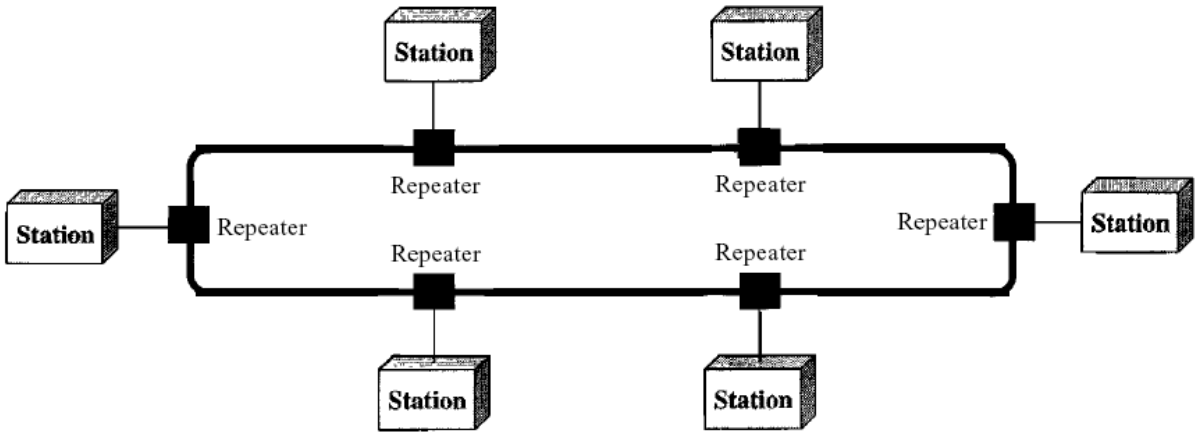


Figure 1.9 : Bus Topology

1.4.5.4 Ring

In a ring topology as given in Figure 1.10, each device has a dedicated point-to-point connection with only the two devices on either side of it. A signal is passed along the ring in one direction, from device to device, until it reaches its destination. Each device in the ring incorporates a repeater. When a device receives a signal intended for another device, its repeater regenerates the bits and passes them along.

Figure 1.10 : Ring Topology



A ring is relatively easy to install and reconfigure. Each device is linked to only its immediate neighbors (either physically or logically). To add or delete a device requires changing only two connections. The only constraints are media and traffic considerations (maximum ring length and number of devices). In addition, fault isolation is simplified. Generally in a ring, a signal is circulating at all times. If one device does not receive a signal within a specified period, it can issue an alarm. The alarm alerts the network operator to the problem and its location. However, unidirectional traffic can be a disadvantage. In a simple ring, a break in the ring (such as a disabled station) can disable the entire network. This weakness can be solved by using a **dual ring** or a switch capable of closing off the break. Ring topology was prevalent when IBM introduced its local-area network Token Ring. Today, the need for higher-speed LANs has made this topology less popular.

1.4.5.5 Hybrid

In real world, based on requirments, we build network with multiple topologies. Its called as hybrid. For example, we can have a main star topology with each branch connecting several stations in a bus topology as shown in Figure 1.11.

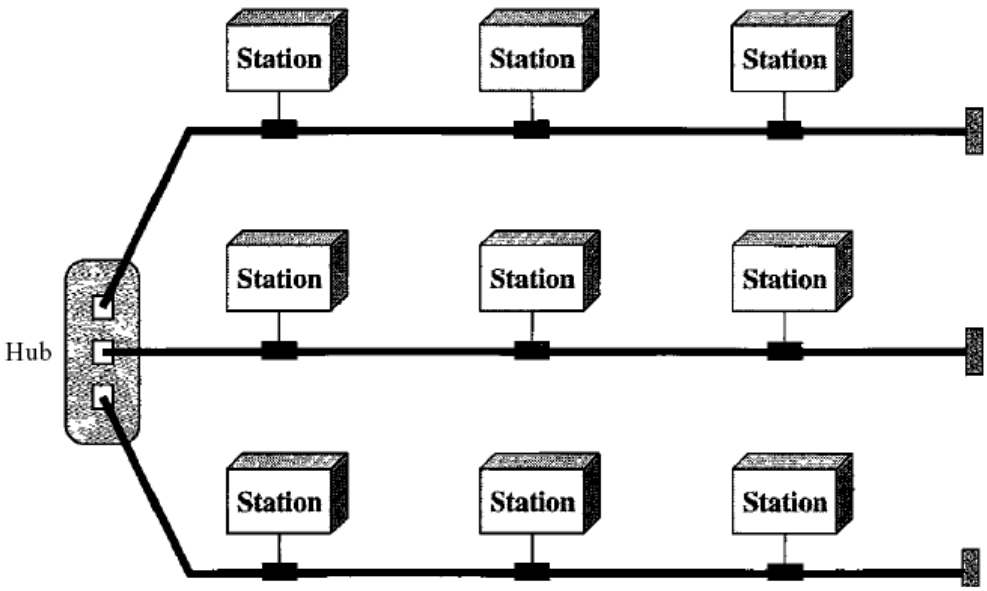


Figure 1.11 : Hybrid Topology

1.5 THE INTERNET

The Internet has revolutionized many aspects of our daily lives. It has affected the way we do business as well as the way we spend our leisure time. Count the ways you've used the Internet recently. Perhaps you've sent electronic mail (e- ail) to a business associate, paid a utility bill, read a newspaper from a distant city, or looked up a local movie schedule by using the Internet Or maybe you researched a medical topic, booked a hotel reservation, chatted with a fellow Trekkie, or comparison-shopped for a car. The Internet is a communication system that has brought a wealth of information to our fingertips and organized it for our use.

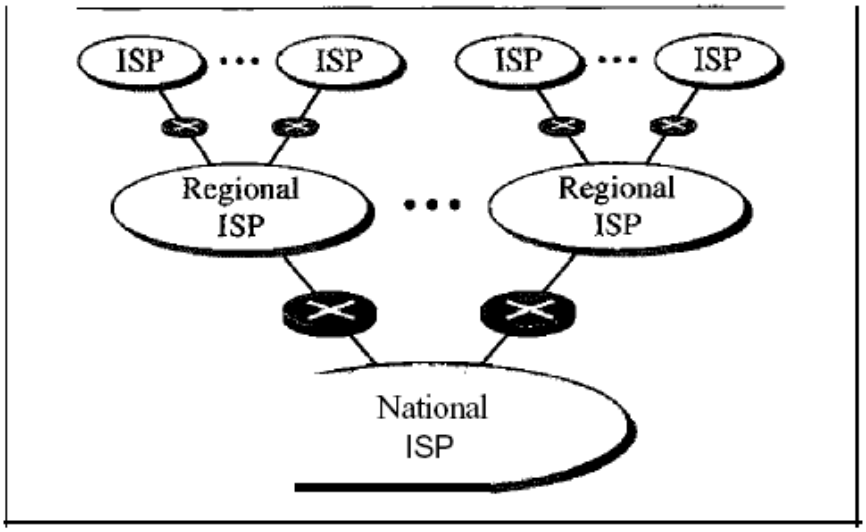
1.5.1 A Brief History

A network is a group of connected communicating devices such as computers and printers. An internet (note the lowercase letter i) is two or more networks that can communicate with each other. The most notable internet is called the Internet (uppercase letter I), a collaboration of more than hundreds of thousands of interconnected networks. Private individuals as well as various organizations such as government agencies, schools, research facilities, corporations, and libraries in more than 100 countries use the Internet. Millions of people are users. Yet this extraordinary communication system only came into being in 1969.

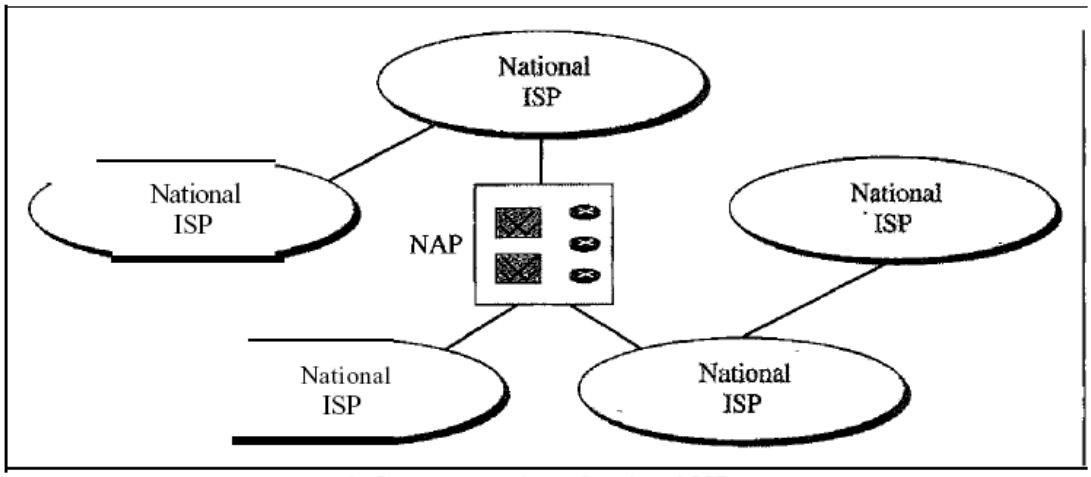
In the mid-1960s, mainframe computers in research organizations were standalone devices. Computers from different manufacturers were unable to communicate with one another. The Advanced Research Projects Agency (ARPA) in the Department of Defense (DoD) was interested in finding a way to connect computers so that the researchers they funded could share their findings, thereby reducing costs and eliminating duplication of effort. In 1967, at an Association for Computing Machinery (ACM) meeting, ARPA presented its ideas for ARPANET, a small network of connected computers. The idea was that each host computer (not necessarily from the same manufacturer) would be attached to a specialized computer, called an interface message processor (IMP). The IMPs, in turn, would be connected to one another. Each IMP had to be able to communicate with other IMPs as well as with its own attached host. By 1969, ARPANET was a

reality. Four nodes, at the University of California at Los Angeles (UCLA), the University of California at Santa Barbara (UCSB), Stanford Research Institute (SRI), and the University of Utah, were connected via the IMPs to form a network. Software called the Network Control Protocol (NCP) provided communication between the hosts.

In 1972, Vint Cerf and Bob Kahn, both of whom were part of the core ARPANET group, collaborated on what they called the Internet Project. Cerf and Kahn's landmark 1973 paper outlined the protocols to achieve end-to-end delivery of packets. This paper on Transmission Control Protocol (TCP) included concepts such as encapsulation, the datagram, and the functions of a gateway. Shortly thereafter, authorities made a decision to split TCP into two protocols: Transmission Control Protocol (TCP) and Internetworking Protocol (lP). IP would handle datagram routing while TCP would be responsible for higher-level functions such as segmentation, reassembly, and error detection. The internetworking protocol became known as TCPI/IP.



a. Structure of a national ISP



b. Interconnection of national ISPs

Figure 1.12: Conceptual Vie

w of Internet

1.5.2 The Internet Today

The Internet has come a long way since the 1960s. The Internet today is not a simple hierarchical structure. It is made up of many wide and local area networks joined by connecting devices and switching stations. It is difficult to give an accurate representation of the Internet because it is continually changing-new networks are being added, existing networks are adding addresses, and networks of defunct companies are being removed. Today most end users who want Internet connection use the services of Internet service providers (lSPs). There are international service providers, national service providers, regional service providers, and local service providers. The Internet today is run by private companies, not the government. Figure 1.12 shows a conceptual (not geographic) view of the Internet.

1.5.2.1 International Internet Service Providers:

At the top of the hierarchy are the international service providers that connect nations together.

1.5.2.2 National Internet Service Providers:

The national Internet service providers are backbone networks created and maintained by specialized companies. There are many national ISPs operating in North America; some of the most well known are SprintLink, PSINet, UUNet Technology, AGIS, and internet Mel. To provide connectivity between the end users, these backbone networks are connected by complex switching stations (normally run by a third party) called network access points (NAPs). Some national ISP networks are also connected to one another by private switching stations called peering points. These normally operate at a high data rate (up to 600 Mbps).

1.5.2.3 Regional Internet Service Providers:

Regional internet service providers or regional ISPs are smaller ISPs that are connected to one or more national ISPs. They are at the third level of the hierarchy with a smaller data rate.

1.5.2.4 Local Internet Service Providers:

Local Internet service providers provide direct service to the end users. The local ISPs can be connected to regional ISPs or directly to national ISPs. Most end users are connected to the local ISPs. Note that in this sense, a local ISP can be a company that just provides Internet services, a

corporation with a network that supplies services to its own employees, or a nonprofit organization, such as a college or a university, that runs its own network. Each of these local ISPs can be connected to a regional or national service provider.

1.6 PROTOCOLS AND STANDARDS

In computer networks, communication occurs between entities in different systems. An entity is anything capable of sending or receiving information. However, two entities cannot simply send bit streams to each other and expect to be understood. For communication to occur, the entities must agree on a protocol. A protocol is a set of rules that govern data communications. A protocol defines what is communicated, how it is communicated, and when it is communicated. The key elements of a protocol are syntax, semantics, and timing.

i. Syntax

The term syntax refers to the structure or format of the data, meaning the order in which they are presented. For example, a simple protocol might expect the first 8 bits of data to be the address of the sender, the second 8 bits to be the address of the receiver, and the rest of the stream to be the message itself.

ii. Semantics

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

iii. Timing

The term timing refers to two characteristics: when data should be sent and how fast they can be sent. For example, if a sender produces data at 100 Mbps but the receiver can process data at only 1 Mbps, the transmission will overload the receiver and some data will be lost.

iv. Standards

Standards are essential in creating and maintaining an open and competitive market for equipment manufacturers and in guaranteeing national and international interoperability of data and telecommunications technology and processes. Standards provide guidelines to manufacturers,

vendors, government agencies, and other service providers to ensure the kind of interconnectivity necessary in today's marketplace and in international communications. Data communication standards fall into two categories: de facto (meaning "by fact" or "by convention") and de jure (meaning "by law" or "by regulation"). Standards that have not been approved by an organized body but have been adopted as standards through widespread use are de facto standards. De facto standards are often established originally by manufacturers who seek to define the functionality of a new product or technology. Those standards that have been legislated by an officially recognized body are de jure standards.

1.7 LAYERED TASKS

We use the concept of layers in our daily life. As an example, let us consider two friends who communicate through postal mail. The process of sending a letter to a friend would be complex if there were no services available from the post office. Below Figure 1.12 shows the steps in this task.

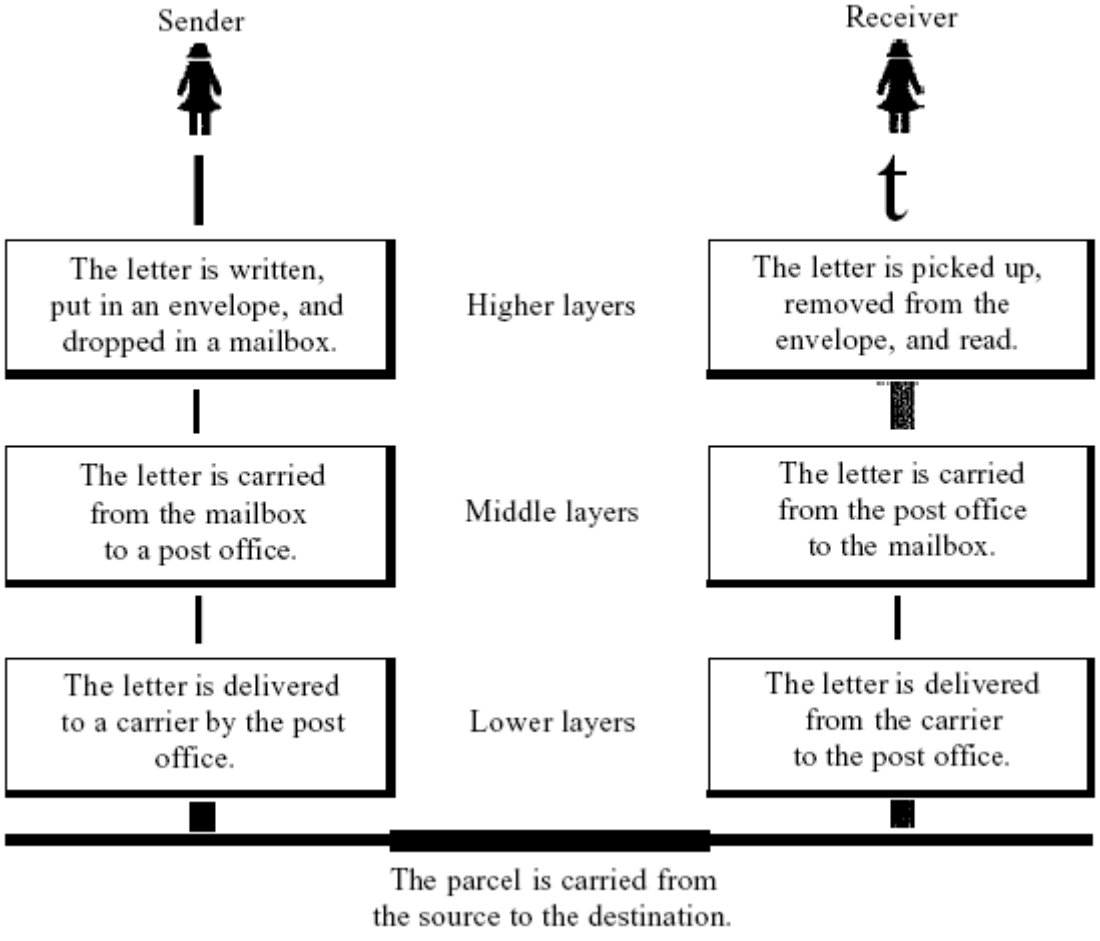


Figure 1.13: Postal Communication Tasks as layers Sender, Receiver, and Carrier

In Figure 1.13, we have a sender, a receiver, and a carrier that transports the letter. There is a hierarchy of tasks.

At the Sender Site

Let us first describe, in order, the activities that take place at the sender site.

Higher layer : The sender writes the letter, inserts the letter in an envelope, writes the sender and receiver addresses, and drops the letter in a mailbox.

Middle layer : The letter is picked up by a letter carrier and delivered to the post office.

Lower layer : The letter is sorted at the post office; a carrier transports the letter.

On the Way: The letter is then on its way to the recipient. On the way to the recipient's local post office, the letter may actually go through a central office. In addition, it may be transported by truck, train, airplane, boat, or a combination of these.

At the Receiver Site

Lower layer : The carrier transports the letter to the post office.

Middle layer : The letter is sorted and delivered to the recipient's mailbox.

Higher layer : The receiver picks up the letter, opens the envelope, and reads it.

1.8 The OSI Reference Model

The OSI model (minus the physical medium) is shown in Figure 1.14. This model is based on a proposal developed by the International Standards Organization (ISO) as a first step toward international standardization of the protocols used in the various layers (Day and Zimmermann, 1983). It was revised in 1995(Day, 1995). The model is called the ISO-OSI (Open Systems Interconnection) Reference Model because it deals with connecting open systems—that is, systems that are open for communication with other systems. The OSI model has seven layers. The principles that were applied to arrive at the seven layers can be briefly summarized as follows:

- A layer should be created where a different abstraction is needed.
- Each layer should perform a well-defined function.
- The function of each layer should be chosen with an eye toward defining internationally standardized protocols.
- The layer boundaries should be chosen to minimize the information flow across the interfaces.

- The number of layers should be large enough that distinct functions need not be thrown together in the same layer out of necessity and small enough that the architecture does not become unwieldy.

The Physical Layer

The physical layer is concerned with transmitting raw bits over a communication channel. The design issues have to do with making sure that when one side sends a 1 bit, it is received by the other side as a 1 bit, not as a 0 bit.

The Data Link Layer

The main task of the data link layer is to transform a raw transmission facility into a line that appears free of undetected transmission errors to the network layer. It accomplishes this task by having the sender break up the input data into data frames (typically a few hundred or a few thousand bytes) and transmits the frames sequentially. If the service is reliable, the receiver confirms correct receipt of each frame by sending back an acknowledgement frame. Another issue that arises in the data link layer (and most of the higher layers as well) is how to keep a fast transmitter from drowning a slow receiver in data. Some traffic regulation mechanism is often needed to let the transmitter know how much buffer space the receiver has at the moment. Frequently, this flow regulation and the error handling are integrated.

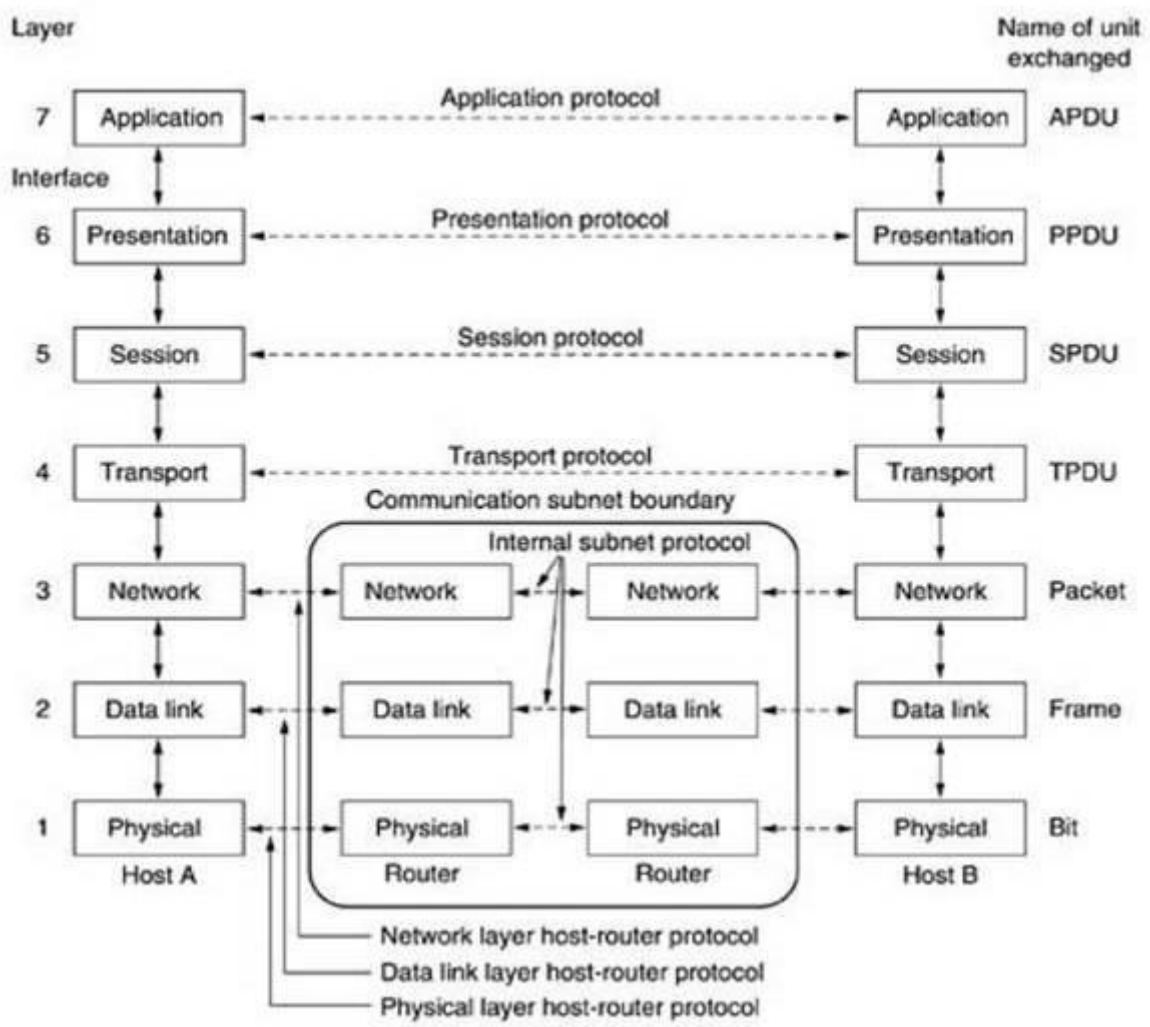


Figure 1.14 : ISO OSI Reference Model

The Network Layer

The network layer controls the operation of the subnet. A key design issue is determining how packets are routed from source to destination. Routes can be based on static tables that are "wired into" the network and rarely changed. They can also be determined at the start of each conversation, for example, a terminal session (e.g., a login to a remote machine). Finally, they can be highly dynamic, being determined anew for each packet, to reflect the current network load.

If too many packets are present in the subnet at the same time, they will get in one another's way, forming bottlenecks. The control of such congestion also belongs to the network layer. More generally, the quality of service provided (delay, transit time, jitter, etc.) is also a network layer issue. When a packet has to travel from one network to another to get to its destination, many problems can arise. The addressing used by the second network may be different from the first one. The second one may not accept the packet at all because it is too large. The protocols may differ,

and so on. It is up to the network layer to overcome all these problems to allow heterogeneous networks to be interconnected. In broadcast networks, the routing problem is simple, so the network layer is often thin or even nonexistent.

The Transport Layer

The basic function of the transport layer is to accept data from above, split it up into smaller units if need be, pass these to the network layer, and ensure that the pieces all arrive correctly at the other end. Furthermore, all this must be done efficiently and in a way that isolates the upper layers from the inevitable changes in the hardware technology. The transport layer also determines what type of service to provide to the session layer, and, ultimately, to the users of the network. The most popular type of transport connection is an error-free point-to-point channel that delivers messages or bytes in the order in which they were sent. However, other possible kinds of transport service are the transporting of isolated messages, with no guarantee about the order of delivery, and the broadcasting of messages to multiple destinations. The type of service is determined when the connection is established. The transport layer is a true end-to-end layer, all the way from the source to the destination. In other words, a program on the source machine carries on a conversation with a similar program on the destination machine, using the message headers and control messages. In the lower layers,the protocols are between each machine and its immediate neighbours, and not between the ultimate source and destination machines, which may be separated by many routers.

The Session Layer

The session layer allows users on different machines to establish sessions between them. Sessions offer various services, including dialog control (keeping track of whose turn it is to transmit), token management (preventing two parties from attempting the same critical operation at the same time), and synchronization (check pointing long transmissions to allow them to continue from where they were after a crash).

The Presentation Layer

The presentation layer is concerned with the syntax and semantics of the information transmitted. In order to make it possible for computers with different data representations to communicate, the data structures to be exchanged can be defined in an abstract way, along with a standard encoding to be used "on the wire." The presentation layer manages these abstract data structures and allows higher-level data structures (e.g., banking records), to be defined and exchanged.

The Application Layer

The application layer contains a variety of protocols that are commonly needed by users. One widely-used application protocol is HTTP (Hypertext Transfer Protocol), which is the basis for the World Wide Web. When a browser wants a Web page, it sends the name of the page it wants to the server using HTTP. The server then sends the page back. Other application protocols are used for file transfer, electronic mail, and network news.

1.9 The TCP/IP Reference Model

The TCP/IP reference model was developed prior to OSI model. The major design goals of this model were,

- To connect multiple networks together so that they appear as a single network.
- To survive after partial subnet hardware failures.
- To provide a flexible architecture.

Unlike OSI reference model, TCP/IP reference model has only 4 layers. They are,

- Host-to-Network Layer
- Internet Layer
- Transport Layer
- Application Layer

Host-to-Network Layer

The TCP/IP reference model does not really say much about what happens here, except to point out that the host has to connect to the network using some protocol so it can send IP packets to it. This protocol is not defined and varies from host to host and network to network.

Internet Layer

This layer, called the internet layer, is the linchpin that holds the whole architecture together. Its job is to permit hosts to inject packets into any network and have they travel independently to the destination (potentially on a different network). They may even arrive in a different order than they were sent, in which case it is the job of higher layers to rearrange them, if in-order delivery is

desired. Note that "internet" is used here in a generic sense, even though this layer is present in the Internet. The internet layer defines an official packet format and protocol called IP (Internet Protocol). The job of the internet layer is to deliver IP packets where they are supposed to go. Packet routing is clearly the major issue here, as is avoiding congestion. For these reasons, it is reasonable to say that the TCP/IP internet layer is similar in functionality to the OSI network layer. Figure 1.15 shows this correspondence.

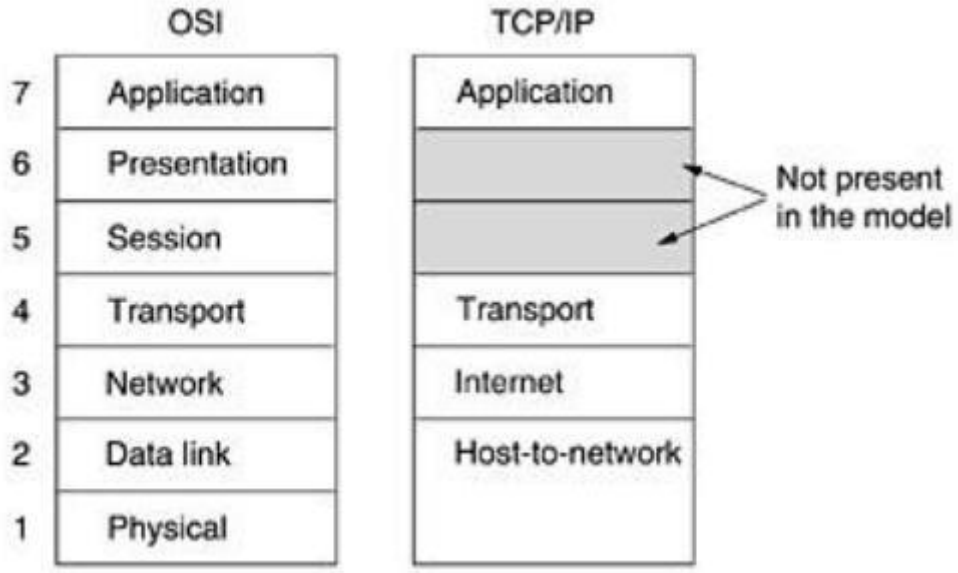


Figure 1.15 : TCP/IP Model

The Transport Layer

The layer above the internet layer in the TCP/IP model is now usually called the transport layer. It is designed to allow peer entities on the source and destination hosts to carry on a conversation, just as in the OSI transport layer. Two end-to-end transport protocols have been defined here. The first one, TCP (Transmission Control Protocol), is a reliable connection-oriented protocol that allows a byte stream originating on one machine to be delivered without error on any other machine in the internet. It fragments the incoming byte stream into discrete messages and passes each one on to the internet layer. At the destination, the receiving TCP process reassembles the received messages into the output stream. TCP also handles flow control to make sure a fast sender cannot swamp a slow receiver with more messages than it can handle.

The second protocol in this layer, UDP (User Datagram Protocol), is an unreliable, connectionless protocol for applications that do not want TCP's sequencing or flow control and wish to provide their own. It is also widely used for one-shot, client-server-type request-reply queries and

applications in which prompt delivery is more important than accurate delivery, such as transmitting speech or video. The relation of IP, TCP, and UDP is shown in Figure 1.16. Since the model was developed, IP has been implemented on many other networks.

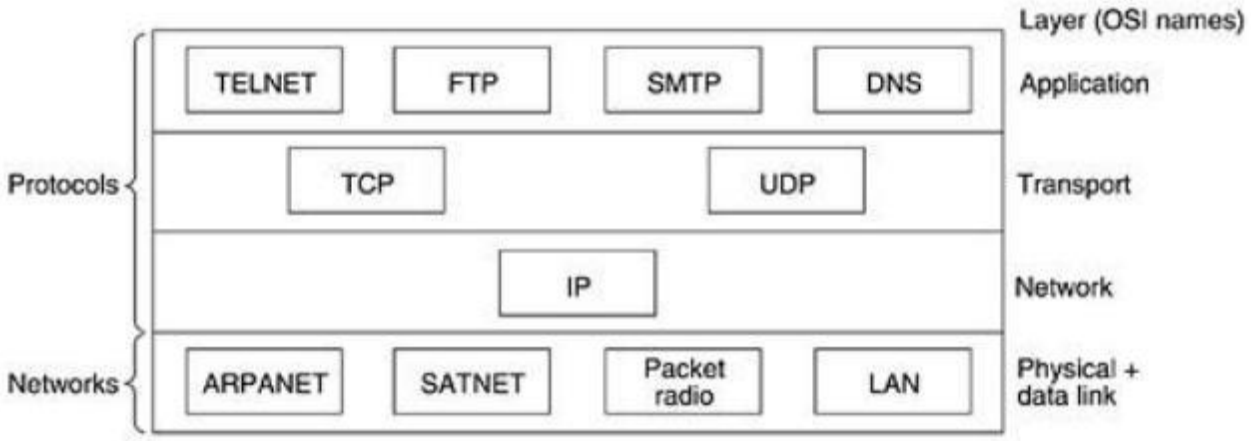


Figure 1.16 : Protocols in TCP/IP Model

The Application Layer

The TCP/IP model does not have session or presentation layers. On top of the transport layer is the application layer. It contains all the higher-level protocols. The early ones included virtual terminal (TELNET), file transfer (FTP), and electronic mail (SMTP). The virtual terminal protocol allows a user on one machine to log onto a distant machine and work there. The file transfer protocol provides a way to move text file efficiently from one machine to another. Electronic mail was originally just a kind of file transfer, but later a specialized protocol (SMTP) was developed for it. Many other protocols have been added to these over the years: the Domain Name System (DNS) for mapping host names onto their network addresses, NNTP, the protocol for moving USENET news articles around, and HTTP, the protocol for fetching pages on the World Wide Web, and many others.

1.10 Comparison of the OSI and TCP/IP Reference Models

The OSI and TCP/IP reference models have much in common. Both are based on the concept of a stack of independent protocols. Also, the functionality of the layers is roughly similar. For example, in both models the layers up through and including the transport layer are there to provide an end-to-end, network-independent transport service to processes wishing to communicate. These layers form the transport provider. Again in both models, the layers above transport are

application-oriented users of the transport service. Despite these fundamental similarities, the two models also have many differences. Three concepts are central to the OSI model:

- Services.
- Interfaces.
- Protocols.

Probably the biggest contribution of the OSI model is to make the distinction between these three concepts explicit. Each layer performs some services for the layer above it. The service definition tells what the layer does, not how entities above it access it or how the layer works. It defines the layer's semantics. A layer's interface tells the processes above it how to access it. It specifies what the parameters are and what results to expect. It, too, says nothing about how the layer works inside. Finally, the peer protocols used in a layer are the layer's own business. It can use any protocols it wants to, as long as it gets the job done (i.e., provides the offered services). It can also change them at will without affecting software in higher layers.

The TCP/IP model did not originally clearly distinguish between service, interface, and protocol, although people have tried to retrofit it after the fact to make it more OSI-like. For example, the only real services offered by the internet layer are SEND IP PACKET and RECEIVE IP PACKET.

As a consequence, the protocols in the OSI model are better hidden than in the TCP/IP model and can be replaced relatively easily as the technology changes. Being able to make such changes is one of the main purposes of having layered protocols in the first place. The OSI reference model was devised before the corresponding protocols were invented. This ordering means that the model was not biased toward one particular set of protocols, a fact that made it quite general. The downside of this ordering is that the designers did not have much experience with the subject and did not have a good idea of which functionality to put in which layer.

Another difference is in the area of connectionless versus connection-oriented communication. The OSI model supports both connectionless and connection-oriented communication in the network layer, but only connection-oriented communication in the transport layer, where it counts (because the transport service is visible to the users). The TCP/IP model has only one mode in the network layer (connectionless) but supports both modes in the transport layer, giving the users a choice. This choice is especially important for simple request-response protocols.

1.11 Network Components

Network formation happen at two levels. First, we need to establish connection between systems physically. When we are connecting just two systems, we dont need much hardwares like Switches but when we have to connect more than two systems, we need lot of additional hardwares. After connecting the systems physically, we need to install network softwares and configure. Then we can use network. In first step, while connecting the systems physically, we need some of the following devices.

1.11.1 Network Interface Card

A network interface card (NIC) is a hardware component, typically a circuit board or chip, which is installed on a computer so that it can connect to a network. Modern NICs provide functionality to computers such as support for I/O interupt, direct memory access (DMA) interfaces, data transmission, network traffic engineering partitioning. A NIC provides a computer with a dedicated, full-time connection to a network by implementing the physical layer circuitry necessary for communicating with a data link layer standard, such asEthernet orWi-Fi. Each card represents a device and can prepare, transmit and control the flow of data on the network. The NIC uses the OSI Model to send signals at the physical layer, transmit data packets at the network layer and operate as an interface at the TCP/IP layer. The network card operates as a middleman between a computer and a data network. For example, when a user requests a web page, the computer will pass the request to the network card which converts it into electrical impulses. Those impulses are received by a web server on the internet and by sending the web page back to the network card as electrical signals. The card gets these signals and translates them into the data that the computer displays. Originally, network controllers were implemented as expansion cards that could be plugged into a computer port,router or USB device. However, more modern controllers are built directly into the computer motherboard chipset. Expansion card NICs can be purchased online or in retail stores if additional independent network connections are needed. When purchasing a NIC, specifications should correspond with the standard of the network. The term network interface card is often considered interchangeable with the terms **network interface controller, network adapter and LAN adapter**.

Figure 1.17 : A Network Interface Card

1.11.2 Types of network interface cards

While the standard NIC is a plastic circuit board that slides into a computer to connect with the motherboard, there are multiple ways this connection can occur:

Wireless- These are NICs that use an antenna to provide wireless reception through radio frequency waves. Wireless NICs are designed for connection.

Wired- These are NICs that have input jacks made for cables. The most popular wired LAN technology is Ethernet.

USB- These are NICs that provide network connections through a device plugged into the USB port.

Fiber Optic - These are expensive and more complex NICs that are used as a high-speed support system for network traffic handling on server computers. This could also be accomplished by combining multiple NICs.

1.11.3 Components of network interface cards

Speed- All NICs have a speed rating in terms of Mbps that suggests the general performance of the card when implemented in a computer network with ample bandwidth. If the bandwidth is lower than the NIC or multiple computers are connected with the same controller, this will slow down the labeled speed. The average Ethernet NICs are offered in 10 Mbps, 100 Mbps, 1000 Mbps 1Gbps varieties.

Driver- This is the required software that passes data between the computer's operating system and the NIC. When a NIC is installed on a computer, the corresponding driver software is also downloaded. Drivers must stay updated and uncorrupted to ensure optimal performance from the NIC.

MAC address- Unique, unchangeable MAC addressess, also known as a physical network address, assigned to NICs that is used to deliver Ethernet packets to the computer.

Connectivity LED- Most NICs have an LED indicator integrated into the connector to notify the user of when the network is connected and data is being transmitted.

Router- A router is also sometimes needed to allow communication between a computer and other devices. In this case, the NIC connects to the router which is connected to the internet.

1.11.2 Media

Network media refers to the communication channels used to interconnect nodes on a computer network. Typical examples of network media include copper coaxial cable, copper twisted pair cables and optical fiber cables used in wired networks, and radio waves used in wireless data communications networks. A transmission medium is a physical path between the transmitter and the receiver i.e it is the channel through which data is sent from one place to another. Transmission Media is broadly classified into the following types.

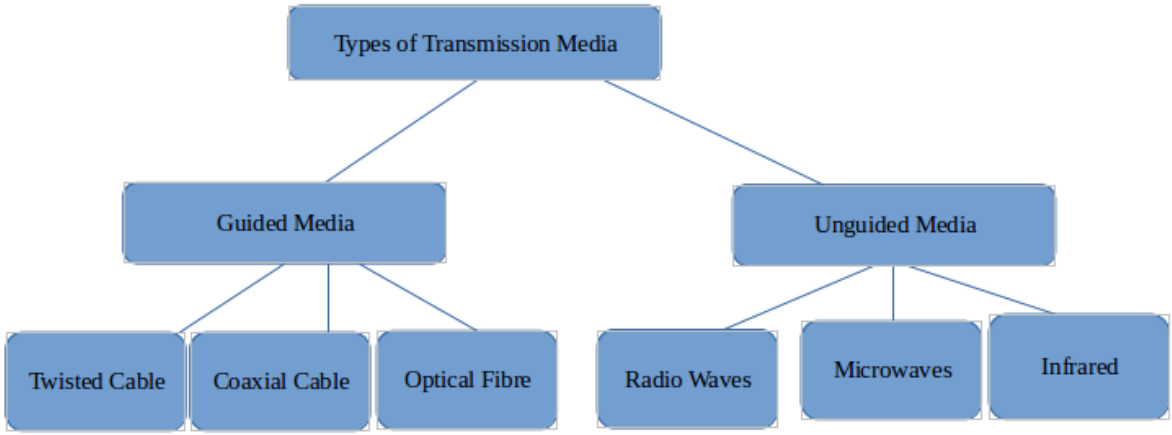


Figure 1.17 : Different types of Media

1.11.2.1 Guided Media

It is also referred to as Wired or Bounded transmission media. Signals being transmitted are directed and confined in a narrow pathway by using physical links.

Features:

- High Speed
- Secure
- Used for comparatively shorter distances

There are 3 major types of Guided Media:

1.11.2.1.1 Twisted Pair Cable

It consists of 2 separately insulated conductor wires wound about each other. Generally,

several such pairs are bundled together in a protective sheath. They are the most widely used Transmission Media. Twisted Pair is of two types:

Unshielded Twisted Pair (UTP):

This type of cable has the ability to block interference and does not depend on a physical shield for this purpose. It is used for telephonic applications.

Advantages:

- Least expensive
- Easy to install
- High speed capacity

Disadvantages:

- Susceptible to external interference
- Lower capacity and performance in comparison to STP
- Short distance transmission due to attenuation

Shielded Twisted Pair (STP):

This type of cable consists of a special jacket to block external interference. It is used in fast-data-rate Ethernet and in voice and data channels of telephone lines.

Advantages:

- Better performance at a higher data rate in comparison to UTP
- Eliminates crosstalk
- Comparitively faster

Disadvantages:

- Comparitively difficult to install and manufacture
- More expensive
- Bulky

1.11.2.1.2 Coaxial Cable

It has an outer plastic covering containing 2 parallel conductors each having a separate insulated protection cover. Coaxial cable transmits information in two modes: Baseband mode(dedicated cable bandwidth) and Broadband mode(cable bandwidth is split into separate ranges). Cable TVs and analog television networks widely use Coaxial cables.

Advantages:

- High Bandwidth
- Better noise Immunity
- Easy to install and expand
- Inexpensive

Disadvantages:

- Single cable failure can disrupt the entire network

1.11.2.1.3 Optical Fibre Cable

It uses the concept of reflection of light through a core made up of glass or plastic. The core is surrounded by a less dense glass or plastic covering called the cladding. It is used for transmission of large volumes of data.

Advantages:

- Increased capacity and bandwidth
- Light weight
- Less signal attenuation
- Immunity to electromagnetic interference
- Resistance to corrosive materials

Disadvantages:

- Difficult to install and maintain
- High cost
- Fragile
- unidirectional, ie, will need another fibre, if we need bidirectional communication

1.11.2.2 Unguided Media

It is also referred to as Wireless or Unbounded transmission media.No physical medium is required for the transmission of electromagnetic signals.

Features:

- Signal is broadcasted through air
- Less Secure
- Used for larger distances

There are 3 major types of Unguided Media:

1.11.2.2.1 Radiowaves

These are easy to generate and can penetrate through buildings. The sending and receiving antennas need not be aligned. Frequency Range:3KHz – 1GHz. AM and FM radios and cordless phones use Radiowaves for transmission.

Further Categorized as (i) Terrestrial and (ii) Satellite.

1.11.2.2.2 Microwaves

It is a line of sight transmission i.e. the sending and receiving antennas need to be properly aligned with each other. The distance covered by the signal is directly proportional to the height of the antenna. Frequency Range:1GHz – 300GHz. These are majorly used for mobile phone communication and television distribution.

1.11.2.2.3 Infrared

Infrared waves are used for very short distance communication. They cannot penetrate through obstacles. This prevents interference between systems. Frequency Range:300GHz – 400THz. It is used in TV remotes, wireless mouse, keyboard, printer, etc.

1.11.3 Repeater

A repeater operates at the physical layer. Its job is to regenerate the signal over the same network before the signal becomes too weak or corrupted so as to extend the length to which the signal can be transmitted over the same network. An important point to be noted about repeaters is that they do not amplify the signal. When the signal becomes weak, they copy the signal bit by bit and regenerate it at the original strength. It is a 2 port device.

1.11.4 Hub

In Star topology, all the systems are directly connected to a central system called Hub. A hub is basically a multiport repeater. A hub connects multiple wires coming from different branches, for example, the connector in star topology which connects different stations. Hubs cannot filter data, so data packets are sent to all connected devices. In other words, collision domain of all hosts connected through Hub remains one. Also, they do not have intelligence to find out best path for data packets which leads to inefficiencies and wastage. There are two types of Hubs.

Figure 1.18 :Typical Hub

- **Active Hub :-** These are the hubs which have their own power supply and can clean , boost and relay the signal along the network. It serves both as a repeater as well as wiring center. These are used to extend maximum distance between nodes.
- **Passive Hub :-** These are the hubs which collect wiring from nodes and power supply from active hub. These hubs relay signals onto the network without cleaning and boosting them and can't be used to extend distance between nodes.

1.11.5 Switch

A switch is a multi port bridge with a buffer and a design that can boost its efficiency (large number of ports imply less traffic) and performance. Switch is data link layer device. Switch can perform error checking before forwarding data, that makes it very efficient as it does not forward packets that have errors and forward good packets selectively to correct port only. In other words, switch divides collision domain of hosts, but broadcast domain remains same.

Figure 1.19 : A 24 Ports Switch

1.11.6 Router

A router is a device like a switch that routes data packets based on their IP addresses. Router is mainly a Network Layer device. Routers normally connect LANs and WANs together and have a dynamically updating routing table based on which they make decisions on routing the data packets. Router divide broadcast domains of hosts connected through it.

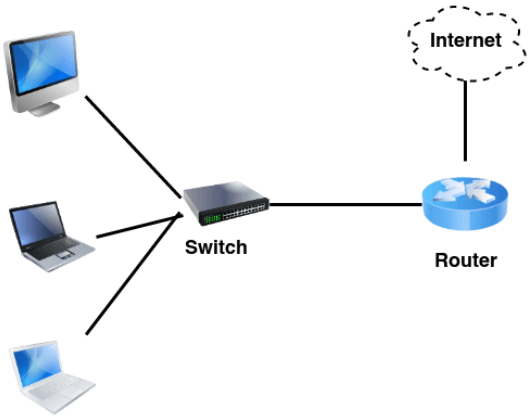


Figure 1.20 : Switch & Router in a Network

1.11.7 Bridge

A bridge operates at data link layer. A bridge is a repeater, with add on functionality of filtering content by reading the MAC addresses of source and destination. It is also used for interconnecting two LANs working on the same protocol. It has a single input and single output port, thus making it a 2 port device.

Types of Bridges

- **Transparent Bridges :-** These are the bridges in which the stations are completely unaware of the bridge's existence i.e. whether or not a bridge is added or deleted from the network, reconfiguration of the stations is unnecessary. These bridges make use of two processes i.e. bridge forwarding and bridge learning.

- **Source Routing Bridges :-** In these bridges, routing operation is performed by source station and the frame specifies which route to follow by discovering frame by sending a specical frame called discovery frame, which spreads through the entire network using all possible paths to destination.

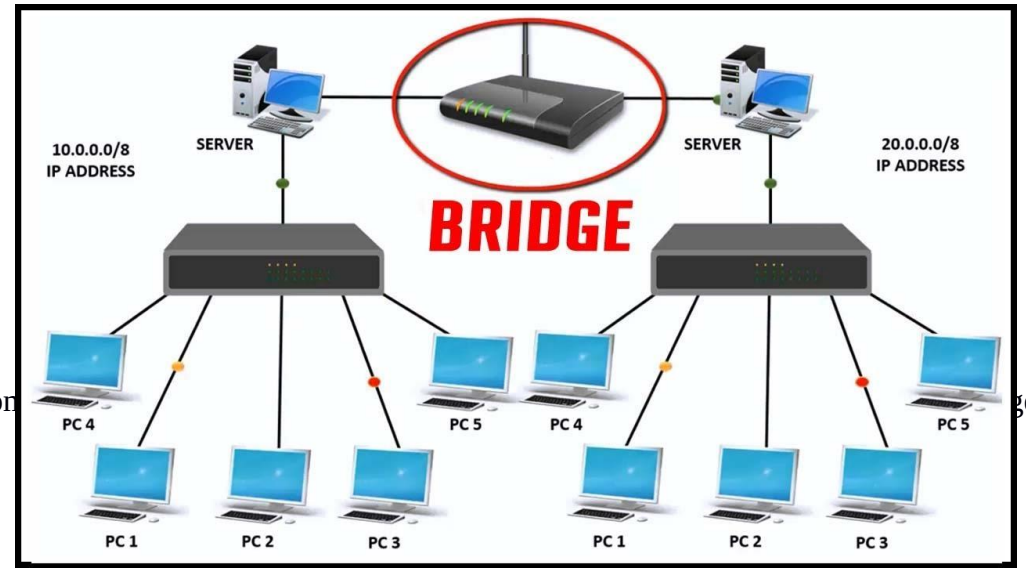


Figure 1.21 : A Bridge connecting similar networks

1.11.8 Gateway

A gateway, as the name suggests, is a passage to connect two networks together that may work upon different networking models. They basically works as the messenger agents that take data from one system, interpret it, and transfer it to another system. Bridge connects two or more similar networks where as Gateways connect two or more different networks. Gateways are also called protocol converters and can operate at any network layer. Gateways are generally more complex than switch or router.

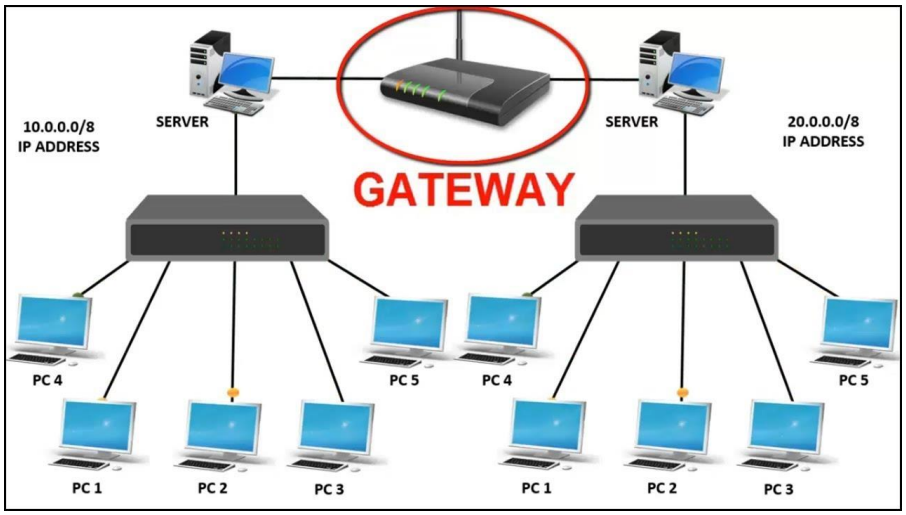


Figure 1.21 : A Bridge connecting similar networks

1.11.9 Brouter

It is also known as bridging router is a device which combines features of both bridge and router. It can work either at data link layer or at network layer. Working as router, it is capable of

routing packets across networks and working as bridge, it is capable of filtering local area network traffic.

1.12 Types of Data Transmission

Data Transmission can happen in two ways. Either in Synchronous or in Asynchronous Transmission. In Synchronous Transmission, data is sent in form of blocks or frames. This transmission is the full duplex type. Between sender and receiver the synchronization is compulsory. In Synchronous transmission, There is no gap present between data. It is more efficient and more reliable than asynchronous transmission to transfer the large amount of data. In Asynchronous Transmission, data is sent in the form of byte or character. This transmission is the half duplex type transmission. In this transmission start bits and stop bits are added with data. It does not require synchronization.

Difference between two types are given below.

Synchronous Transmission	Asynchronous Transmission
In Synchronous transmission, Data is sent in form of blocks or frames.	In asynchronous transmission, Data is sent in form of byte or character.
Synchronous transmission is fast.	Asynchronous transmission is slow.
Synchronous transmission is costly.	Asynchronous transmission economical.
In Synchronous transmission, time interval of transmission is constant.	In asynchronous transmission, time interval of transmission is not constant, it is random.
In Synchronous transmission, There is no gap present between data.	In asynchronous transmission, There is present gap between data.

1.13 Asynchronous Transfer Mode (ATM) in Computer Network

It is an International Telecommunication Union- Telecommunications Standards Section (ITU-T) efficient for call relay and it transmits all information including multiple service types such as data, video or voice which is conveyed in small fixed size packets called cells. Cells are transmitted asynchronously and the network is connection oriented.

ATM is a technology which has some event in the development of broadband ISDN in 1970s and 1980s, which can be considered an evolution of packet switching. *Each cell is 53 bytes*

long – 5 bytes header and 48 bytes payload. Making an ATM call requires first sending a message to set up a connection.

Subsequently all cells follow the same path to the destination. It can handle both constant rate traffic and variable rate traffic. Thus it can carry multiple types of traffic with **end-to-end** quality of service. ATM is independent of transmission medium, they maybe sent on a wire or fiber by themselves or they may also be packaged inside the payload of other carrier systems. ATM networks use "Packet" or "cell" Switching with virtual circuits. It's design helps in the implementation of high performance multimedia networking.

1.13.1 Why ATM networks?

1. Driven by the integration of services and performance requirements of both telephony and data networking : "broadband integrated service vision" (B-ISON).
2. Telephone networks support a single quality of service and is expensive to boot.
3. Internet supports no quality of service but is flexible and cheap.
4. ATM networks were meant to support a range of service qualities at a reasonable cost-intended to subsume both the telephone network and the Internet.

1.13.2 ATM vs DATA Networks (Internet) –

- ATM is a "virtual circuit" based: the path is reserved before transmission. While, Internet Protocol (IP) is connectionless and end-to-end resource reservations not possible. RSVP is a new signaling protocol in the internet.
- ATM Cells: Fixed or small size and Tradeoff is between voice or data. While, IP packets are of variable size.
- Addressing: ATM uses 20-byte global NSAP addresses for signaling and 32-bit locally assigned labels in cells. While, IP uses 32-bit global addresses in all packets.

1.14 Data and Signal

What we want to transmit in network is Data. What we send is Signal. We represent data that we wish to send in the form of signals. It's called as encoding. Number of data bits we send per second is called as Data Rate. It is represented as bits per second or bps. Number of signal components we send per second is known as Baud Rate. It is represented as signal components per second. Signals are of two types – Analog and Digital. Anlog signal is one which takes any value among infinite set of values in any range. For example, when we speak, our voice is analog in

nature. We can speak at lowest pitch to highest pitch. In this range, any value is valid. Digital signal is one that takes one of certain discrete values in the range. Signal takes any of the discrete value at any point of time. Data can be sent either in analog or digital form in networks. Digital signals are more robust compared to analog signal. A composite signal is a combination of two or more sinusoidal signal. When we extract individual signals from a composite signal, we get two or more periodic signals of different frequency, different amplitude etc. A digital signal is a composite analog signal. A digital signal, in the time domain, comprises connected vertical and horizontal line segments. A vertical line in the time domain means a frequency of infinity (sudden change in time). A horizontal line in the time domain means a frequency of zero (no change in time). Fourier analysis can be used to decompose a digital signal. If the digital signal is periodic, the decomposed signal has a frequency domain representation with an infinite bandwidth and discrete frequencies. If the digital signal is non-periodic, the decomposed signal has a frequency domain representation with an infinite bandwidth and continuous frequencies.

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

$$Bit\ Length = Propagation\ Speed\ X\ Bit\ Duration$$

There exists two methods of transmitting a digital signal. Baseband and Broadband. Baseband transmission means sending a digital signal over a channel without changing the digital signal to an analog signal. Baseband transmission requires that we have a low-pass channel. Low-pass channel means a channel with a bandwidth that starts from zero. For example, we can have a dedicated medium with a bandwidth constituting only one channel.

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

1.15 TRANSMISSION IMPAIRMENT

Signals travel through transmission media, which are not perfect. The imperfection causes signal-impairment.This means that signal at beginning of the medium is not the same as the signal at end of medium. What is sent is not what is received. There are 3 causes of impairment are

1) Attenuation

2) Distortion &

3) Noise.

Attenuation : As signal travels through the medium, its strength decreases as distance increases. This is called attenuation. As the distance increases, attenuation also increases.For example: Voice-data becomes weak over the distance & loses its contents beyond a certain distance. To compensate for this loss, amplifiers are used to amplify the signal. The decibel (dB) measures the relative strengths of 2 signals or one signal at 2 different points. The decibel is negative if a signal is attenuated. The decibel is positive if a signal is amplified.

$DB = 10\log_{10} (p2/p1)$

Variables P 1 and P 2 are the powers of a signal at points 1 and 2, respectively. To show that a signal has lost or gained strength, engineers use the unit of decibel.

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

Noise : Noise is defined as an unwanted data.In other words, noise is the external energy that corrupts a signal. Due to noise, it is difficult to retrieve the original data/information.There are four types of noise

i) Thermal Noise : It is random motion of electrons in wire which creates extra signal not originally sent by transmitter.

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

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

iv) Impulse Noise : Impulse Noise is a spike that comes from power-lines, lightning, and so on.

Signal-to-Noise Ratio (SNR) : SNR is used to find the theoretical bit-rate limit.SNR is defined as

ratio of Signal Power to that of Noise Power. SNR is actually the ratio of what is wanted (signal) to what is not wanted (noise). A high-SNR means the signal is less corrupted by noise.A low-SNR means the signal is more corrupted by noise. Because SNR is the ratio of 2 powers, it is often described in decibel units, SNR dB , defined as

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

1.16 DATA RATE LIMITS

Data-rate depends on 3 factors : Bandwidth available, Level of the signals and Quality of channel (the level of noise). Two theoretical formulas can be used to calculate the data-rate:

1) Nyquist for a noiseless channel and

2) Shannon for a noisy channel.

1.16.1 Noiseless Channel: Nyquist Bit Rate

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

$$BitRate = 2 \text{ X } Bandwidth \text{ X } \log_2 L$$

where Bandwidth = bandwidth of the channel

L = number of signal-levels used to represent data

BitRate = bitrate of channel in bps

According to the formula,

- By increasing number of signal-levels, we can increase the bit-rate.
- Although the idea is theoretically correct, practically there is a limit.
- When we increase the number of signal-levels, we impose a burden on the receiver.
- If no. of levels in a signal is 2, the receiver can easily distinguish b/w 0 and 1.
- If no. of levels is 64, the receiver must be very sophisticated to distinguish b/w 64 different levels.
- In other words, increasing the levels of a signal reduces the reliability of the system.

1.16.2 Noisy Channel: Shannon Capacity

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

$$Capacity = Bandwidth \text{ X } \log_2 (1+SNR)$$

where Bandwidth = bandwidth of channel in bps.

SNR = Signal-to-Noise ratio and

Capacity = Capacity of channel in bps.

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

1.17 PERFORMANCE

We evaluate the performance of a network on certain parameters. Based on that, we can tell which is a better network option for different applications. Following are the parameters.

i) Bandwidth

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

a) Bandwidth of an Analog Signal (in hz) : Bandwidth of an analog signal is expressed in terms of its frequencies. Bandwidth is defined as the range of frequencies that the channel can carry. It is calculated by the difference b/w the maximum frequency and the minimum frequency. For example, when the channel allows signal of a minimum frequency of F1 = 1000Hz and maximum frequency of F2 = 5000Hz, the bandwidth is given by F2 – F1= 5000 - 1000 = 4000 Hz.

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

Relationship between (a) and (b)

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

ii) Throughput

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

1) The bandwidth is a potential measurement of a link.

2) The throughput is an actual measurement of how fast we can send data.

For example:

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

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

iii) Latency (Delay)

The latency defines how long it takes for an entire message to completely arrive at the destination from the time the first bit is sent out from the source.

Latency = Propagation Time + Transmission Time + Queing Time + Processing Time

a) Propagation Time : Propagation time is defined as the time required for a bit to travel from source to destination. Propagation time is given by

Distance / Propagation Speed

Propagation speed of electromagnetic signals depends on medium and frequency of the signal.

b) Transmission Time : The time required for transmission of a message depends on

→ size of the message and

→ bandwidth of the channel.

 The transmission time is given by

Message Size/Bandwidth

c) Queuing Time : Queuing-time is the time needed for each intermediate-device to hold the message before it can be processed(Intermediate device may be a router or a switch). The queuing-time is not a fixed factor. This is because

i) Queuing-time changes with the load imposed on the network.

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

 An intermediate-device

→ queues the arrived messages and

→ processes the messages one by one.

 If there are many messages, each message will have to wait.

d) Processing Delay

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

iv) Bandwidth Delay Product

Two performance-metrics of a link are 1) Bandwidth and 2) Delay

The bandwidth-delay product is very important in data-communications.

• Let us elaborate on this issue, using 2 hypothetical cases as examples.

Let us assume,

Bandwidth of the link = 1 bps

Delay of the link = 5s.

Bandwidth-Delay product is 1 x 5 = 5. Thus, there can be maximum 5 bits on the line.

There can be no more than 5 bits at any time on the link.

Let us assume,

Bandwidth of the link = 4 bps

Delay of the link = 5s.

Bandwidth-Delay product is 5 x 5 = 25. Thus, there can be maximum 25 bits on the line.

At each second, there are 5 bits on the line, thus the duration of each bit is 0.20s.

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

v) Jitter

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

• For example:

If the delay for the first packet is 20 ms

the delay for the second is 45 ms and

the delay for the third is 40 ms

then the real-time application that uses the packets suffers from jitter.

Means, difference in delay exhibited by different packet of same message is known as Jitter.

Module 2

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

The possible encodings are:

1) Digital data to digital signal
2) Digital data to analog signal
3) Analog data to digital signal
4) Analog data to analog signal

2.1 Digital Data to Digital Signal : LINE CODING

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

We all believe that we transmit data in computer networks. In fact, we transmit signals. Data and signals are different. We represent data as signals and then these signals are sent along the medium. Means Data is just assumption. What we want to send is data but what we send is signal. Number of data bits we send in a second is called bit rate. But, number of signal components we send in a second is baud rate. We represent a bit or sequence of bits with a signal component. The process of representing data bits as signal components at source is called as encoding. The reverse process happen at the destination where databits are built using signal components received which is called as decoding. Data as well as signals that represents data can either be digital or analog. Line coding is the process of converting digital data to digital signals. By this technique we converts a sequence of bits to a digital signal.

We can roughly divide line coding schemes into five categories:

1. Unipolar (eg. NRZ scheme).
2. Polar (eg. NRZ-L, NRZ-I, RZ, and Biphase – Manchester and differential Manchester).
3. Bipolar (eg. AMI and Pseudoternary).
4. Multilevel
5. Multitransition

But, before learning difference between first three schemes we should first know the **characteristic** of these line coding techniques:

- There should be **self-synchronizing** i.e., both receiver and sender clock should be synchronized.
- There should have some error-detecting capability.
- There should be immunity to noise and interference.
- There should be less complexity.
- There should be no low frequency component (**DC-component**) as long distance transfer is not feasible for low frequency component signal.
- There should be less base line wandering.

Different characteristics of digital signal are

a) Signal Element Vs Data Element

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

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

b) Data Rate Vs Signal Rate

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

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

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

where N = data-rate (in bps)

c = case factor, which varies for each case

S = number of signal-elements and

r = previously defined factor.

This relationship depends on

→ value of r.

→ data pattern.

(If we have a data pattern of all 1s or all 0s, the signal-rate may be different from a data pattern of alternating Os and 1s).

c) Bandwidth

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

The bandwidth refers to range of frequencies used for transmitting a signal. Relationship b/w baud rate (signal-rate) and the bandwidth (range of frequencies) is given as

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

where N = data-rate (in bps)

c = case factor, which varies for each case

r = previously defined factor

B_{min} = minimum bandwidth

d) Baseline Wandering

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

e) DC Components

When the voltage-level in a digital-signal is constant for a while, the spectrum creates very low frequencies. These frequencies around zero are called DC (direct-current) components. DC

components present problems for a system that cannot pass low frequencies. For example, Telephone line cannot pass frequencies below 200 Hz. For Telephone systems, we need a scheme with no DC component.

f) Built-in Error Detection

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

g) Self-synchronization

To correctly interpret the signals received from the sender, the receiver's bit intervals must correspond exactly to the sender's bit intervals. If the receiver clock is faster or slower, the bit intervals are not matched and the receiver might misinterpret the signals. A self-synchronizing digital-signal includes timing-information in the data being transmitted. This can be achieved if there are transitions in the signal that alert the receiver to the beginning, middle, or end of the pulse. If the receiver's clock is out-of-synchronization, these points can reset the clock.

h) Immunity to Noise and Interference

The code should be immune to noise and other interferences.

i) Complexity

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

2.1.1 Unipolar-Scheme –

In this scheme, all the signal levels are either above or below the axis.

- **Non return to zero (NRZ) –** It is unipolar line coding scheme in which positive voltage defines bit 1 and the zero voltage defines bit 0. Signal does not return to zero at the middle of the bit thus it is called NRZ. For example: Data = 10110.

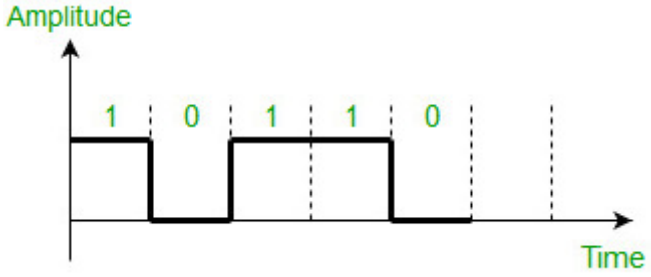


Fig. 2.1 Unipolar NRZ representation for the data 10110

But this scheme uses more power as compared to polar scheme to send one bit per unit line resistance. Moreover for continuous set of zeros or ones there will be self-synchronization and base line wandering problem.

2.1.2 Polar-Schemes –

In polar schemes, the voltages are on the both sides of the axis.

- **NRZ-L and NRZ-I –** These are somewhat similar to unipolar NRZ scheme but here we use two levels of amplitude (voltages). For **NRZ-L(NRZ-Level)**, the level of the voltage determines the value of the bit, typically binary 1 maps to logic-level high, and binary 0 maps to logic-level low, and for **NRZ-I(NRZ-Invert)**, two-level signal has a transition at a boundary if the next bit that we are going to transmit is a logical 1, and does not have a transition if the next bit that we are going to transmit is a logical 0.

  Note – For NRZ-I we are assuming in the example that previous signal before starting of data set "01001110" was positive. Therefore, there is no transition at the beginning and first bit "0" in current data set "01001110" is starting from +V. Example: Data = 01001110.

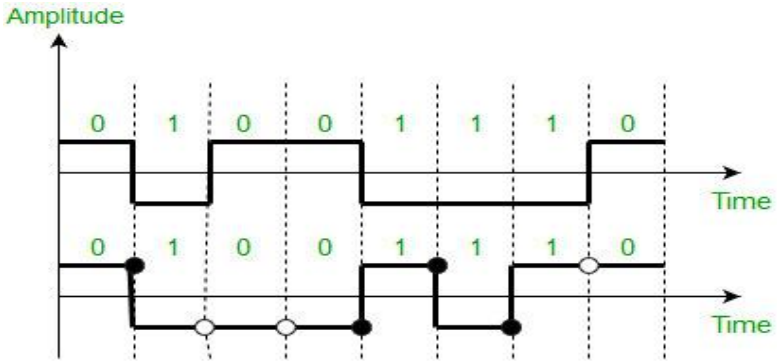


Fig. 2.2 : Polar NRZ-L and NRZ-I for the data 01001110

Comparison between NRZ-L and NRZ-I: Baseline wandering is a problem for both of them, but for NRZ-L it is twice as bad as compared to NRZ-I. This is because of transition at the boundary for NRZ-I (if the next bit that we are going to transmit is a logical 1). Similarly self-synchronization problem is similar in both for long sequence of 0's, but for long sequence of 1's it is more severe in NRZ-L.

- **Return to zero (RZ) –** One solution to NRZ problem is the RZ scheme, which uses three values positive,negative,and zero. In this scheme signal goes to 0 in the middle of each bit. **Note –** The logic we are using here to represent data is that for bit 1 half of the signal is represented by +V and half by zero voltage and for bit 0 half of the signal is represented by -V and half by zero voltage. Example: Data = 01001.

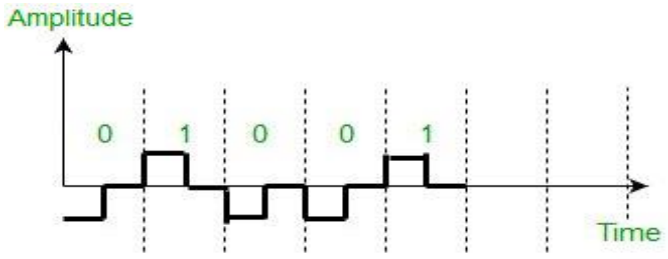


Fig. 2.3 : Polar RZ on data 01001

Main disadvantage of RZ encoding is that it requires greater bandwidth. Another problem is the complexity as it uses three levels of voltage. As a result of all these deficiencies, this scheme is not used today. Instead, it has been replaced by the better-performing Manchester and differential Manchester schemes.

2.1.3 Biphase (Manchester and Differential Manchester) – Manchester encoding is somewhat combination of the RZ (transition at the middle of the bit) and NRZ-L schemes. The duration of the bit is divided into two halves. The voltage remains at one level during the first half and moves to the other level in the second half. The transition at the middle of the bit provides synchronization.

Differential Manchester is somewhat combination of the RZ and NRZ-I schemes. There is always a transition at the middle of the bit but the bit values are determined at the beginning of the bit. If the next bit is 0, there is a transition, if the next bit is 1, there is no transition.

Note –

1. The logic we are using here to represent data using Manchester is that for bit 1 there is transition form -V to +V volts in the middle of the bit and for bit 0 there is transition from +V to -V volts in the middle of the bit.

2. For differential Manchester we are assuming in the example that previous signal before starting of data set “010011” was positive. Therefore there is transition at the beginning and first bit “0” in current data set “010011” is starting from -V. Example: Data = 010011.

The Manchester scheme overcomes several problems associated with NRZ-L, and differential Manchester overcomes several problems associated with NRZ-I as there is no baseline wandering and no DC component because each bit has a positive and negative voltage contribution.

Only limitation is that the minimum bandwidth of Manchester and differential Manchester is twice that of NRZ.

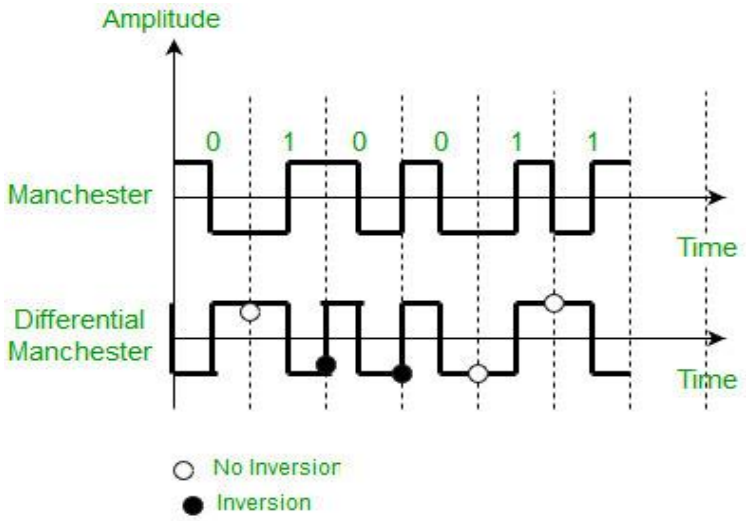


Fig. 2.4 : Bi Phase Manchester and Differential Manchester Schemes on data 010011

2.1.4 Bipolar-Schemes

In this scheme there are three voltage levels positive, negative, and zero. The voltage level for one data element is at zero, while the voltage level for the other element alternates between positive and negative.

- **Alternate Mark Inversion (AMI) –** A neutral zero voltage represents binary 0. Binary 1's are represented by alternating positive and negative voltages.
- **Pseudoternary** – Bit 1 is encoded as a zero voltage and the bit 0 is encoded as alternating positive and negative voltages i.e., opposite of AMI scheme. Example: Data = 010010.

The bipolar scheme is an alternative to NRZ.This scheme has the same signal rate as NRZ,but there is no DC component as one bit is represented by voltage zero and other alternates every time.

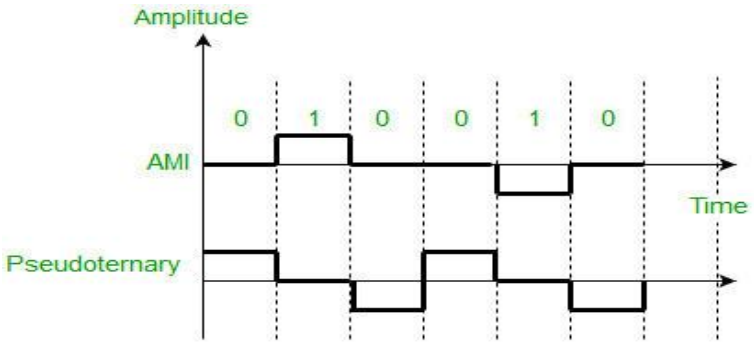


Fig 2.5 : BiPolar AMI & Pseudoternary Schemes on data 010010

2.2 ANALOG TO DIGITAL CONVERSION

An analog-signal may be created by a microphone or camera. To change an analog-signal to digital-data, we use PCM (pulse code modulation).After the digital-data are created (digitization), then we convert the digital-data to a digital-signal.

2.2.1 PCM (Pulse Code Modulation)

PCM is a technique used to change an analog signal to digital data (digitization). PCM has encoder at the sender and decoder at the receiver. The encoder has 3 processes,

a) Sampling

b) Quantization

c) Encoding

a) Sampling

We convert the continuous time signal (analog) into the discrete time signal (digital). Pulses from the analog-signal are sampled every T s sec, where T s is the sample-interval or period. The inverse of the sampling-interval is called the sampling-frequency (or sampling-rate). Sampling-frequency is given by

$$f_s = 1/t_s$$

There are three sampling methods:

1) Ideal Sampling

This method is difficult to implement.

2) Natural Sampling

A high-speed switch is turned ON for only the small period of time when the sampling occurs. The result is a sequence of samples that retains the shape of the analog-signal.

3) Flat Top Sampling

This is the most common sampling method creates flat-top samples.This method is sometimes referred to as PAM (pulse amplitude modulation).

Sampling Rate

According to Nyquist theorem,“The sampling-rate must be at least 2 times the highest frequency, not the bandwidth”.If the analog-signal is low-pass, the bandwidth and the highest frequency are the

same value. If the analog-signal is bandpass, the bandwidth value is lower than the value of the maximum frequency.

b) Quantization

The sampled-signal is quantized. Result of sampling is a set of pulses with amplitude-values b/w max & min amplitudes of the signal.There exists 4 steps in quantization, as below.

i) We assume that the original analog-signal has amplitudes between V min & V max .

ii) We divide the range into L zones, each of height Δ(delta).

$$\Delta(\text{delta}) = (V_{max} - V_{min}) / L$$ where L is the number of levels.

iii) We assign quantized values of 0 to (L-1) to the midpoint of each zone.

iv) We approximate the value of the sample amplitude to the quantized values.

• For example: Let V min =-20

V max =+20 V

let L = 8 Therefore, Δ = [+20-(-20)]/8= 5 V

• In the case,

1) First row is normalized-PAM-value for each sample.

2) Second row is normalized-quantized-value for each sample.

3) Third row is normalized error (which is diff. b/w normalized PAM value & quantized values).

4) Fourth row is quantization code for each sample.

5) Fifth row is the encoded words (which are the final products of the conversion).

Quantization Level

Let L = number of levels.

The choice of L depends on

→ range of the amplitudes of the analog-signal and

→ how accurately we need to recover the signal.

• If the signal has only 2 amplitude values, we need only 2 quantization-levels. If the signal (like voice) has many amplitude values, we need more quantization-levels.

• In audio digitizing, L is normally chosen to be 256. In video digitizing, L is normally thousands.

• Choosing lower values of L increases the quantization-error.

Quantization Error

• Quantization-error is the difference b/w normalized PAM value & quantized values. Quantization is an approximation process. The input values to the quantizer are the real values.The output values from the quantizer are the approximated values.The output values are chosen to be the middle value

in the zone.If the input value is also at the middle of the zone,then, there is no error. Otherwise, there is an error.

In the previous example,the normalized amplitude of the third sample is 3.24, but the normalized quantized value is 3.50. This means that there is an error of +0.26.

Uniform vs. Non Uniform Quantization

Nonuniform quantization can be done by using a process called companding and expanding.

1) The signal is companded at the sender before conversion.

2) The signal is expanded at the receiver after conversion.

Companding means reducing the instantaneous voltage amplitude for large values. Expanding means increasing the instantaneous voltage amplitude for small values. It has been proved that non-uniform quantization effectively reduces the SNR_{dB} of quantization.

c) Encoding

• The quantized values are encoded as n-bit code word. In the previous example, a quantized value 2 is encoded as 010, a quantized value 5 is encoded as 101. Relationship between number of quantization-levels (L) & number of bits (n) is given by

$n=\log_2 L$

or

$2^n = L$

The bit-rate is given by

Bit Rate + Sampling Rate X Number of Bits per Sample = f_s X n

PCM Bandwidth

The minimum bandwidth of a line-encoded signal is $B_{min} = c \times N \times (1/r)$.

When 1/r = 1 (for a NRZ or bipolar signal) and c = (1/2) (the average situation), the minimum bandwidth is $B_{min} = n_b \times B_{analog}$.

This means the minimum bandwidth of the digital-signal is n_b times greater than the bandwidth of the analog-signal.

Maximum Data Rate of a Channel

The Nyquist theorem gives the data-rate of a channel as $N_{max} = 2 \times B \times \log_2 L$

We can deduce above data-rate from the Nyquist sampling theorem by using the following arguments.

i) We assume that the available channel is low-pass with bandwidth B.

ii) We assume that the digital-signal we want to send has L levels, where each level is a signal-element. This means r = $1/\log_2 L$.

iii) We first pass digital-signal through a low-pass filter to cut off the frequencies above B Hz.

iv) We treat the resulting signal as an analog-signal and sample it at 2 x B samples per second and quantize it using L levels.

5) The resulting bit-rate is $N=f_s \text{ X } n_b = 2 \text{ X } B \text{ X } \log_2 L$.

This is the maximum bandwidth; if the case factor c increases, the data-rate is reduced.

$$N_{max} = 2 \text{ X } B \text{ X } \log_2 L \text{ bps.}$$

Minimum Required Bandwidth

The previous argument can give us the minimum bandwidth if the data-rate and the number of signal-levels are fixed. We can say

$$B_{min} = N/(2 \text{ X } \log_2 L) \text{ Hz}$$

2.3 TRANSMISSION MODES

There are two ways by which we can transmit data over a link. i) Parallel mode. ii) Serial mode. Under serial, there are 3 sub ways. Asynchronous, Synchronous and Isochronous.

2.3.1 PARALLEL TRANSMISSION

Multiple bits are sent with each clock-tick.'n' bits in a group are sent simultaneously.'n' wires are used to send 'n' bits at one time.Each bit has its own wire.Typically, the 8 wires are bundled in a cable with a connector at each end.

Advantage: Parallel transmission can increase the transfer speed by a factor of n over serial transmission.

• Disadvantage: Parallel transmission requires n communication lines just to transmit the data-stream. Because this is expensive, parallel transmission is usually limited to short distances.

2.3.2 SERIAL TRANSMISSION

Here, one bit is sent with each clock-tick using only a single link.

Advantage: Serial transmission reduces cost of transmission over parallel by a factor of n.

Disadvantage:Since communication within devices is parallel, following 2 converters are required at interface:i) Parallel-to-serial converter and ii) Serial-to-parallel converter.

There are 3 types of serial transmission: asynchronous, synchronous, and isochronous.

2.3.2.1 Asynchronous Transmission

Asynchronous transmission is so named because the timing of a signal is not important. Prior to data transfer, both sender & receiver agree on pattern of information to be exchanged. Normally, patterns are based on grouping the bit-stream into bytes. The sender transmits each group to the link without regard to a timer. As long as those patterns are followed, the receiver can retrieve the information without regard to a timer.There may be a gap between bytes.We send

→ 1 start bit (0) at the beginning of each byte.

→ 1 stop bit (1) at the end of each byte.

Start bit alerts the receiver to the arrival of a new group. Stop bit lets the receiver know that the byte is finished. Here, the term asynchronous means "asynchronous at the byte level". However, the bits are still synchronized & bit-durations are the same.

Advantages:

1) Cheap & effective.

2) Useful for low-speed communication.

Disadvantage:

1) Slower than synchronous transmission. (Because of stop bit, start bit and gaps)

2.3.2.2 Synchronous Transmission

We send bits one after another without start or stop bits or gaps. The receiver is responsible for grouping the bits. The bit-stream is combined into longer "frames," which may contain multiple bytes. If the sender wants to send data in separate bursts, the gaps between bursts must be filled with a special sequence of 0s & 1s (that means idle).

Advantages:

1) Speed: Faster than asynchronous transmission. ('.' of no stop bit, start bit and gaps).

2) Useful for high-speed applications such as transmission of data from one computer to another.

2.3.2.3 Isochronous

Synchronization between characters is not enough; the entire stream of bits must be synchronized. The isochronous transmission guarantees that the data arrive at a fixed rate. In real-time audio/video, jitter is not acceptable. Therefore, synchronous transmission fails.For example: TV

images are broadcast at the rate of 30 images per second. The images must be viewed at the same rate.

2.4 DIGITAL TO ANALOG CONVERSION

Digital-to-analog conversion is the process of changing one of the characteristics of an analog-signal based on the information in digital-data. A sine wave can be defined by 3 attributes,

a) Amplitude

b) Frequency &

c) Phase.

When anyone of the 3 attributes of a wave is varied, a different version of the wave will be created. So, by changing one attribute of an analog signal, we can use it to represent digital-data. Four methods of digital to analog conversion are

i) Amplitude shift keying (ASK)

ii) Frequency shift keying (FSK)

iii) Phase shift keying (PSK)

iv) Quadrature amplitude modulation (QAM).

QAM is a combination of ASK and PSK i.e. QAM combines changing both the amplitude and phase. I is the most efficient of these 4 methods. QAM is the method commonly used today.

2.4.1 Amplitude Shift Keying (ASK)

The amplitude of the carrier-signal is varied to represent different signal-elements. Both frequency and phase remain constant for all signal-elements.

Binary ASK (BASK)

BASK is implemented using only 2 levels. This is also known as OOK (On-Off Keying). Implementation of BASK

Here, line coding unipolar NRZ method is used.

The unipolar NRZ signal is multiplied by the carrier-frequency coming from an oscillator.

1) When amplitude of the NRZ signal = 0, amplitude of the carrier-signal = 0.

2) When amplitude of the NRZ signal = 1, the amplitude of the carrier-signal is held.

Bandwidth for ASK

Here, the bandwidth (B) is proportional to the signal-rate (S). The bandwidth is given by

$$B=(1+d) \text{ X } S$$

where d(0<d<1)= this factor depends on modulation and filtering-process.

2.4.2 Frequency Shift Keying (FSK)

The frequency of the carrier-signal is varied to represent different signal-elements. The frequency of the modulated-signal is constant for the duration of one signal-element, but changes for the next signal-element if the data-element changes. Both amplitude and phase remain constant for all signal-elements.

Binary FSK (BFSK)

This uses 2 carrier-frequencies: f1 and f2.

1) When data-element = 1, first carrier frequency(f1) is used.

2) When data-element = 0, second carrier frequency(f2) is used.

Implementation

• Here, line coding method used = unipolar NRZ.

• Two implementations of BFSK: i) Coherent and ii) Coherent BFSK

Coherent BFSK	Non Coherent BFSK
he phase continues through the boundary of two signal-elements	There may be discontinuity in the phase when one signal-element ends and the next begins.
This is implemented by using one voltage-controlled oscillator (VCO). VCO changes frequency according to the input voltage.When the amplitude of NRZ signal = 0, the VCO	This is implemented by → treating BFSK as 2 ASK modulations and → using 2 carrier-frequencies
When the amplitude of NRZ signal = 0, the VCO keeps its regular frequency. When the amplitude of NRZ signal = 0, the VCO increases its frequency.	

Bandwidth for BFSK

FSK has two ASK signals, each with its own carrier-frequency f1 or f2.

• The bandwidth is given by $B= (1+d)XS + 2\Delta f$

where $2\Delta f$ is the difference between f1 and f2

2.4.3 Phase Shift Keying (PSK)

The phase of the carrier-signal is varied to represent different signal-elements. Both amplitude and frequency remain constant for all signal-elements.

Binary PSK (BPSK)

We have only two signal-elements:

1) First signal-element with a phase of 0°.

2) Second signal-element with a phase of 180°.

ASK vs. PSK

In ASK, the criterion for bit detection is the amplitude of the signal.

In PSK, the criterion for bit detection is the phase.

• Advantages:

1) PSK is less susceptible to noise than ASK.

2) PSK is superior to FSK because we do not need 2 carrier-frequencies.

• Disadvantage:

PSK is limited by the ability of the equipment to distinguish small differences in phase.

Implementation

The implementation of BPSK is as simple as that for ASK. The signal-element with phase 180° can be seen as the complement of the signal-element with phase 0°. Here, line coding method used-polar NRZ.The polar NRZ signal is multiplied by the carrier-frequency coming from an oscillator.

1) When data-element = 1, the phase starts at 0°.

2) When data-element = 0, the phase starts at 180°.

Bandwidth for BPSK

The bandwidth is the same as that for BASK, but less than that for BFSK. No bandwidth is wasted for separating 2 carrier-signals.

2.4.4 Quadrature PSK (QPSK)

The scheme is called QPSK because it uses 2 separate BPSK modulations.

1) First modulation is in-phase,

2) Second modulation is quadrature (out-of-phase).

• A serial-to-parallel converter

→ accepts the incoming bits

→ sends first bit to first modulator and

→ sends second bit to second modulator.

• The bit to each BPSK signal has one-half the frequency of the original signal.

• Advantages:

1) Decreases the baud rate.

2) Decreases the required bandwidth.

The 2 composite-signals created by each multiplier are 2 sine waves with the same frequency, but different phases. When the 2 sine waves are added, the result is another sine wave, with 4 possible phases: 45°, -45°, 135° and -135°. There are 4 kinds of signal-elements in the output signal (L=4), so we can send 2 bits per signal element (r=2).

2.4.5 Quadrature Amplitude Modulation (QAM)

This is a combination of ASK and PSK. Main idea: Using 2 carriers, one in-phase and the other quadrature, with different amplitude levels for each carrier.

There are many variations of QAM.

Bandwidth for QAM

• The bandwidth is same as in ASK and PSK transmission.

• QAM has the same advantages as PSK over ASK.

Module 3

In real life, we have links with limited bandwidths. The wise use of these bandwidths has been one of the main challenges of electronic communications. However, the meaning of *wise* may depend on the application. Sometimes it is necessary to combine several low-bandwidth channels to make use of one channel with a larger bandwidth. Sometimes it is necessary to expand the bandwidth of a channel to achieve goals such as privacy and antijamming. There are two broad categories of bandwidth utilization: multiplexing and spectrum spreading. In multiplexing, the main goal is efficiency; it combines several channels into one. In spectrum spreading, its goals are privacy and antijamming.

MULTIPLEXING:

The set of techniques that allow the simultaneous transmission of multiple signals across a single data link is called ***Multiplexing***.

In a multiplexed system, *n* lines share the bandwidth of one link. Figure 3.1 shows the basic format of a multiplexed system. At the sending side, many lines direct their transmission streams to a **multiplexer (MUX),** which combines them into a single stream (many-to one).At the receiving end, that stream is fed into a **DE multiplexer (DEMUX),** which separates the stream back into its component transmissions (one-to-many) and directs them to their corresponding lines. In the figure, the word **link** refers to the physical path. The word **channel** refers to the portion of a link that carries a transmission between a given pair of lines. One link can have many (*n*) channels.

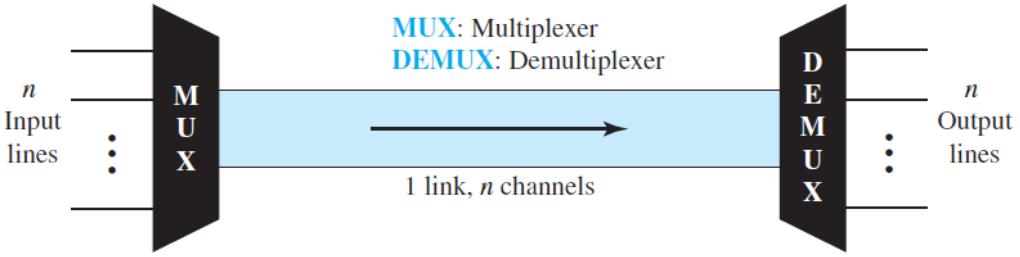


Fig.3.1 dividing a link into channels

There are three basic multiplexing techniques:

- *Frequency-division multiplexing,*
- *Wavelength-division multiplexing, and*
- *Time-division multiplexing*

The first two are techniques designed for analog signals, the third, for digital signals

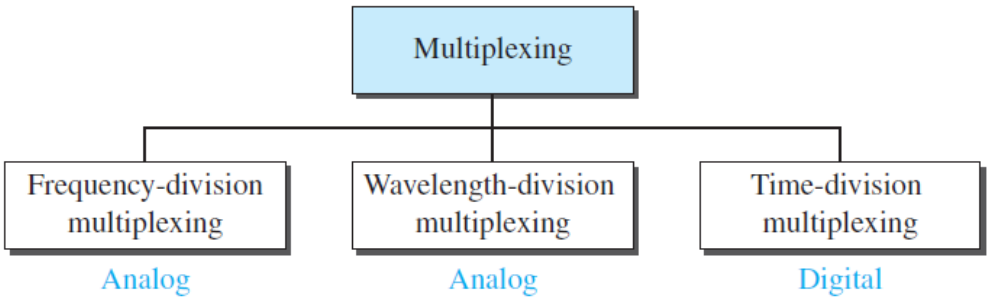


Fig.3.2 categories of multiplexing

frequency-division multiplexing(FDM):

FDM is an ***analog multiplexing technique*** that combines analog signals. It can be applied when the bandwidth of a link is greater than the combined bandwidths of the signals to be transmitted. In FDM, signals generated by each sending device modulate different carrier frequencies. These modulated signals are then combined into a single composite signal that can be transported by the link. Carrier frequencies are separated by sufficient bandwidth to accommodate the modulated signal. These bandwidth ranges are the channels which can be separated by strips of unused bandwidth—**guard bands**—to prevent signals from overlapping. Carrier frequencies must not interfere with the original data frequencies.

Figure 3.3 gives a conceptual view of FDM. In this illustration, the transmission path is divided into three parts, each representing a channel that carries one transmission.

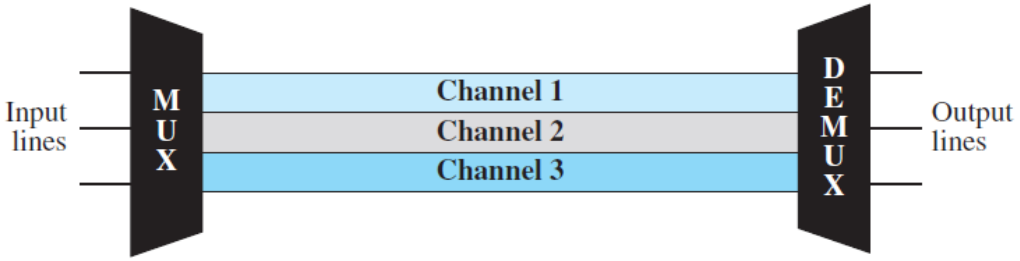


Fig.3.3 FDM

Multiplexing Process

Figure 3.4 is a conceptual illustration of the multiplexing process. Each source generates a signal of a similar frequency range. Inside the multiplexer, these similar signals modulate different carrier frequencies (f1, *f*2, and *f*3). The resulting modulated signals are then combined into a single composite signal that is sent out over a media link that has enough bandwidth to accommodate it.

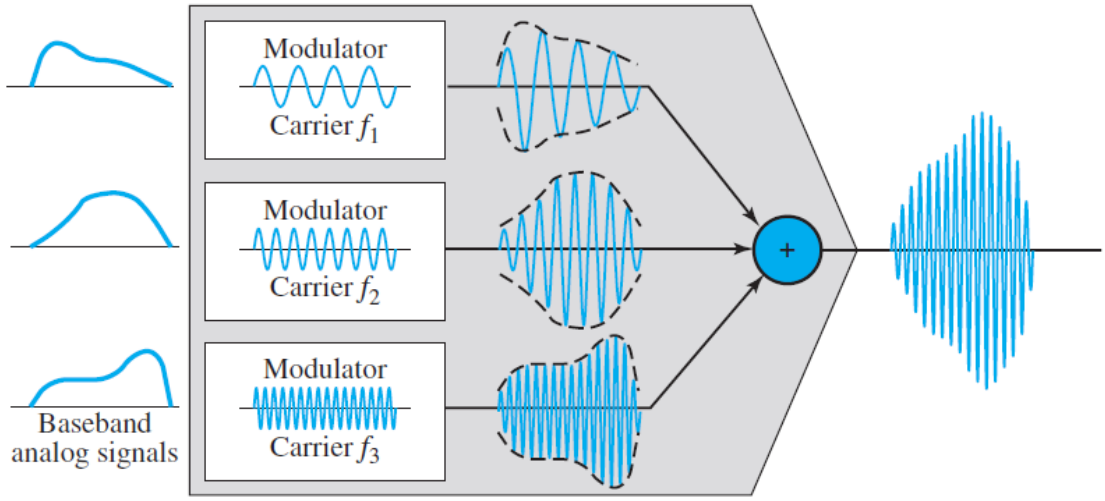


Fig.3.4 FDM Multiplexing example

Demultiplexing Process

The demultiplexer uses a series of filters to decompose the multiplexed signal into its constituent component signals. The individual signals are then passed to a demodulator that separates them from their carriers and passes them to the output lines. Figure 3.5 is a conceptual illustration of demultiplexing process.

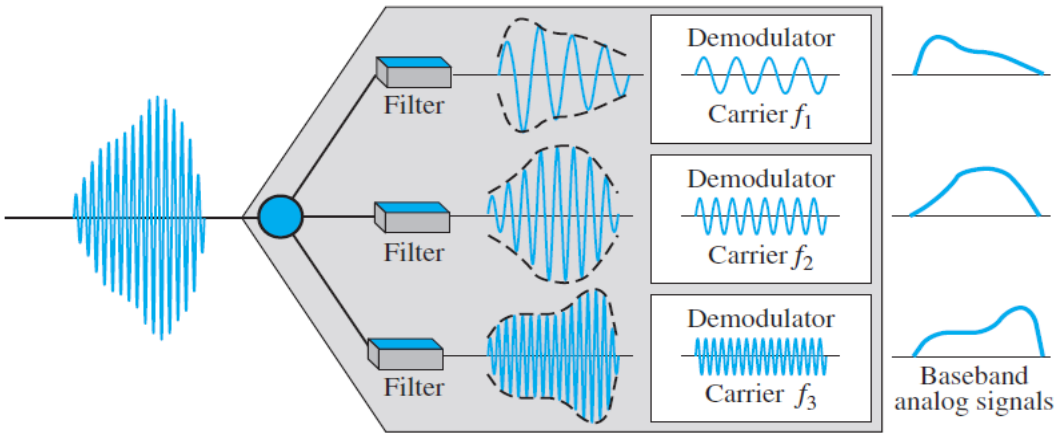


Fig.3.5 FDM Demultiplexing example

Other Applications of FDM

A very common application of FDM *is AM and FM radio broadcasting*. Radio uses the air as the transmission medium. A special band from 530 to 1700 kHz is assigned to AM radio. All radio stations need to share this band. However, FM has a wider band of 88 to 108 MHz because each station needs a bandwidth of 200 kHz.

Another common use of FDM is in *television broadcasting*. Each TV channel has its own bandwidth of 6 MHz.

The first generation of *cellular telephones* also uses FDM. Each user is assigned two 30-kHz channels, one for sending voice and the other for receiving. The voice signal, which has a bandwidth of 3 kHz is modulated by using FM.

Wavelength-Division Multiplexing(WDM):

WDM is an analog multiplexing technique to combine optical signals. Here, the multiplexing and demultiplexing involve optical signals transmitted through fiber-optic channels. It combines different signals of different frequencies. The difference is that the frequencies are very high. It is designed to use the high-data-rate capability of fiber-optic cable.

Figure 3.6 gives a conceptual view of a WDM multiplexer and demultiplexer. Very narrow bands of light from different sources are combined to make a wider band of light. At the receiver, the signals are separated by the demultiplexer.

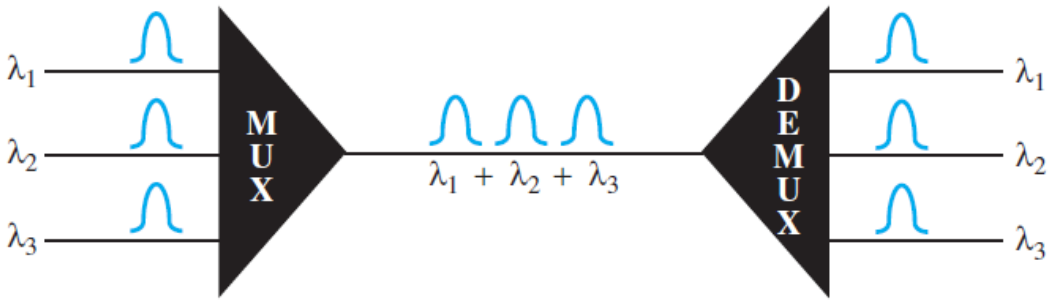


Fig.3.6 WDM

The basic idea of WDM is very simple. It combines multiple light sources into one single light at the multiplexer and do the reverse at the demultiplexer. The combining and splitting of light sources are easily handled by a prism. Using this technique, a multiplexer can be made to combine several input beams of light, each containing a narrow band of frequencies,

into one output beam of a wider band of frequencies. A demultiplexer can also be made to reverse the process. Figure 3.7 shows the concept.

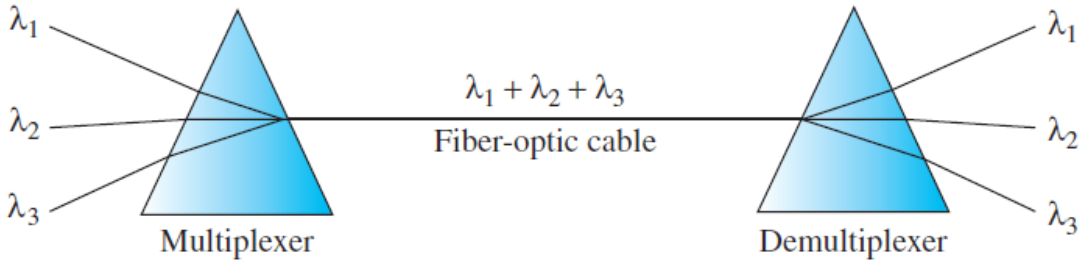


Fig 3.7 prisms in wavelength division multiplexing and demultiplexing

One application of WDM is the SONET network, in which multiple optical fiber lines are multiplexed and demultiplexed.

Time-Division Multiplexing:

Time-division multiplexing (TDM) is a digital multiplexing technique for combining several low-rate channels into one high-rate one. It allows several connections to share the high bandwidth of a link on time basis. Each connection occupies a portion of time in the link. Figure 3.8 gives a conceptual view of TDM. In the figure, portions of signals 1, 2, 3, and 4 occupy the link sequentially.

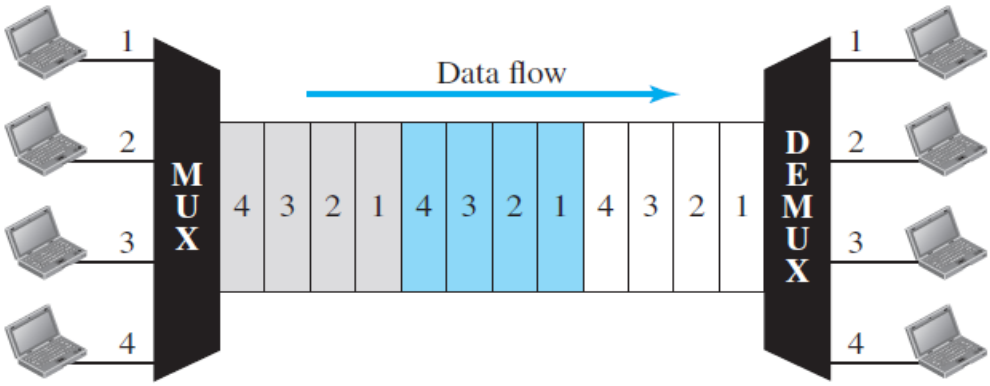


Fig 3.8 TDM

TDM can be divided into two different schemes: **Synchronous and Statistical.**

Synchronous TDM

In synchronous TDM, each input connection has an allotment in the output even if it is not sending data.

Time Slots and Frames

In synchronous TDM, the data flow of each input connection is divided into units, where each input occupies one input time slot. A unit can be 1 bit, one character, or one block of data. Each input unit becomes one output unit and occupies one output time slot. The duration of an output time slot is *n* times shorter than the duration of an input time slot. If an input time slot is *T* s, the output time slot is *T*/*n* s, where *n* is the number of connections. Figure 3.9 is an example of synchronous TDM where *n* is 3.

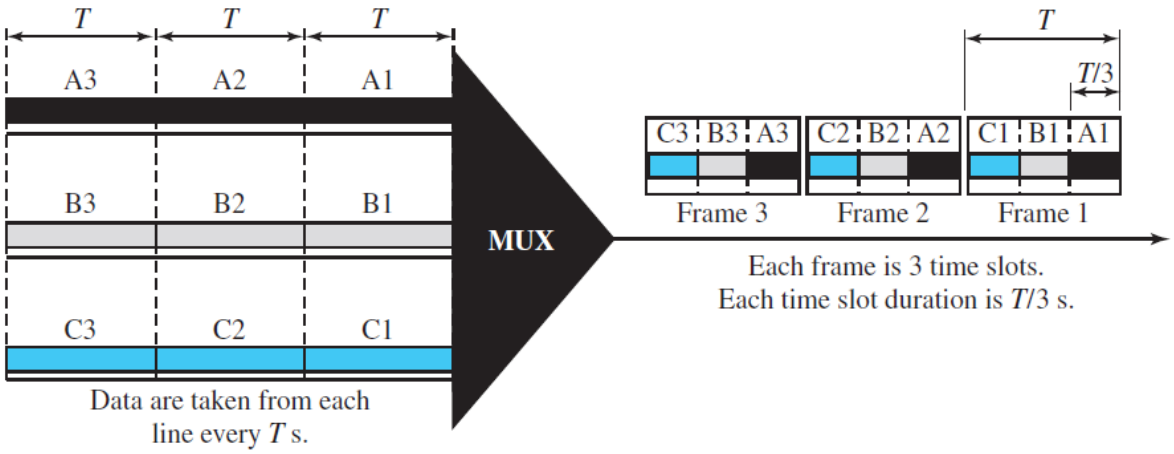


Fig 3.9 Synchronous TDM

In synchronous TDM, if there are *n* connections, a frame is divided into *n* time slots and one slot is allocated for each unit, one for each input line. If the duration of the input unit is *T*, the duration of each slot is *T*/*n* and the duration of each frame is *T*.

In synchronous TDM, the data rate of the link is *n* times faster, and the unit duration is *n* times shorter.

Interleaving

TDM can be visualized as two fast-rotating switches, one on the multiplexing side and the other on the demultiplexing side. The switches are synchronized and rotate at the same speed, but in opposite directions. On the multiplexing side, as the switch opens in front of a connection, that connection has the opportunity to send a unit onto the path. This process is called **interleaving.** On the demultiplexing side, as the switch opens in front of a connection, that connection has the opportunity to receive a unit from the path. Figure 3.10 shows the interleaving process for the connection shown in Figure 3.9.

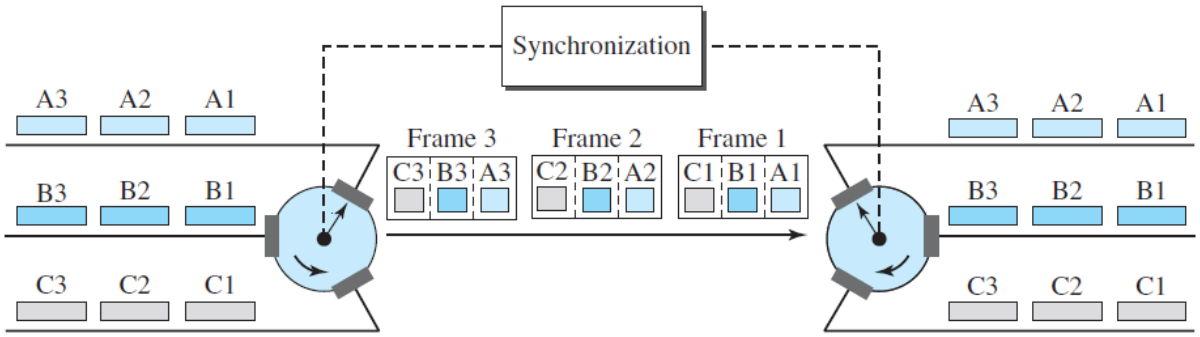


Fig 3.10 Interleaving

Data Rate Management

One problem with TDM is how to handle a disparity in the input data rates that means data rates of all input lines are not the same. Hence three strategies can be used. Three strategies are ***multilevel multiplexing, multiple-slot allocation,*** *and* ***pulse stuffing.***

Multilevel Multiplexing

Multilevel multiplexing is a technique used when the data rate of an input line is a multiple of others.

For example, in Figure 3.11, it has two inputs of 20 kbps and three inputs of 40 kbps.The first two input lines can be multiplexed together to provide a data rate equal to the last three. A second level of multiplexing can create an output of 160 kbps.

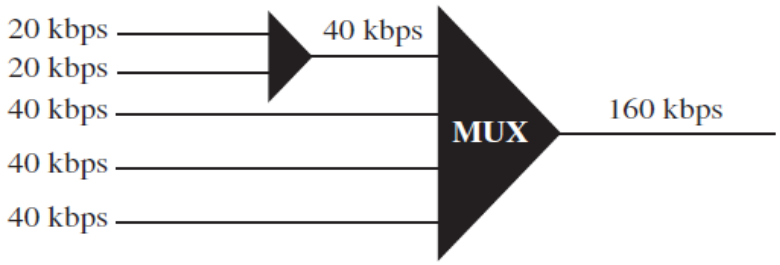


Fig 3.11 multilevel multiplexing

Multiple-Slot Allocation

It allots more than one slot in a frame to a single input line.

For example, if an input line has a data rate that is a multiple of another input. In Figure 3.12, the input line with a 50-kbps data rate can be given two slots in the output. Then a demultiplexer is inserted in the line to make two inputs out of one.

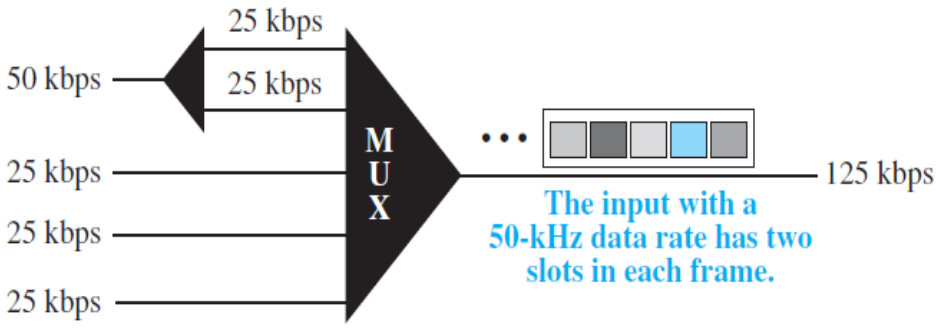


Fig 3.12 Multiple slot allocation

Pulse Stuffing

Sometimes the bit rates of sources are not multiple integers of each other. Therefore, neither of the above two techniques can be applied. One solution is to make the highest input data rate the dominant data rate and then add dummy bits to the input lines with lower rates. This will increase their rates. This technique is called ***pulse stuffing, bit padding***, or ***bit stuffing***. The idea is shown in Figure 3.13. The input with a data rate of 46 is pulse-stuffed to increase the rate to 50 kbps.

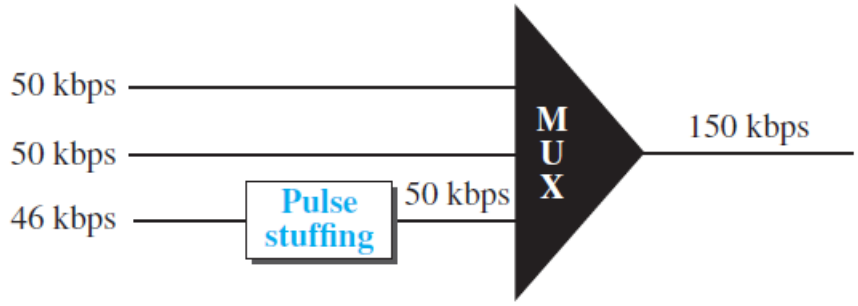


Fig 3.13 Pulse stuffing

Frame Synchronizing

Synchronization between the multiplexer and demultiplexer is a major issue. If the multiplexer and the demultiplexer are not synchronized, a bit belonging to one channel may be received by the wrong channel. For this reason, one or more synchronization bits are usually added to the beginning of each frame. These bits, called **framing bits that** allows the demultiplexer to synchronize with the incoming stream so that it can separate the time slots accurately. This synchronization information consists of 1 bit per frame, alternating between 0 and 1, as shown in Figure 3.14.

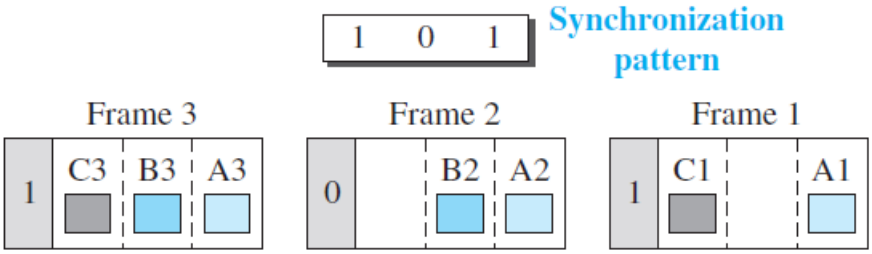


Fig 3.14 Framing bits

Statistical TDM

In synchronous TDM, each input has a reserved slot in the output frame. Hence it can be inefficient if some input lines have no data to send. This drawback of synchronous TDM can be overcome by statistical TDM, in which slots are dynamically allocated to improve bandwidth efficiency. In statistical multiplexing, the number of slots in each frame is less than the number of input lines. The multiplexer checks each input line in round robin fashion; it allocates a slot for an input line if the line has data to send; otherwise, it skips the line and checks the next line.

Figure 3.15 shows a synchronous and a statistical TDM example.

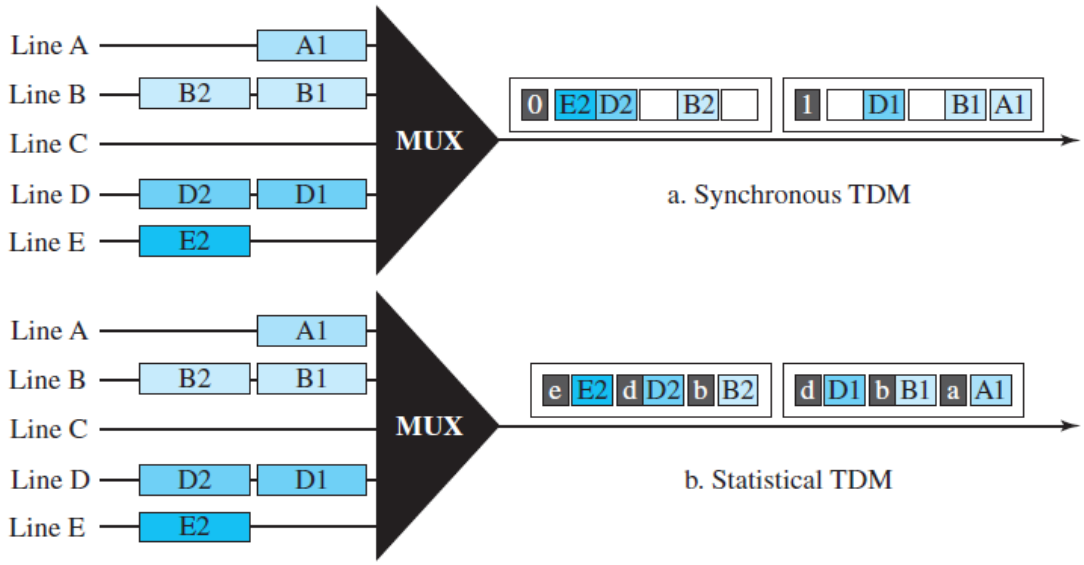


Fig 3.15 TDM slot comparison

Addressing

An output slot in synchronous TDM is totally occupied by data; hence there is no need for addressing. Synchronization and preassigned relationships between the inputs and outputs serve as an address.

In statistical TDM, a slot needs to carry ***data as well as the address*** of the destination. Here, there is no fixed relationship between the inputs and outputs because there are no preassigned or reserved slots. It is necessary to include the address of the receiver inside each slot to show where it is to be delivered. The addressing in its simplest form can be n bits to define N different output lines with $n = \log_2 N$.

Slot Size

Since a slot carries both data and an address in statistical TDM, the ratio of the data size to address size must be reasonable to make transmission efficient.

No Synchronization Bit

There is another difference between synchronous and statistical TDM, but this time it is at the frame level. The frames in statistical TDM need not be synchronized, so we do not need synchronization bits.

Bandwidth

In statistical TDM, the capacity of the link is normally less than the sum of the capacities of each channel. The designers of statistical TDM define the capacity of the link based on the statistics of the load for each channel.

SPREAD SPECTRUM:

Spread spectrum is designed to be used in wireless applications (LANs and WANs). It also combines signals from different sources to fit into a larger bandwidth, but the goals are privacy and antijamming. To achieve these goals, spread spectrum techniques add redundancy; they spread the original spectrum needed for each station. If the required bandwidth for each station is B, spread spectrum expands it to B_{ss}, such that $B_{ss} >> B$. The expanded bandwidth allows the source to wrap its message in a protective envelope for a more secure transmission.

Figure 3.16 shows the idea of spread spectrum. Spread spectrum achieves its goals through two principles:

1. The bandwidth allocated to each station needs to be, by far, larger than what is needed. This allows redundancy.
2. The expanding of the original bandwidth B to the bandwidth B_{ss} must be done by a process that is independent of the original signal. In other words, the spreading process occurs after the signal is created by the source.

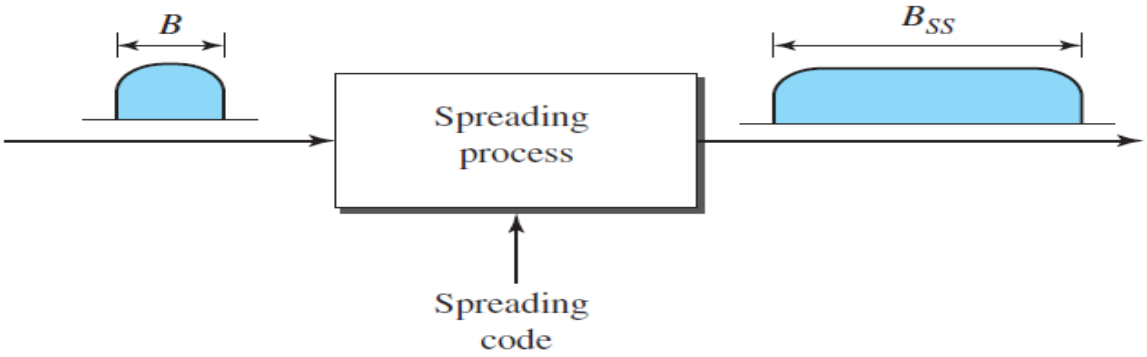


Fig.3.16 Spread spectrum

After the signal is created by the source, the spreading process uses a spreading code and spreads the bandwidth. The above figure shows the original bandwidth *B* and the spread bandwidth *B*SS. The spreading code is a series of numbers that look random but are a pattern. There are two techniques to spread the bandwidth: *Frequency hopping spread spectrum (FHSS) and Direct sequence spread spectrum (DSSS).*

Frequency hopping spread spectrum (FHSS)

The **frequency hopping spread spectrum (FHSS)** technique uses *M* different carrier frequencies that are modulated by the source signal. The modulation is done using one carrier frequency at a time, *M* frequencies are used in the long run. The bandwidth occupied by a source after spreading is $B_{FHSS} >> B$.

Figure 3.17 shows the general layout for FHSS. A **pseudorandom code generator,** called ***pseudorandom noise*** **(PN),** creates a *k*-bit pattern for every **hopping period** T_h. The frequency table uses the pattern to find the frequency to be used for this hopping period and passes it to the frequency synthesizer. The frequency synthesizer creates a carrier signal of that frequency, and the source signal modulates the carrier signal.

In this case, *M* is 8 and *k* is 3. The pseudorandom code generator will create eight different 3-bit patterns. These are mapped to eight different frequencies in the frequency table (see Figure 3.18). The pattern for this station is 101, 111, 001, 000, 010, 011, 100. Note that this pattern is pseudorandom.it is repeated after eight hopping.

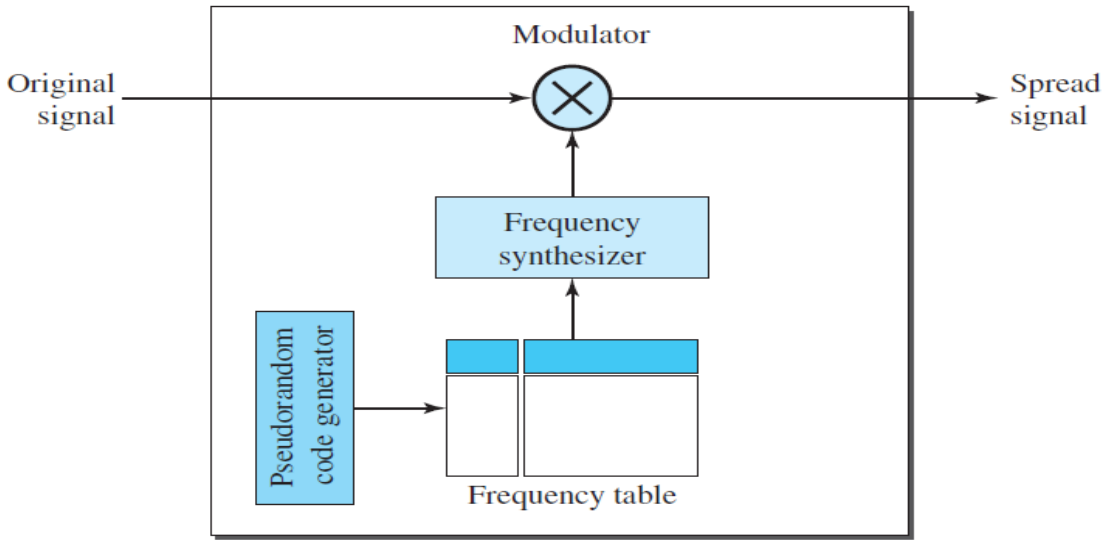


Fig.3.17 FHSS

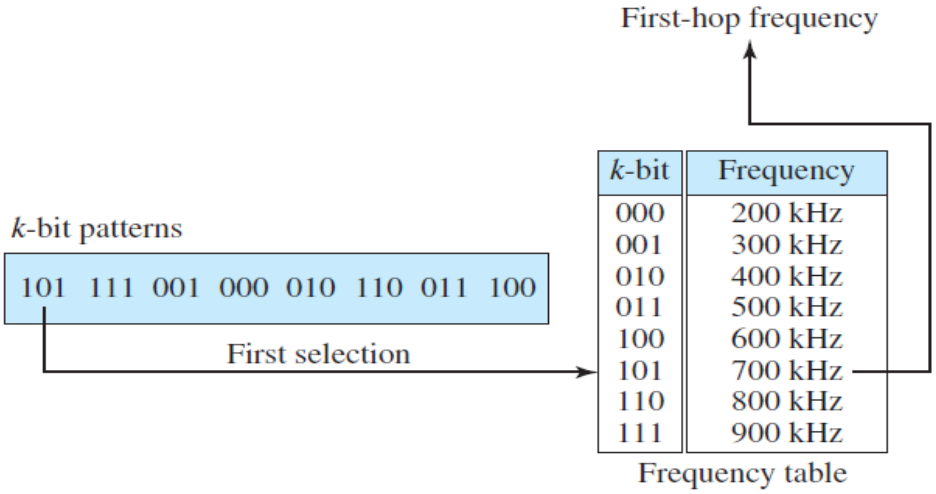


Fig.3.18 Frequency selection in FHSS

Figure 3.19 shows how the signal hops around from carrier to carrier. Assume the required bandwidth of the original signal is 100 kHz.

It can be shown that this scheme can accomplish the previously mentioned goals. If there are many *k*-bit patterns and the hopping period is short, a sender and receiver can have privacy. The scheme also has an antijamming effect. A malicious sender may be able to send noise to jam the signal for one hopping period (randomly), but not for the whole period.

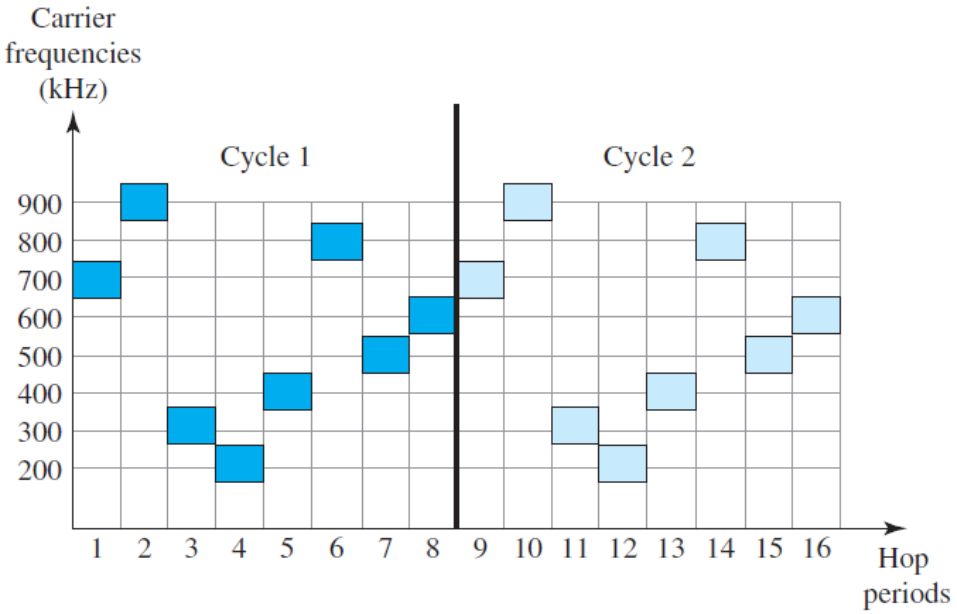


Fig.3.19 FHSS cycles

Direct Sequence Spread Spectrum (DSSS)

The **direct sequence spread spectrum (DSSS)** technique also expands the bandwidth of the original signal, replacing each data bit with *n* bits using a spreading code. Each bit is assigned a code of *n* bits, called ***chips,*** where the chip rate is *n* times that of the data bit. Figure 3.20 shows the concept of DSSS.

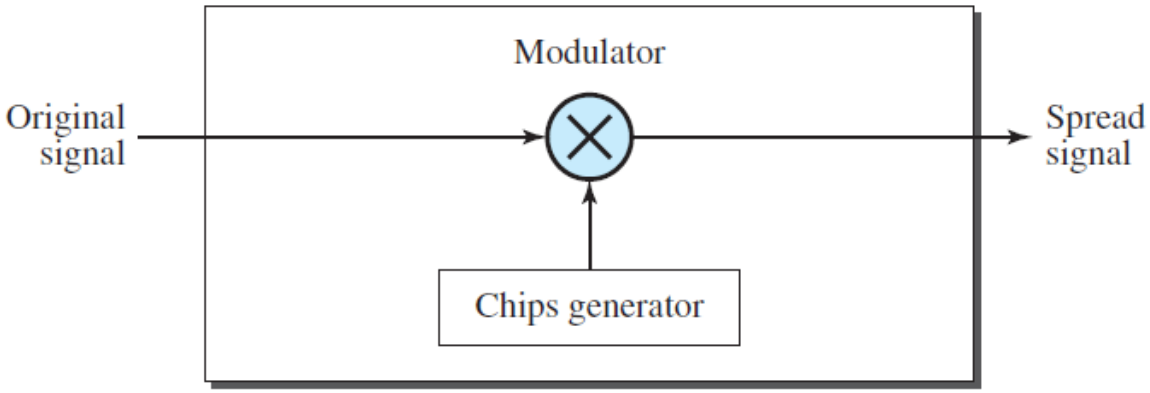


Fig.3.20 DSSS

For example, let us consider the sequence used in a wireless LAN, the famous **Barker sequence,** where *n* is 11. Assume that the original signal and the chips in the chip generator use polar NRZ encoding. Figure 3.21 shows the chips and the result of multiplying the original data by the chips to get the spread signal.

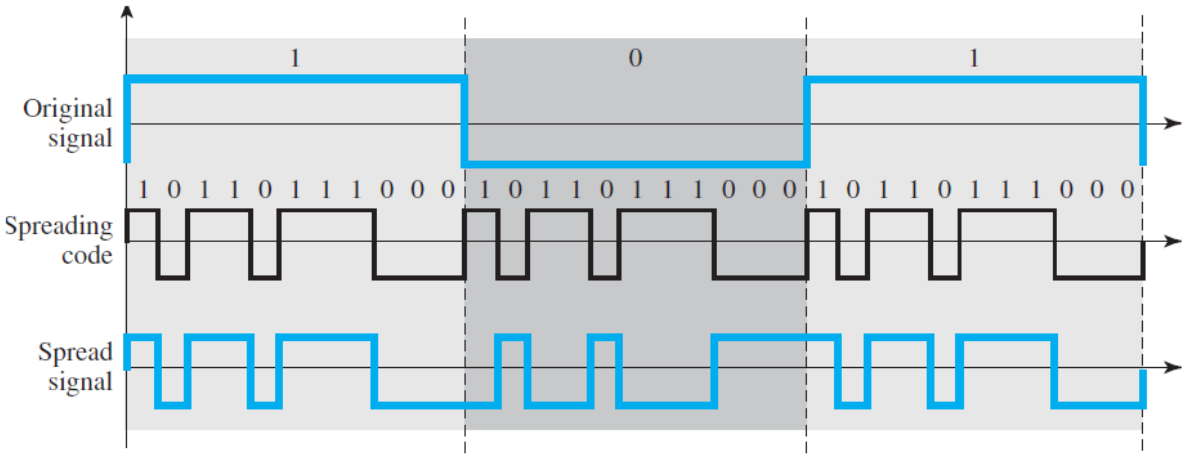


Fig.3.21 DSSS example

In Figure 3.21, the spreading code is 11 chips having the pattern 10110111000 (in this case). If the original signal rate is *N*, the rate of the spread signal is 11*N*. The spread signal can provide privacy if the intruder does not know the code. It can also provide immunity against interference if each station uses a different code.

SWITCHING

Switching is the solution to connect multiple devices in a network to make one-to-one communication. A switched network consists of a series of interlinked nodes, called ***switches.*** Switches are devices capable of creating temporary connections between two or more devices linked to the switch. Figure 3.22 shows a switched network.

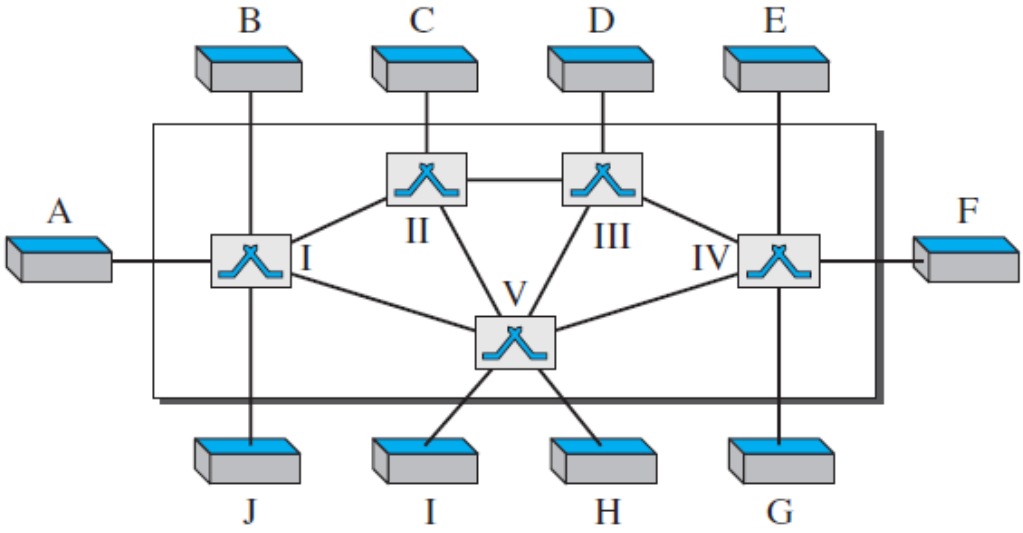
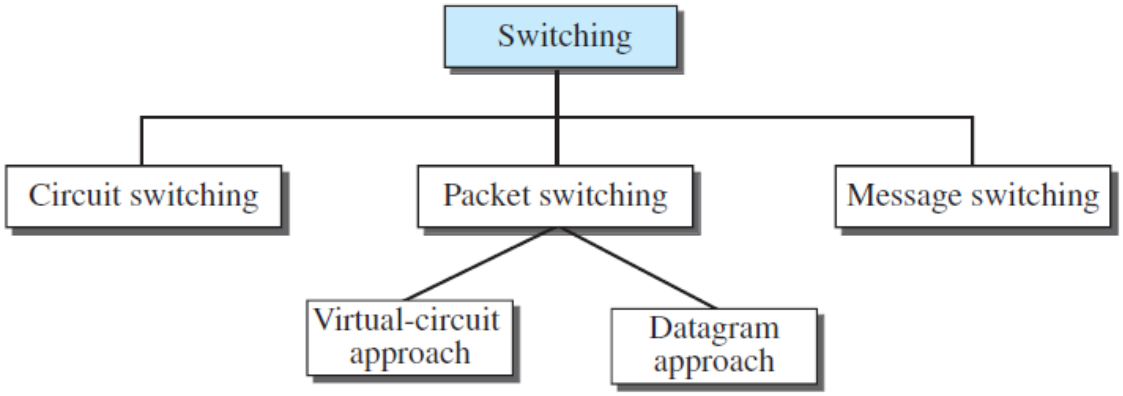


Fig.3.22 Switched network

The **end systems** (communicating devices) are labeled A, B, C, D, and so on, and the switches are labeled I, II, III, IV, and V. Each switch is connected to multiple links.

There are three methods of switching: circuit switching, packet switching and message switching. Packet switching can further be divided into two subcategories—virtual circuit approach and datagram approach.



Circuit-switched network:

A circuit-switched network is made of a set of switches connected by physical links, in which each link is divided into *n* channels. It occurs at physical layer. A connection between two stations is a dedicated path made of one or more links. In circuit switching, the resources need to be reserved during the setup phase; the resources remain dedicated for the entire duration of data transfer until the teardown phase.

The actual communication in a circuit-switched network requires three phases: connection setup, data transfer, and connection teardown.

Setup Phase

Before the two parties can communicate, a dedicated circuit needs to be established. The end systems are normally connected through dedicated lines to the switches, so connection setup means creating dedicated channels between the switches. For example, in Figure 3.23 when system A needs to connect to system M, it sends a setup request that includes the address of system M, to switch I. Switch I finds a channel between itself and switch IV that can be dedicated for this purpose. Switch I then send the request to switch IV, which finds a

dedicated channel between itself and switch III. Switch III informs system M of system A's intention currently.

In the next step to making a connection, system M will send an acknowledgment in the opposite direction to system A. Only after system A receives this acknowledgment is the connection established.

The end-to-end addressing is required for creating a connection between the two end systems. These can be, the addresses of the computers assigned by the administrator in a TDM network, or telephone numbers in an FDM network.

Data-Transfer Phase

After the establishment of the dedicated circuit (channels), the two parties can transfer data.

Teardown Phase

When one of the parties needs to disconnect, a signal is sent to each switch to release the resources.

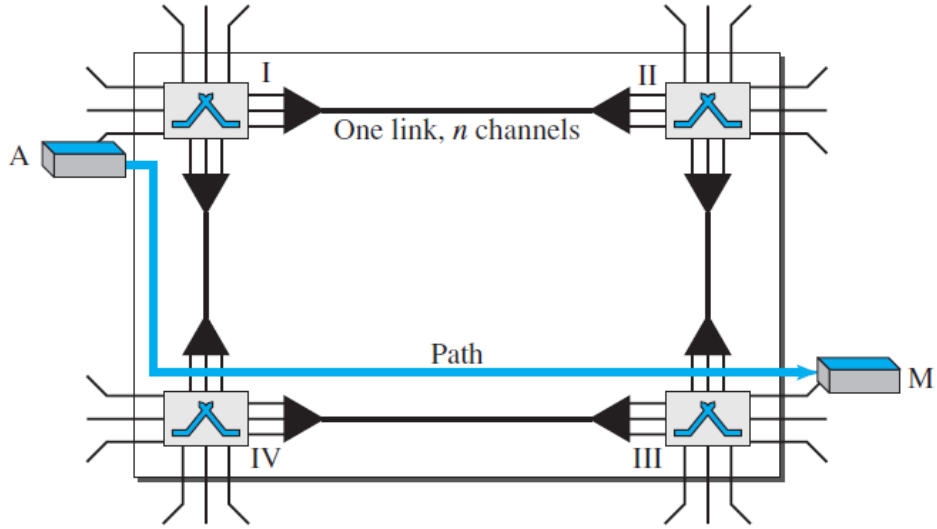


Fig.3.23 Circuit Switched network

Efficiency

The circuit-switched networks are not as efficient as the other two types of networks because resources are allocated during the entire duration of the connection. These resources are unavailable to other connections.

Delay

In circuit-switched network the delay is minimal. During data transfer the data are not delayed at each switch; the resources are allocated for the duration of the connection. Figure 3.24 shows the idea of delay in a circuit-switched network when only two switches are involved.

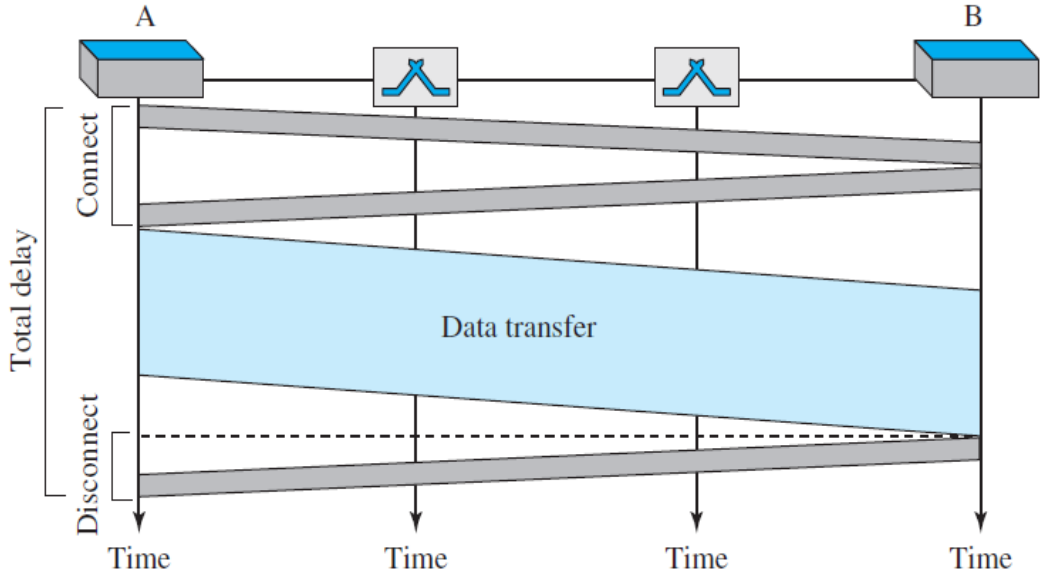


Fig.3.24 Delay in a Circuit Switched network

As figure 3.24 shows, there is no waiting time at each switch. **The total delay is due to the time needed to create the connection, transfer data, and disconnect the circuit.**

- The delay caused by the setup=the propagation time of the source computer request + the request signal transfer time +the propagation time of the acknowledgment from the destination computer+ the signal transfer time of the acknowledgment
- The delay due to data transfer =the propagation time +data transfer time
- The time needed to tear down the circuit.

Packet-switched network

In packet-switched network, messages are divided into packets of fixed or variable size. The size of the packet is determined by the network and the governing protocol.

In a packet-switched network, there is no resource reservation, resources are allocated on demand.

There are two types of packet-switched networks: **datagram networks and virtual circuit networks.**

Datagram Networks

In a **datagram network**, each packet is treated independently of all others. Packets in this approach are referred to as ***datagrams***.

Datagram switching is normally done at the network layer. Figure 3.25 shows how the datagram approach is used to deliver four packets from station A to station X. The switches in a datagram network are traditionally referred to as routers.

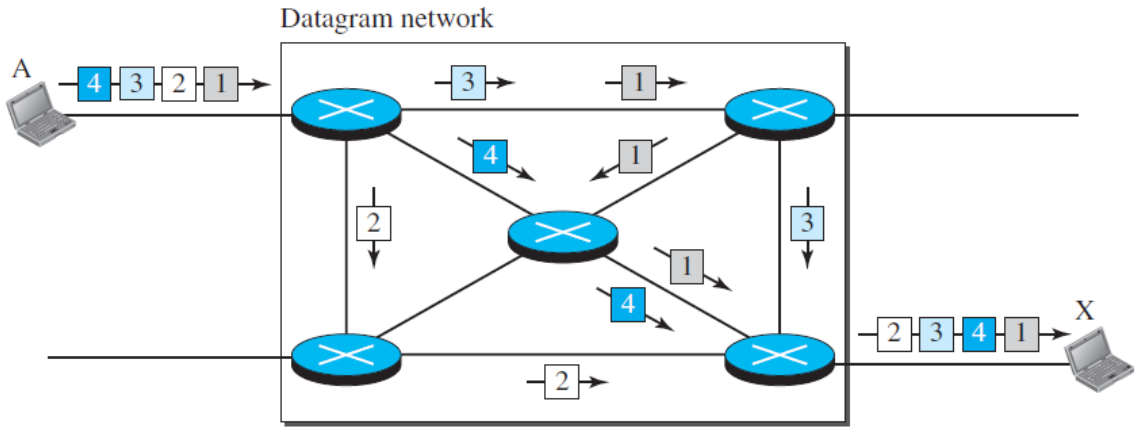


Fig.3.25 Datagram network with 4 switches

In this example, all four packets (or datagrams) belong to the same message but may travel different paths to reach their destination. This approach can cause the datagrams of a transmission to arrive at their destination out of order with different delays between the packets. Packets may also be lost or dropped because of a lack of resources.

The datagram networks are sometimes referred to as ***connectionless networks***. The term ***connectionless*** here means that the switch (packet switch) does not keep information about the connection state. There are no setup or teardown phases. Each packet is treated the same by a switch regardless of its source or destination.

Routing Table

In this type of network, each switch (or packet switch) has a routing table which is based on the destination address. The routing tables are dynamic and are updated periodically. The destination addresses and the corresponding forwarding output ports are recorded in the tables. Figure 3.26 shows the routing table for a switch.

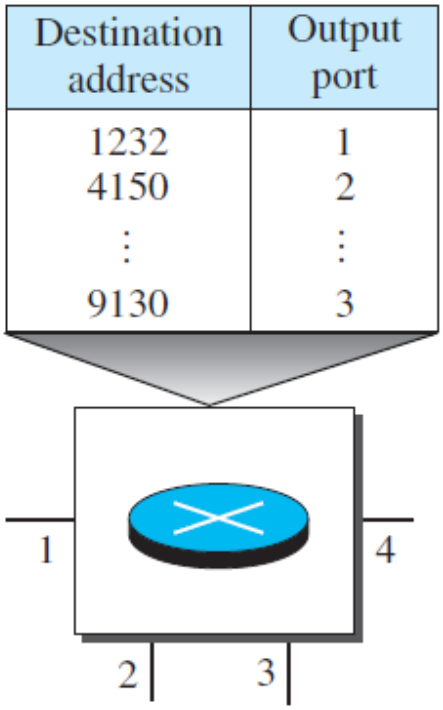


Fig.3.26 routing table in a datagram network

A switch in a datagram network uses a routing table that is based on the destination address.

Destination Address

Every packet in a datagram network carries a header that contains, among other information, the destination address of the packet. When the switch receives the packet, it examines the destination address and consult the routing table to find the corresponding port through which the packet should be forwarded. This address remains the same during the entire journey of the packet.

Efficiency

The efficiency of a datagram network is better than that of a circuit-switched network, because resources are allocated only when there are packets to be transferred.

Delay

Delay in a datagram network is greater than in a virtual-circuit network. Each packet may experience a wait at a switch before it is forwarded. Since not all packets in a message necessarily travel through the same switches, the delay is not uniform for the packets of a message. Figure 3.27 gives an example of delay in a datagram network for one packet.

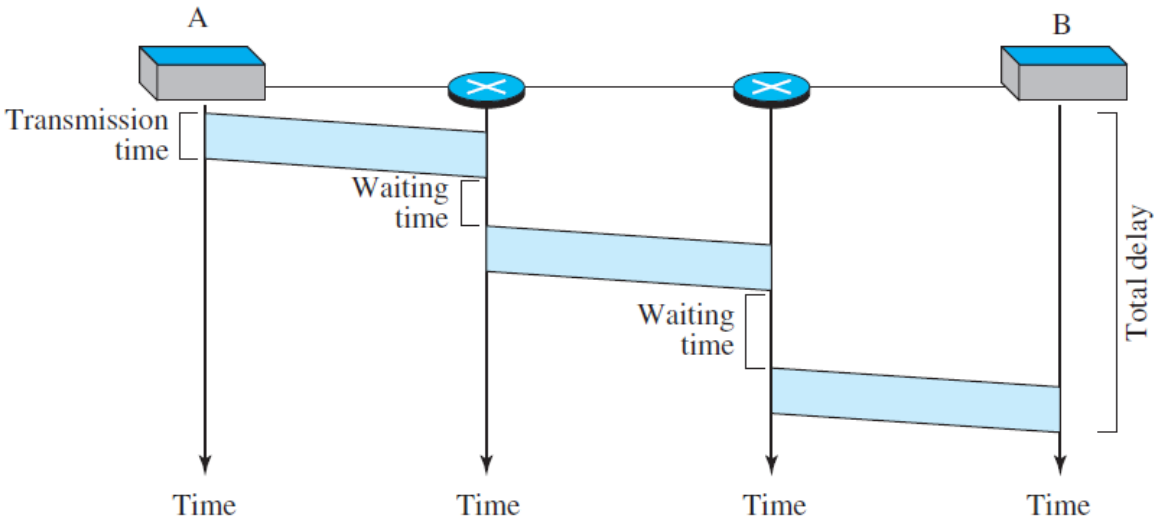


Fig.3.27 Delay in a datagram network

The packet travels through two switches. There are three transmission times (3*T*), three propagation delays (slopes 3τ of the lines), and two waiting times ($\mathbf{w_1}$+ $\mathbf{w_2}$). Processing time in each switch is ignored. The total delay is

$$\textbf{Total delay} = 3T + 3\tau + w_1 + w_2$$

Virtual-Circuit Networks

A **virtual-circuit network** is a cross between a circuit-switched network and a datagram network. It has some characteristics of both.

1. As in a circuit-switched network, there are setup and teardown phases in addition to the data transfer phase.

2. Resources can be allocated during the setup phase, as in a circuit-switched network, or on demand, as in a datagram network.

3. As in a datagram network, data are packetized, and each packet carries an address in the header. However, the address in the header has local jurisdiction, not end-to-end jurisdiction

4. As in a circuit-switched network, all packets follow the same path established during the connection.

5.A virtual-circuit network is normally implemented in the data-link layer, while a circuit-switched network is implemented in the physical layer and a datagram network in the network layer.

Figure 3.28 is an example of a virtual-circuit network. The network has switches that allow traffic from sources to destinations. A source or destination can be a computer, packet switch, bridge, or any other device that connects other networks.

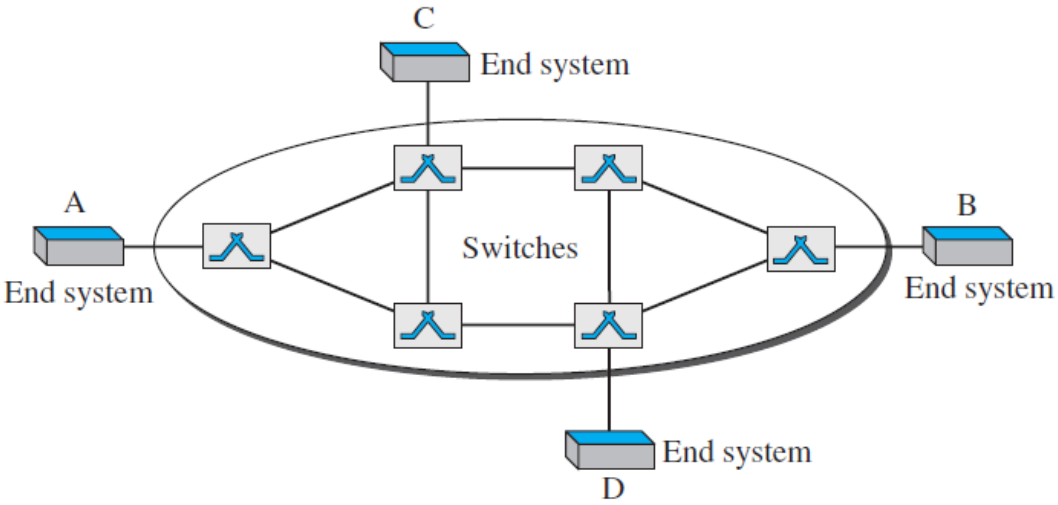


Fig.3.28 Virtual circuit network

Addressing

In a virtual-circuit network, two types of addressing are involved: global and local (virtual-circuit identifier).

Global Addressing

A source or a destination needs to have a global address—an address that can be unique in the scope of the network or internationally if the network is part of an international network.

Virtual-Circuit Identifier

The identifier that is used for data transfer is called the ***virtual-circuit identifier*** **(VCI)** or the ***label.*** A VCI is a small number that has only switch scope; it is used by a frame between two switches. When a frame arrives at a switch, it has a VCI; when it leaves, it has a different VCI. Figure 3.29 shows how the VCI in a data frame changes from one switch to another.

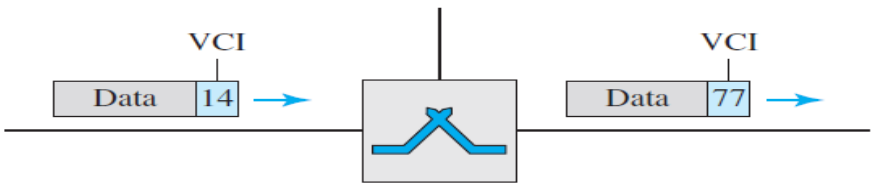


Fig.3.29 Virtual circuit identifier

Three Phases

A source and destination need to go through three phases in a virtual-circuit network: **setup, data transfer, and teardown.**

Setup Phase

In the setup phase, a switch creates an entry for a virtual circuit. The source and destination use their global addresses to help switches make table entries for the connection. For example, suppose source A needs to create a virtual circuit to B. Two steps are required: **the setup request and the acknowledgment**.

Setup Request

A setup request frame is sent from the source to the destination. Figure 3.30 shows the process.

a. Source A sends a setup frame to switch 1.

b. Switch 1 receives the setup request frame. It knows that a frame going from A to B goes out through port 3. The switch can only able to fill three of the four columns. The switch assigns the incoming port (1) and chooses an available incoming VCI (14) and the outgoing port (3). The fourth column, outgoing VCI, which will be found during the acknowledgment step. The switch then forwards the frame through port 3 to switch 2.

c. Switch 2 receives the setup request frame. The same events happen here as at switch 1, three columns of the table are completed: in this case, incoming port (1), incoming VCI (66),

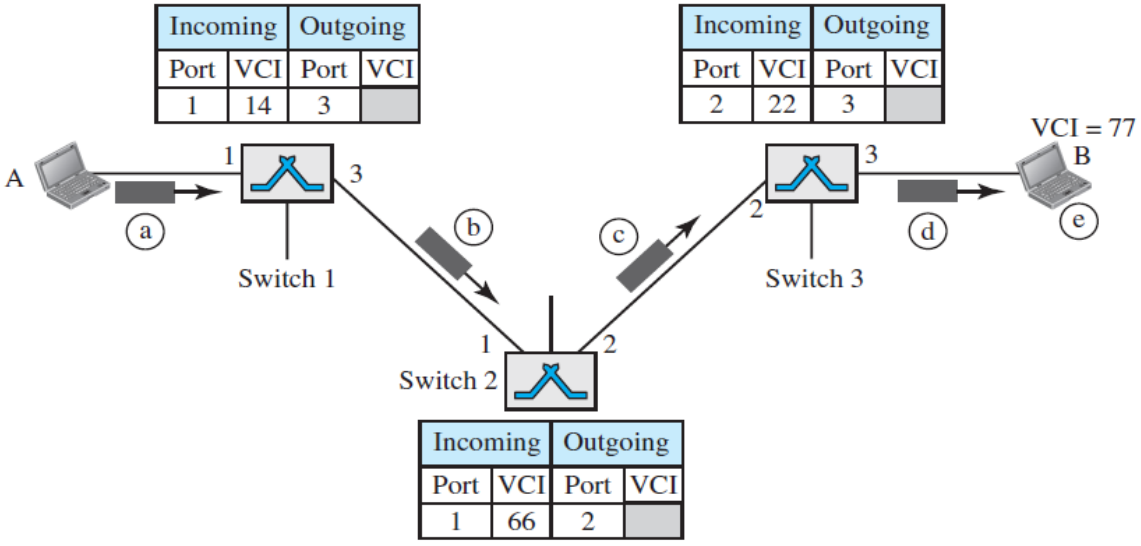


Fig.3.30 set-up request in a Virtual circuit network

and outgoing port (2).

d. Switch 3 receives the setup request frame. Again, three columns are completed: incoming port (2), incoming VCI (22), and outgoing port (3).

e. Destination B receives the setup frame, and if it is ready to receive frames from A, it assigns a VCI to the incoming frames that come from A, in this case 77. This VCI lets the destination know that the frames come from A, and no other sources.

Acknowledgment

A special frame, called the *acknowledgment frame*, completes the entries in the switching tables. Figure 3.31 shows the process.

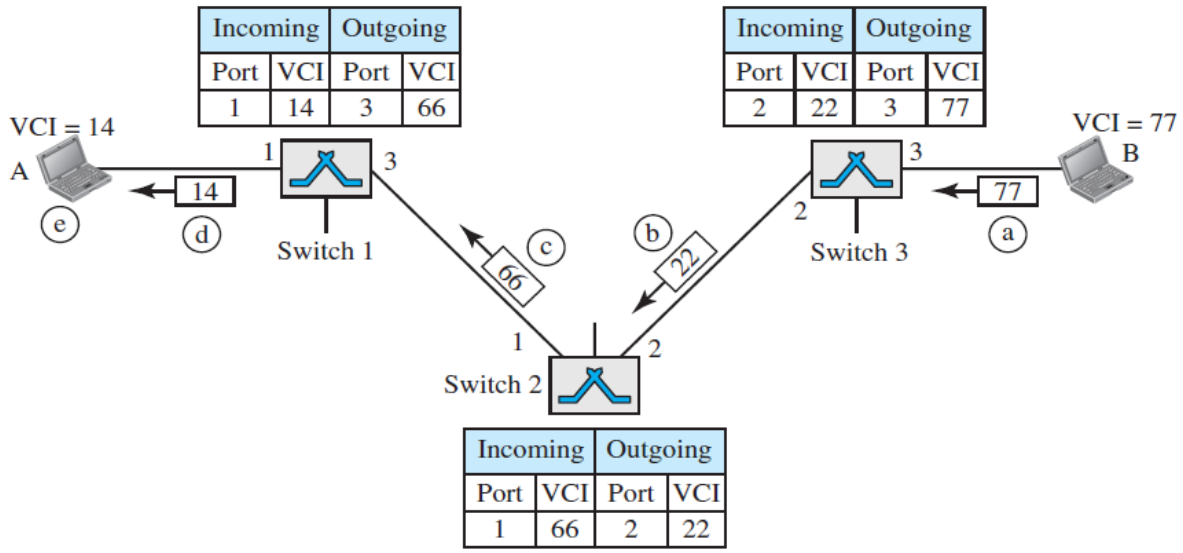


Fig.3.31 set-up acknowledgement in a Virtual circuit network

a. The destination sends an acknowledgment to switch 3. The acknowledgment carries the global source and destination addresses. The frame also carries VCI 77, chosen by the destination as the incoming VCI for frames from A. Switch 3 uses this VCI to complete the outgoing VCI column for this entry.

b. Switch 3 sends an acknowledgment to switch 2 that contains its incoming VCI in the table, chosen in the previous step. Switch 2 uses this as the outgoing VCI in the table.

c. Switch 2 sends an acknowledgment to switch 1 that contains its incoming VCI in the table, chosen in the previous step. Switch 1 uses this as the outgoing VCI in the table.

d. Finally switch 1 sends an acknowledgment to source A that contains its incoming VCI in the table, chosen in the previous step.

e. The source uses this as the outgoing VCI for the data frames to be sent to destination to B.

Teardown Phase

In this phase, source A, after sending all frames to B, sends a special frame called a *teardown request*. Destination B responds with a teardown confirmation frame. All switches delete the corresponding entry from their tables.

Efficiency

In virtual circuit switching, all packets belonging to the same source and destination travel the same path, but the packets may arrive at the destination with different delays if resource allocation is on demand. There is one big advantage in a virtual-circuit network even if resource allocation is on demand. The source can check the availability of the resources, without reserving it.

Delay in Virtual-Circuit Networks

In a virtual-circuit network, there is a one-time delay for setup and a one-time delay for teardown. If resources are allocated during the setup phase, there is no wait time for individual packets. Figure 3.32 shows the delay for a packet traveling through two switches in a virtual-circuit network.

The packet is traveling through two switches (routers). The total delay time is

Total delay = 3T + 3τ + set up delay + tear down delay

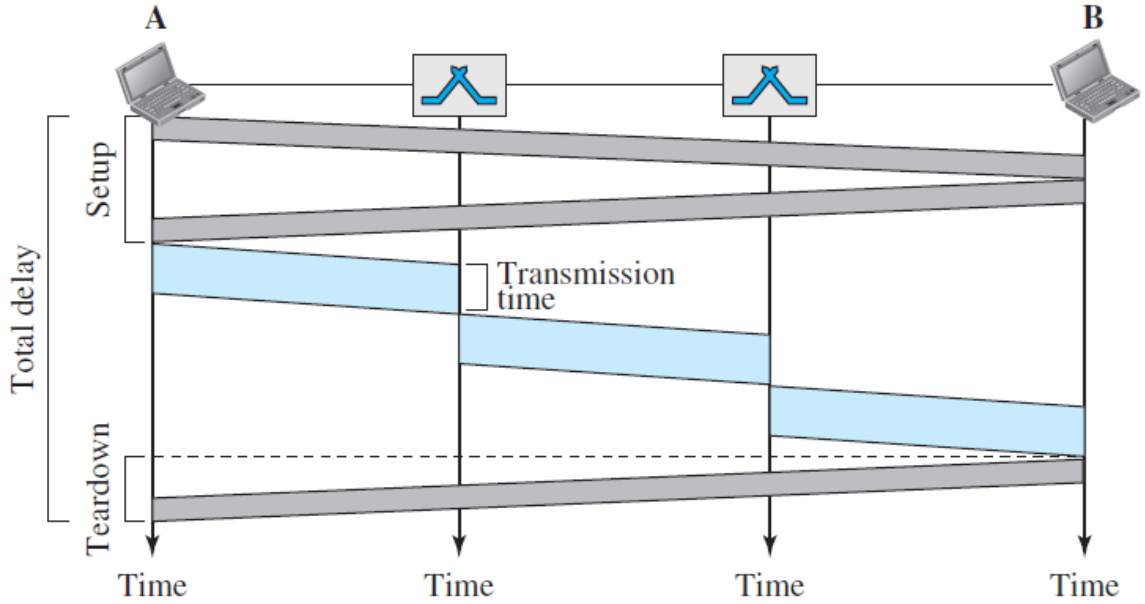


Fig.3.32 delay in a Virtual circuit network

ERROR DETECTION AND CORRECTION

Networks must be able to transfer data from one device to another with acceptable accuracy. For most applications, a system must guarantee that the data received are identical to the data transmitted. Any time data are transmitted from one node to the next, they can become corrupted in passage. Many factors can alter one or more bits of a message. Some applications require a mechanism for detecting and correcting **errors.**

Types of Errors

Whenever bits flow from one point to another, they are subject to unpredictable changes because of **interference.** This interference can change the shape of the signal.

There are 2 types of errors: ***single-bit error and burst error.***

Single-bit error -> only 1 bit of a given data unit (such as a byte, character, or packet) is changed from 1 to 0 or from 0 to 1.

Burst error -> 2 or more bits in the data unit have changed from 1 to 0 or from 0 to 1. Burst error is more likely to occur than a single bit error because the duration of the noise signal is normally longer than the duration of 1 bit

Below figure 3.33 shows the effect of a single-bit and a burst error on a data unit.

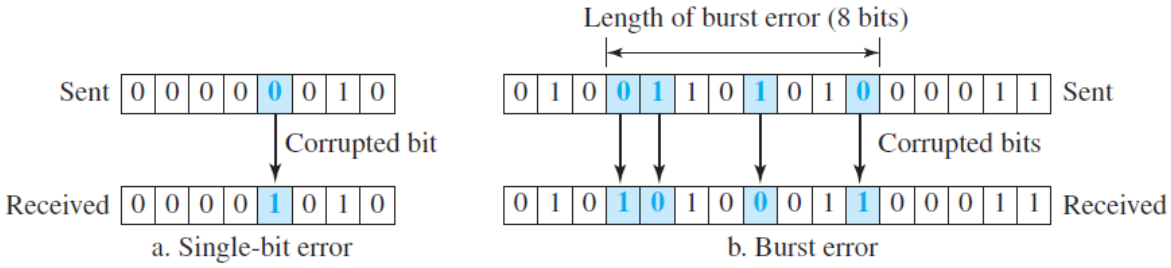


Fig.3.33 single-bit and burst error

The central concept in detecting or correcting errors is **redundancy** which means adding some extra bits to data. These redundant bits are added by the sender and removed by the receiver. The correction of errors is more difficult than the detection. In **error detection**, it only looks to see if any error has occurred. In **error correction**, it is necessary to know the exact number of bits that are corrupted and their location in the message.

Redundancy is achieved through various coding schemes. The sender adds redundant bits through a process that creates a relationship between the redundant bits and the actual data bits. The receiver checks the relationships between the two sets of bits to detect errors.

BLOCK CODING

In block coding, A message is divided into blocks, each of '*k*' bits, called ***data words.*** '*r*' redundant bits are added to each block to make the length $n = k + r$. The resulting *n*-bit blocks are called ***code words.*** With *k* bits, a combination of 2^k data words can be created and with *n* bits, a combination of 2^n codewords can be created. Since $n > k$, the number of possible code words is larger than the number of possible data words. The block coding process is one-to-one; the same data word is always encoded as the same codeword. Hence, out of 2^n codewords,$2^n - 2^k$ codewords are not used. These codewords are invalid or illegal. If the receiver receives an invalid codeword, this indicates that the data was corrupted during transmission.

Error Detection

The receiver can detect a change in the original codeword, if the following two conditions are met:

1. The receiver has a list of valid codewords.

2. The original codeword has changed to an invalid one.

Figure 3.34 shows the role of block coding in error detection. The sender creates codewords out of data words by using a generator that applies the rules and procedures of encoding Each codeword sent to the receiver may change during transmission. If the received codeword is the same as one of the valid codewords, the word is accepted; the corresponding data word is extracted for use. If the received codeword is not valid, it is discarded. However, if the codeword is corrupted during transmission but the received word still matches a valid codeword, the error remains undetected.

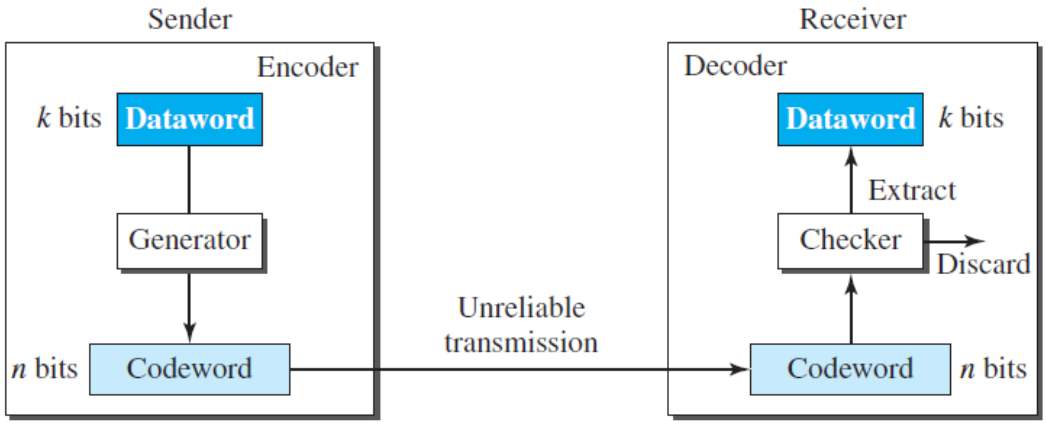


Fig.3.34 Error detection in Block coding

Example: Let us assume that $k = 2$ and $n = 3$. Below Table 3.1 shows the list of data words and codewords.

Dataword	*Codeword*	*Dataword*	*Codeword*
00	000	10	101
01	011	11	110

Table 3.1 A code for error detection

Assume the sender encodes the data word 01 as 011 and sends it to the receiver. Consider the following cases:

1. The receiver receives 011. It is a valid codeword. The receiver extracts the data word 01 from it.
2. The codeword is corrupted during transmission, and 111 is received (the leftmost bit is corrupted). This is not a valid codeword and is discarded.
3. The codeword is corrupted during transmission, and 000 is received (the right two bits are corrupted). This is a valid codeword. The receiver incorrectly extracts the data word 00. Two corrupted bits have made the error undetectable.

Hence, an error-detecting code can detect only the types of errors for which it is designed, other types of errors may remain undetected.

Hamming distance

The **Hamming distance** between two words (of the same size) is the number of differences between the corresponding bits. The Hamming distance between two words x and y as $\boldsymbol{d}(x, y)$. Hamming distance between the received codeword and the sent codeword is the number of bits that are corrupted during transmission.

For example, if the codeword 00000 is sent and 01101 is received, 3 bits are in error and the Hamming distance between the two is $\boldsymbol{d}(00000, 01101) = 3$.

The Hamming distance can easily be found by applying the XOR operation on the two words and count the number of 1s in the result.

The **minimum Hamming distance** is the smallest Hamming distance between all possible pairs of codewords. To guarantee the detection of up to s errors in all cases, the minimum Hamming distance in a block code must be $\boldsymbol{d_{min} = s + 1}$.

Linear Block Codes

a linear block code is a code in which the exclusive OR of two valid codewords creates another valid codeword.

The code in Table 3.1 is a linear block code because the result of XORing any codeword with any other codeword is a valid codeword. For example, the XORing of the second and third codewords creates the fourth one.

Minimum Distance for Linear Block Cods is the number of 1s in the nonzero valid codeword with the smallest number of 1s.

Example: In Table 3.1, the numbers of 1s in the nonzero codewords are 2, 2, and 2. So the minimum Hamming distance is $d_{min} = 2$.

Parity-Check Code

This code is a linear block code. In this code, a *k*-bit data word is changed to an *n*-bit code word, where $n = k + 1$. The extra bit, called the *parity bit*, is selected to make the total number of 1s in the codeword even.

The code in Table 3.2 is a parity-check code with $k = 4$ and $n = 5$.

Dataword	*Codeword*	*Dataword*	*Codeword*
0000	**00000**	1000	**10001**
0001	**00011**	1001	**10010**
0010	**00101**	1010	**10100**
0011	**00110**	1011	**10111**
0100	**01001**	1100	**11000**
0101	**01010**	1101	**11011**
0110	**01100**	1110	**11101**
0111	**01111**	1111	**11110**

Table 3.2 Parity check code

Figure 3.35 shows a possible structure of an encoder (at the sender) and a decoder of simple parity check code. In the above fig., the encoder uses a generator that takes a copy of a 4-bit data word (a_0, a_1, a_2, *and* a_3) and generates a parity bit r_0. The data word bits and the parity bit create the 5-bit codeword. The parity bit that is added makes the number of 1s in the codeword even. This is normally done by adding the 4 bits of the data word (modulo-2); the result is the parity bit.

$$r_0 = a_3 + a_2 + a_1 + a_0 \textbf{ (modulo-2)}$$

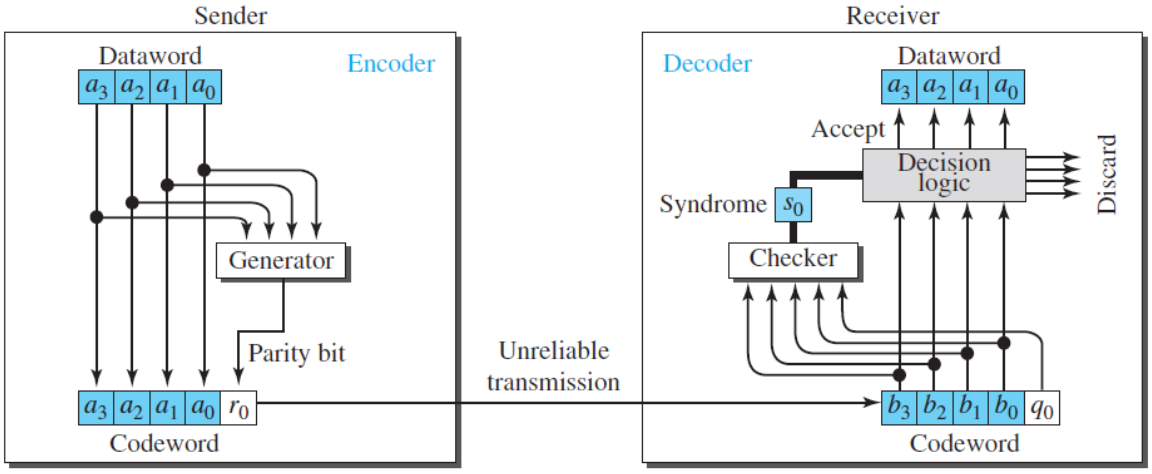


Fig. 3.35 encoder and decoder of simple Parity check code

If the number of 1s is even, the result is 0; if the number of 1s is odd, the result is 1. In both cases, the total number of 1s in the codeword is even. The sender sends the codeword, which may be corrupted during transmission. The receiver receives a 5-bit word. The checker at the receiver do the addition over all 5 bits. The result, which is called the ***syndrome,*** is just 1 bit. The syndrome is 0 when the number of 1s in the received codeword is even; otherwise, it is 1.

$$s_0 = b_3 + b_2 + b_1 + b_0 + q_0 \text{ (modulo-2)}$$

The syndrome is passed to the decision logic analyzer. If the syndrome is 0, there is no detectable error in the received codeword, the data portion of the received codeword is accepted as the data word. If the syndrome is 1, the data portion of the received codeword is discarded. The data word is not created.

Note: A parity-check code can detect an odd number of errors.

CYCLIC CODES

Cyclic codes are special linear block codes with one extra property. In a **cyclic code,** if a codeword is cyclically shifted (rotated), the result is another codeword.

For example, if 1011000 is a codeword and by doing cyclically left shift, then 0110001 is also a codeword.

Cyclic Redundancy Check

A subset of cyclic codes called the **cyclic redundancy check (CRC)**, which is used in networks such as LANs and WANs.

Figure 3.36 shows design for the encoder and decoder of CRC.

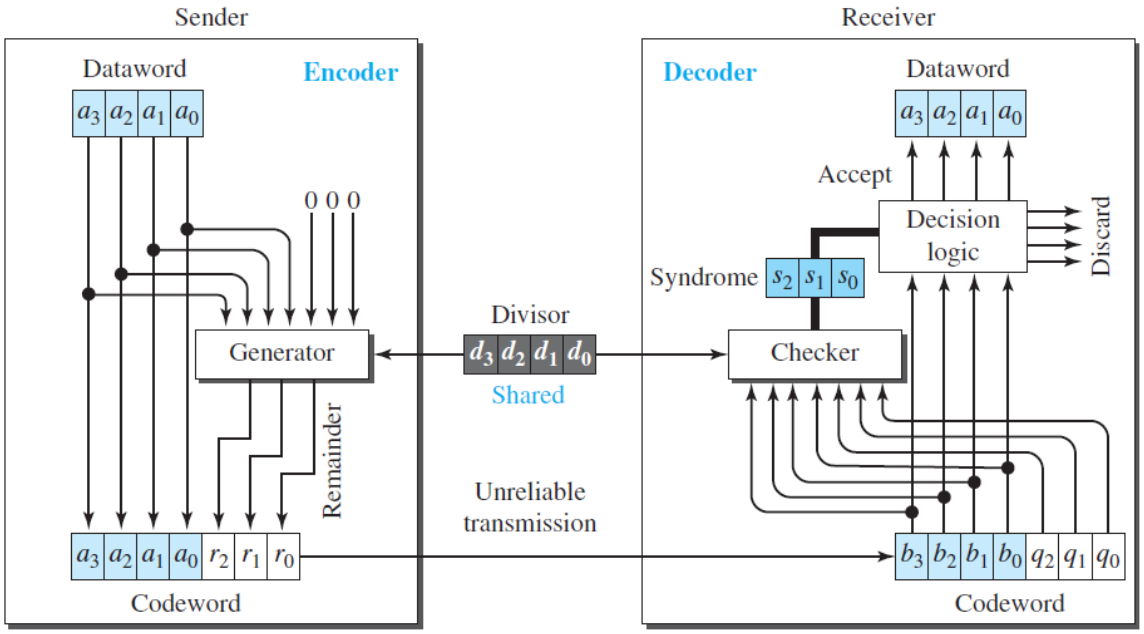


Fig. 3.36 encoder and decoder of CRC

In the encoder, the data word has *k* bits (k=4 here),the codeword has *n* bits (n=7 here).The size of the data word is augmented by adding $n - k$ (3 here) 0s to the right-hand side of the word. The *n*-bit result is fed into the generator. The generator uses a divisor of size $n - k + 1$ (4 here), which is predefined. The generator divides the augmented data word by the divisor (modulo-2 division). The quotient of the division is discarded and the remainder ($r_2r_1r_0$) is appended to the data word to create the codeword.

The decoder receives the codeword A copy of all *n* bits is fed to the checker, which is a replica of the generator. The remainder produced by the checker is a syndrome of $n - k$ (3 here) bits, which is fed to the decision logic analyzer. The analyzer has a simple function. If the syndrome bits are all 0s, the 4 leftmost bits of the codeword are accepted as the data word. Otherwise, the 4 bits are discarded (error).

Note: The divisor in a cyclic code is normally called the *generator polynomial* or simply the *generator*.

For example: The encoder takes a data word and augments it with $n + k$ number of 0s. It then divides the augmented data word by the divisor, as shown in figure 3.37. In each step, a copy of the divisor is XORed with the 4 bits of the dividend. The result of the XOR operation (remainder) is 3 bits (in this case), which is used for the next step after 1 extra bit is pulled down to make it 4 bits long. When there are no bits left to pull down, we have a result. The 3-

bit remainder forms the **check bits (r_2, r_1, and r_0)**. They are appended to the data word to create the codeword.

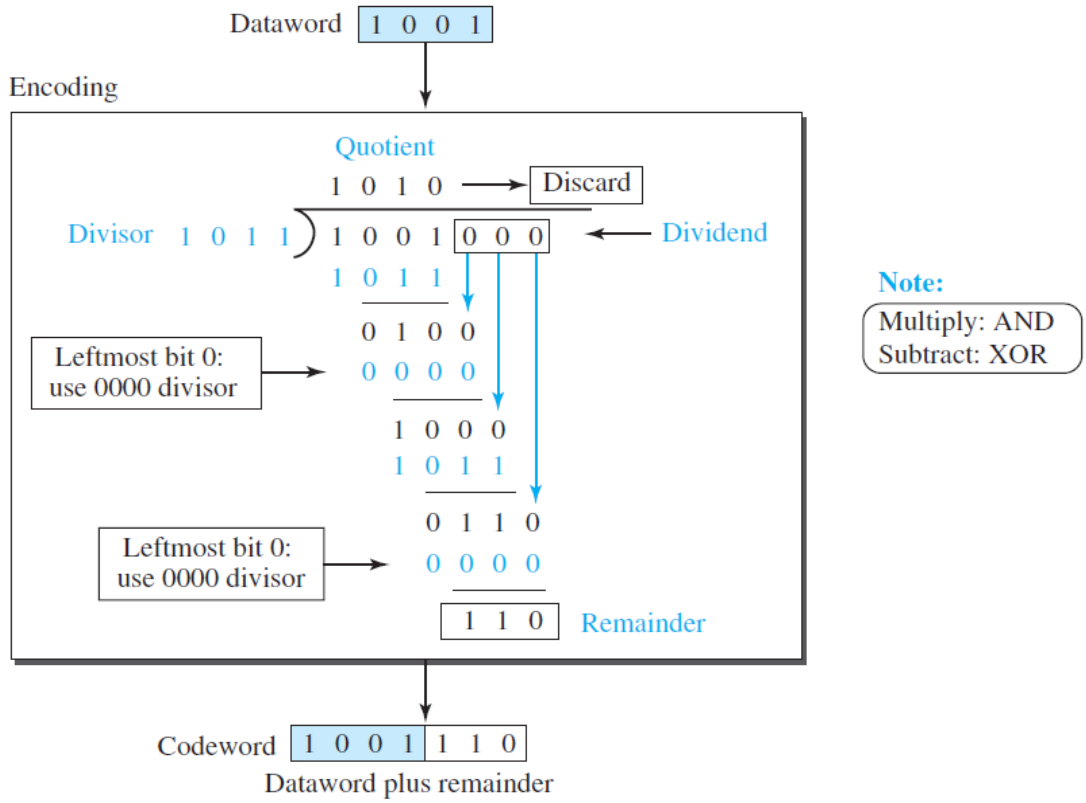


Fig. 3.37 Division in CRC encoder

The decoder does the same division process as the encoder as shown in figure 3.38. The remainder of the division is the syndrome. If the syndrome is all 0s, then there is no error and the data word is separated from the received codeword and accepted. Otherwise, everything is discarded.

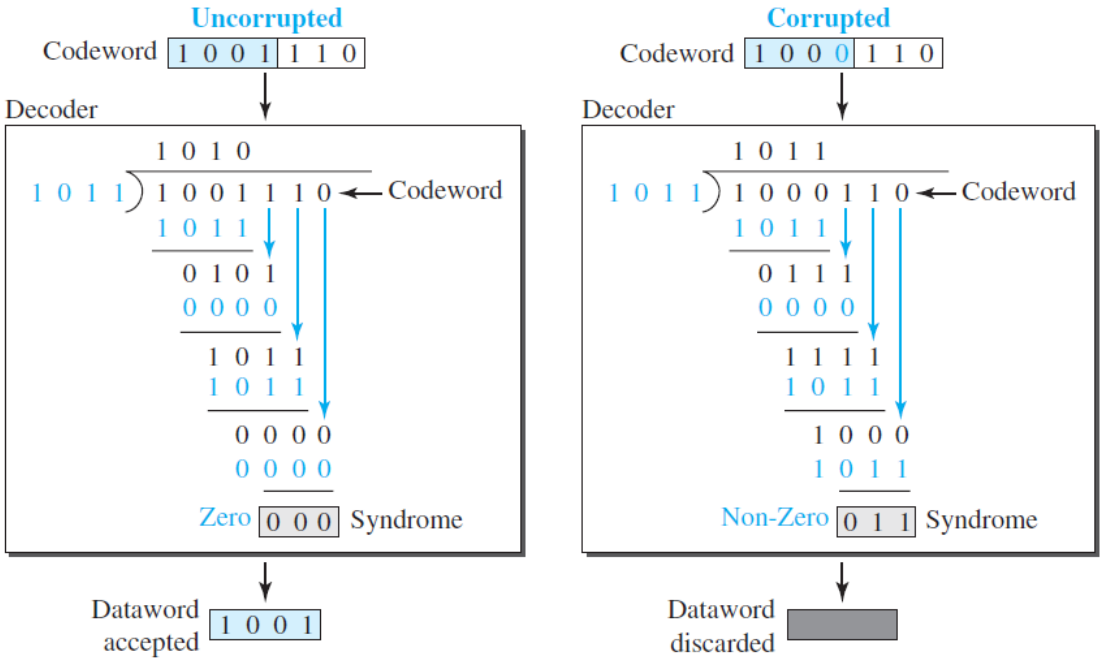


Fig. 3.38 Division in the CRC decoder for two cases

CHECKSUM

Checksum is an error-detecting technique that can be applied to a message of any length. In the Internet, the checksum technique is mostly used at the network and transport layer rather than the data-link layer.

At the source, the message is first divided into *m*-bit units. The generator then creates an extra *m*-bit unit called the ***checksum,*** which is sent with the message. At the destination, the checker creates a new checksum from the combination of the message and sent checksum. If the new checksum is all 0s, the message is accepted; otherwise, the message is discarded as shown in figure 3.39.

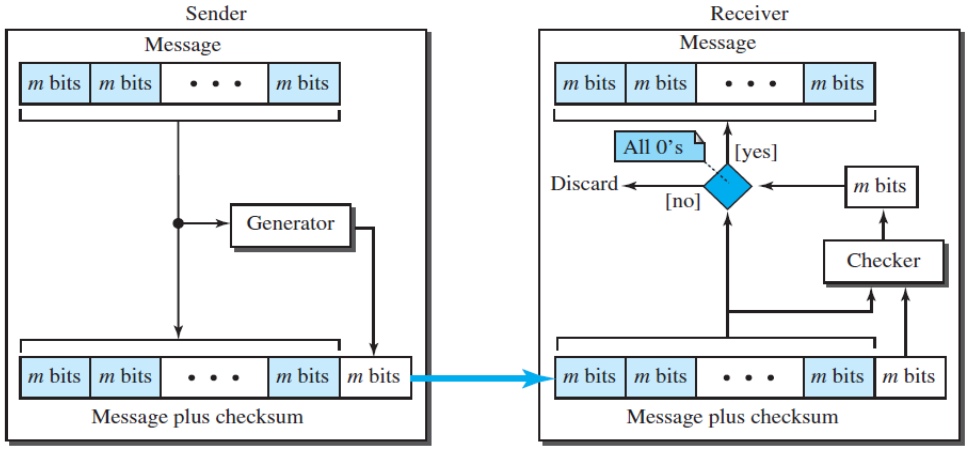


Fig. 3.39 checksum

Example:

Suppose the message is a list of five 4-bit numbers. The set of numbers is (7, 11, 12, 0, 6). The sender adds all five numbers in **one's complement** to get the sum = 6. The sender then complements the result to get the checksum = 9, which is 15 – 6. Note that 6 = (0110)2 and 9 = (1001)2; they are complements of each other. The sender sends the five data numbers and the checksum (7, 11, 12, 0, 6, **9**). If there is no corruption in transmission, the receiver receives (7, 11, 12, 0, 6, **9**) and adds them in one's complement to get 15. The sender complements 15 to get 0. This shows that data have not been corrupted. Figure 3.40 shows the process.

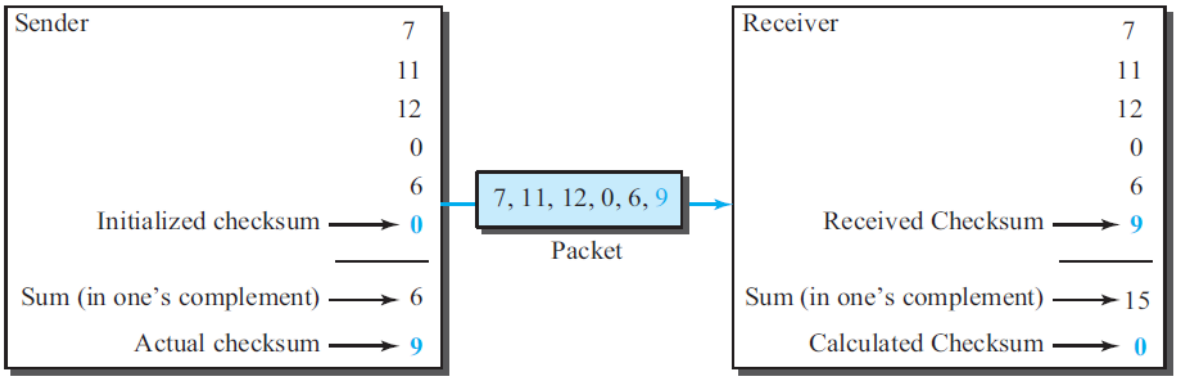


Fig. 3.40 checksum example

one's complement arithmetic

In this arithmetic, unsigned numbers can be represented between 0 and $2^m - 1$ using only *m* bits. If the number has more than *m* bits, the extra leftmost bits need to be added to the *m* rightmost bits (wrapping).

Example: the decimal number 36 in binary is $(100100)_2$. To change it to a 4-bit number, add the extra leftmost bit to the right four bits as shown below.

$$\mathbf{(10)_2 + (0100)_2 = (0110)_2 \rightarrow (6)_{10}}$$

Instead of sending 36 as the sum, we can send 6 as the sum (7, 11, 12, 0, 6, **6**). The receiver can add the first five numbers in one's complement arithmetic. If the result is 6, the numbers are accepted. otherwise, they are rejected.

Module 4

Data link control: DLC services, Data link layer protocols, Point to Point protocol (Framing, Transition phases only). Media Access control: Random Access, Controlled Access and Channelization, Introduction to Data-Link Layer: Introduction, Link-Layer Addressing, ARP IPv4 Addressing and subnetting: Classful and CIDR addressing, DHCP, NAT.

The data link control (DLC) deals with procedures for communication between two adjacent nodes i.e. node-to-node communication. Data link control functions include 1) Framing and 2) Flow control and 3) Error control.

4.1 Framing

A frame is a group of bits. Framing means organizing the bits into a frame that are carried by the physical layer. The data-link-layer needs to form frames, so that each frame is distinguishable from another. Framing separates a message from other messages by adding sender-address & destination-address. The destination-address defines where the packet is to go. The sender-address helps the recipient acknowledge the receipt. Whole message is not packed in one frame because large frame makes flow and error-control very inefficient. Even a single-bit error requires the re-transmission of the whole message. When a message is divided into smaller frames, a single-bit error affects only that small frame. Our postal system practices a type of framing. The simple act of inserting a letter into an envelope separates one piece of information from another; the envelope serves as the delimiter. In addition, each envelope defines the sender and receiver addresses since the postal system is a many-to-many carrier facility.

Frame Size

Two types of frames:

1) Fixed Size Framing

 There is no need for defining boundaries of frames; the size itself can be used as a delimiter. For example: ATM WAN uses frames of fixed size called cells.

2) Variable Size Framing

 We need to define the end of the frame and the beginning of the next frame.

 Two approaches are used: a) Character-oriented approach

b) Bit-oriented approach.

a) Character Oriented Framing

- Data to be carried are 8-bit characters from a coding system such as ASCII.
- The header and the trailer are also multiples of 8 bits.
    1) Header carries the source and destination-addresses and other control information.
    2) Trailer carries error-detection or error-correction redundant bits.
- To separate one frame from the next frame, an 8- bit (I-byte) flag is added at the beginning and the end of a frame.
- The flag is composed of protocol-dependent special characters.
- The flag signals the start or end of a frame.

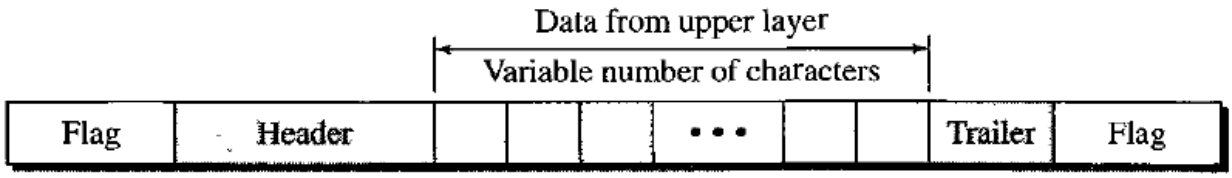


Figure 4.1 : A frame in a character oriented protocol

Problem:

 Character-oriented framing is suitable when only text is exchanged by the data-link-layers.

 However, if we send other type of information (say audio/video), then any pattern used for the flag can also be part of the information.

 If the flag-pattern appears in the data-section, the receiver might think that it has reached the end of the frame.

Solution: A byte-stuffing is used.

(Byte stuffing character stuffing)

- In byte stuffing, a special byte is added to the data-section of the frame when there is a character with the same pattern as the flag.
- The data-section is stuffed with an extra byte. This byte is called the escape character (ESC), which has a predefined bit pattern.
- When a receiver encounters the ESC character, the receiver
    → removes ESC character from the data-section and
    → treats the next character as data, not a delimiting flag.
- Problem:

 What happens if the text contains one or more escape characters followed by a flag?

 The receiver removes the escape character, but keeps the flag, which is incorrectly

interpreted as the end of the frame.

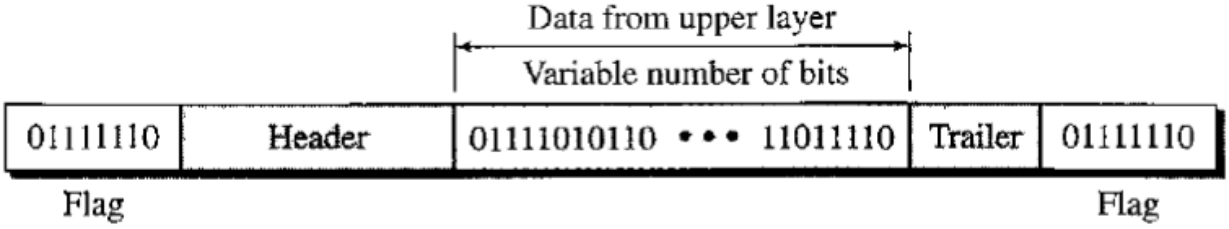
Solution:

 Escape characters part of the text must also be marked by another escape character.

In short, byte stuffing is the process of adding one extra byte whenever there is a flag or escape character in the text.

b) Bit Oriented Framing

The data-section of a frame is a sequence of bits to be interpreted by the upper layer as text, audio, video, and so on. However, in addition to headers and trailers, we need a delimiter to separate one frame from the other. Most protocols use a special 8-bit pattern flag 01111110 as the delimiter to define the beginning and the end of the frame (Figure 4.2).



4.2 : A Frame in a bit oriented protocol

Problem:

If the flag-pattern appears in the data-section, the receiver might think that it has reached the end of the frame.

Solution: A bit-stuffing is used.

In bit stuffing, if a 0 and five consecutive 1 bits are encountered, an extra 0 is added. This extra stuffed bit is eventually removed from the data by the receiver. This guarantees that the flag field sequence does not inadvertently appear in the frame. In short, bit stuffing is the process of adding one extra 0 whenever five consecutive 1s follow a 0 in the data, so that the receiver does not mistake the pattern 0111110 for a flag.

Flow Control and Error Control

One of the responsibilities of the DLC sublayer is flow and error control at the data-link layer.

Flow Control

Whenever an entity produces items and another entity consumes them, there should be a balance between production and consumption rates. If the items are produced faster than they can be consumed, the consumer can be overwhelmed and may need to discard some items. We need to prevent losing the data items at the consumer site. At the sending node, the data-link layer tries to push frames toward the data-link layer at the receiving node.

- If the receiving node cannot process and deliver the packet to its network at the same rate that the frames arrive, it becomes overwhelmed with frames.
- Here, flow control can be feedback from the receiving node to the sending node to stop or slow down pushing frames.

Buffers

Flow control can be implemented by using buffer. A buffer is a set of memory locations that can hold packets at the sender and receiver. Normally, two buffers can be used.

1) First buffer at the sender.

2) Second buffer at the receiver.

- The flow control communication can occur by sending signals from the consumer to the producer.
- When the buffer of the receiver is full, it informs the sender to stop pushing frames.

Error Control

- Error-control includes both error-detection and error-correction.
- Error-control allows the receiver to inform the sender of any frames lost/damaged in transmission.
- A CRC is
  - → added to the frame header by the sender and
  - → checked by the receiver.
- At the data-link layer, error control is normally implemented using one of the following two methods.

1) First method: If the frame is corrupted, it is discarded; If the frame is not corrupted, the packet is delivered to the network layer. This method is used mostly in wired LANs such as Ethernet.

2) Second method: If the frame is corrupted, it is discarded; If the frame is not corrupted, an acknowledgment is sent to the sender. Acknowledgment is used for the purpose of both flow and error control.

Combination of Flow and Error Control

- Flow and error control can be combined.
- The acknowledgment that is sent for flow control can also be used for error control to tell the sender the packet has arrived uncorrupted.
- The lack of acknowledgment means that there is a problem in the sent frame.
- A frame that carries an acknowledgment is normally called an ACK to distinguish it from the data frame.

Connectionless and Connection-Oriented

- A DLC protocol can be either connectionless or connection-oriented.

1) Connectionless Protocol

Frames are sent from one node to the next without any relationship between the frames; each frame is independent. The term connectionless does not mean that there is no physical connection (transmission medium) between the nodes; it means that there is no connection between frames. The frames are not numbered and there is no sense of ordering. Most of the data-link protocols for LANs are connectionless protocols.

2) Connection Oriented Protocol

A logical connection should first be established between the two nodes (setup phase). After all frames that are somehow related to each other are transmitted (transfer phase), the logical connection is terminated (teardown phase). The frames are numbered and sent in order. If the frames are not received in order, the receiver needs to wait until all frames belonging to the same set are received and then deliver them in order to the network layer. Connection oriented protocols are rare in wired LANs, but we can see them in some point-to-point protocols, some wireless LANs, and some WANs.

4.2 DATA LINK LAYER PROTOCOLS

Traditionally 2 protocols have been defined for the data-link layer to deal with flow and error control: 1) Simple Protocol and 2) Stop-and-Wait Protocol.

The behavior of a data-link-layer protocol can be better shown as a finite state machine (FSM).
An FSM is a machine with a finite number of states. The machine is always in one of the states until an event occurs. Each event is associated with 2 reactions:

1) Defining the list (possibly empty) of actions to be performed.

2) Determining the next state (which can be the same as the current state).

One of the states must be defined as the initial state, the state in which the machine starts when it turns on.

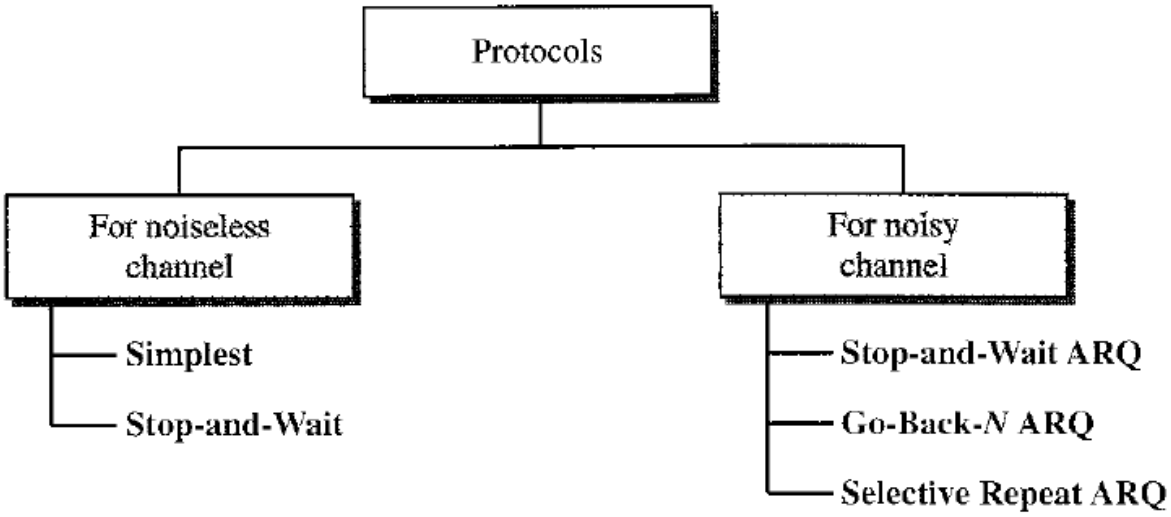


Figure 4.3 : Data Link Layer Protocols for flow control

4.2.1 Simplest Protocol

• Assumptions:

 The protocol has no flow-control or error-control.

 The protocol is a unidirectional protocol (in which frames are traveling in only one direction).

 The receiver can immediately handle any frame it receives.

Design

Here is how it works (Figure 4.4.):

1) At Sender

¤ The data-link-layer

→ gets data from its network-layer

→ makes a frame out of the data and

→ sends the frame.

2) At Receiver

¤ The data-link-layer

→ receives a frame from its physical layer

→ extracts data from the frame and

→ delivers the data to its network-layer.

 Data-link-layers of sender & receiver provide transmission services for their network-

layers.

 Data-link-layers use the services provided by their physical layers for the physical transmission of bits.

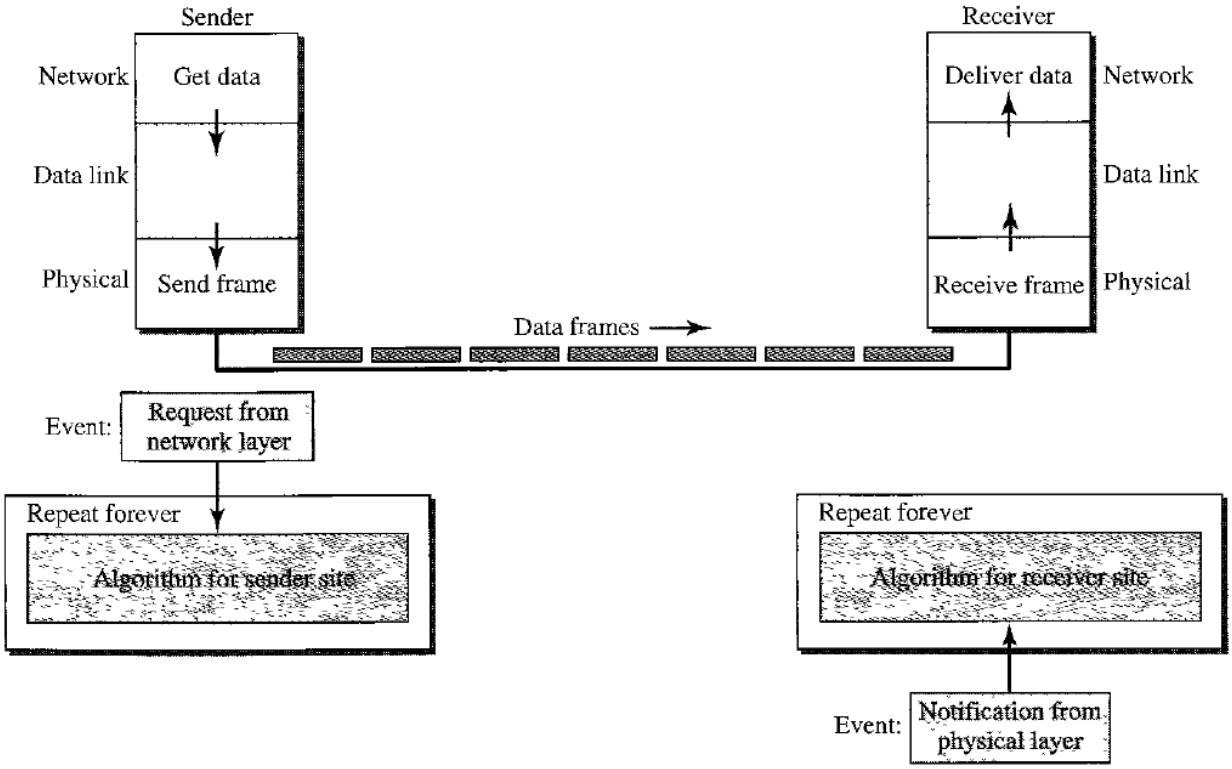


Figure 4.4 : Simplest Protocol for Noiseless Channel

4.2.2 Stop & Wait Protocol

As shown in Figure 4.5, this uses both flow and error control. Normally, the receiver has limited storage-space. If the receiver is receiving data from many sources, the receiver may

→ be overloaded with frames &

→ discard the frames.

To prevent the receiver from being overloaded with frames, we need to tell the sender to slow down.

Design

At Sender

 The sender

→ sends one frame & starts a timer

→ keeps a copy of the sent-frame and

→ waits for ACK-frame from the receiver (okay to go ahead).

 Then,

i) If an ACK-frame arrives before the timer expires, the timer is stopped and the sender sends the next frame. Also, the sender discards the copy of the previous frame.

ii) If the timer expires before ACK-frame arrives, the sender resends the previous frame and restarts the timer

2) At Receiver

- To detect corrupted frames, a CRC is added to each data frame.
- When a frame arrives at the receiver-site, the frame is checked.
- If frame's CRC is incorrect, the frame is corrupted and discarded.
- The silence of the receiver is a signal for the sender that a frame was either corrupted or lost.

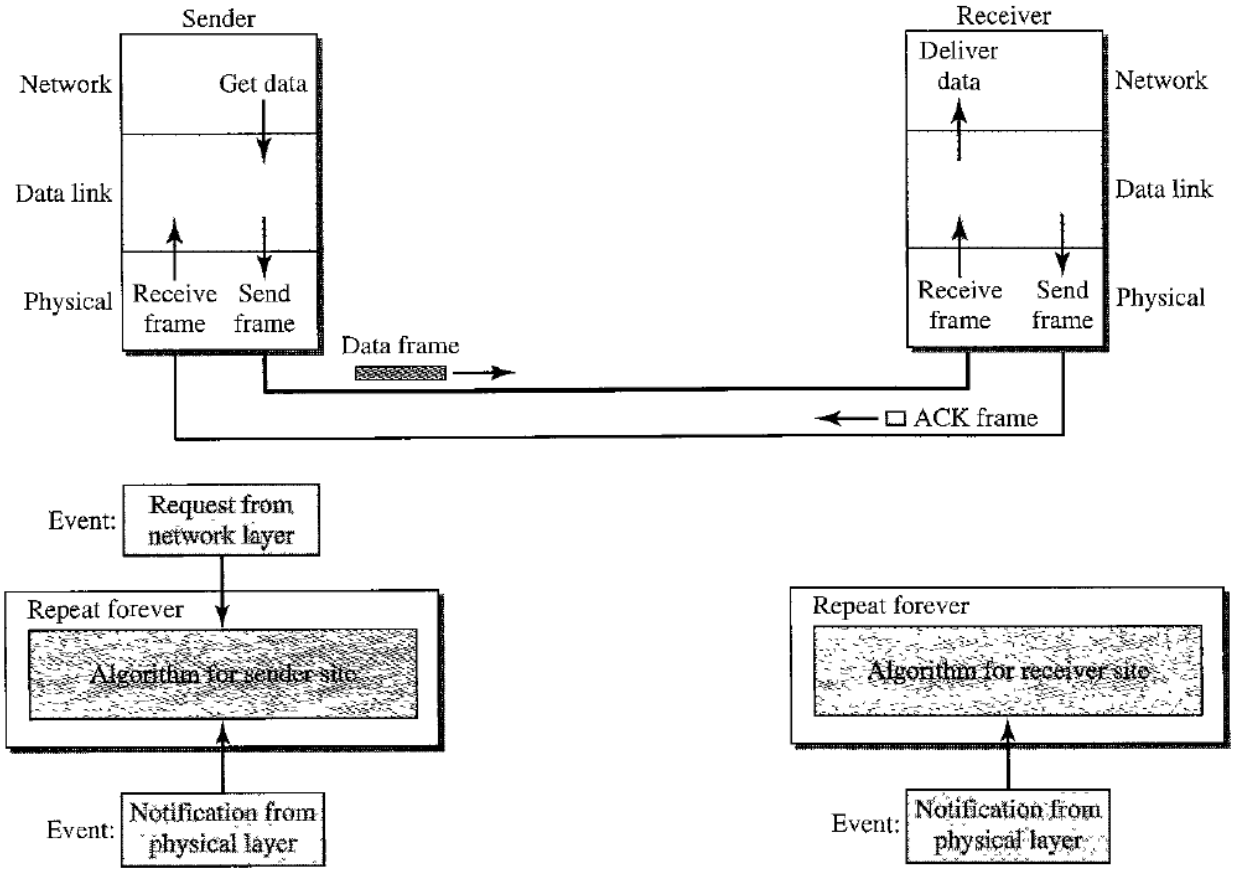


Figure 4.5 : Stop and Wait Method for Noiseless Channel

Noisy Channel

Practically, noiseless channel does not exists. Both protocols we discussed till now are for ideal channel which does not exists. But, understanding them helps in understanding complex algorithms for noisy channel. Frames, are numbered for differentiating with other frames.

Sequence and Acknowledgment Numbers

• Q: How to deal with corrupted-frame?

Ans: If the corrupted-frame arrives at the receiver-site, then the frame is simply discarded.

• Q: How to deal with lost-frames?

Ans: If the receiver receives out-of-order data-frame, then it means that frames were lost'. The lost-frames need to be resent.

• Problem in Stop and Wait protocols:

1) There is no way to identify a frame.

2) The received-frame could be the correct one, or a duplicate, or a frame out of order.

Solution: 1) Use sequence-number for each data frame.

2) Use Acknowledgment-number for each ACK frame.

Sequence Numbers

 Frames need to be numbered. This is done by using sequence-numbers.

 A sequence-number field is added to the data-frame.

Acknowledgment Numbers

An acknowledgment-number field is added to the ACK-frame.

 Sequence numbers are 0, 1, 0, 1, 0, 1, . . .

The acknowledgment numbers can also be 1, 0, 1, 0, 1, 0, ...

 The acknowledgment-numbers always announce the sequence-number of the next frame expected by the receiver.

 For example, If frame-0 has arrived safely, the receiver sends an ACK-frame with acknowledgment-1 (meaning frame-1 is expected next).

Piggybacking

A technique called piggybacking is used to improve the efficiency of the bidirectional protocols. The data in one direction is piggybacked with the acknowledgment in the other direction. In other words, when node A is sending data to node B, Node A also acknowledges the data received from node B.

Module 5

IEEE 802 is a family of IEEE standards dealing with local area networks and metropolitan area networks. LAN market has seen many technologies such as Ethernet, Token Ring, Token Bus, FDDI, ATM LAN. Ethernet known as IEEE 802.3, is one of the most widely used standards for computer networking and general data communications. It is widely used in all forms of data networking from connecting to home Wi-Fi hubs to business data networks and telecommunications networking. The most widely used standards for the Ethernet family are Token Ring, Wireless LAN (Wi-Fi), Bridging and Virtual Bridged LANs. The groups are numbered from 802.1 to 802.12. IEEE 802 splits the OSI Data Link Layer into two sub-layers named logical link control (LLC) and media access control (MAC), so the layers can be listed as LLC & MAC Sublayers. Figure 5.1 shows how IEEE is different than OSI model at datalink layer.

- Data link layer
  - LLC sublayer

    * Flow-control, error-control, and framing duties are grouped into one sublayer called LLC.
    * Framing is handled in both the LLC and the MAC.
    * LLC vs. MAC

      LLC provides one single data-link-control protocol for all IEEE LANs.

      MAC provides different protocols for different LANs.
    * A single LLC protocol can provide interconnectivity between different LANs because it makes the MAC sublayer transparent.

  - MAC sublayer

    * This defines the specific access-method for each LAN.

    For example:

      i) CSMA/CD is used for Ethernet LANs.

      ii) Token-passing method is used for Token Ring and Token Bus LANs.

    * The framing function is also handled by the MAC layer.
    * The MAC contains a number of distinct modules.
    * Each module defines the access-method and the framing-format specific to the corresponding LAN protocol.

LLC handles Error Control, Flow Control and a part of Framing. Framming happens in both LLC and MAC sub layers. LLC provides one single protocol for all IEEE LANs where as there are different MAC protocols for different LANs.

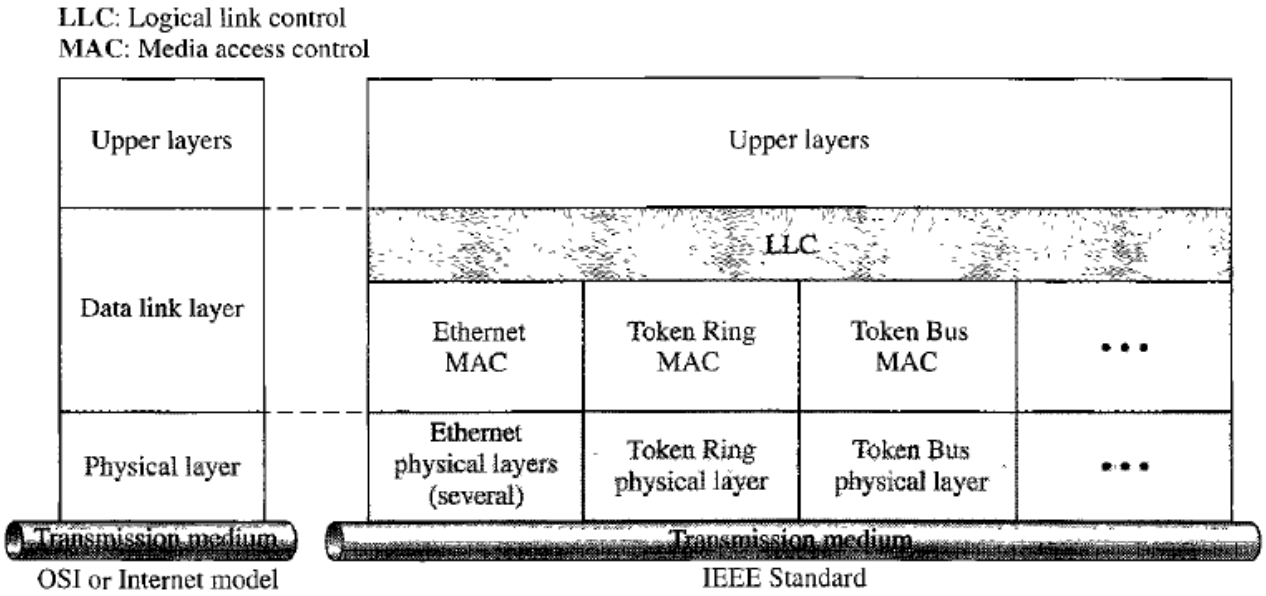


Figure 5.1 : IEEE Standard for LANs

5.1 Ethernet Evolution

There are Four generations of Ethernet (Figure 5.2) and every generation gives higher speed for applications.

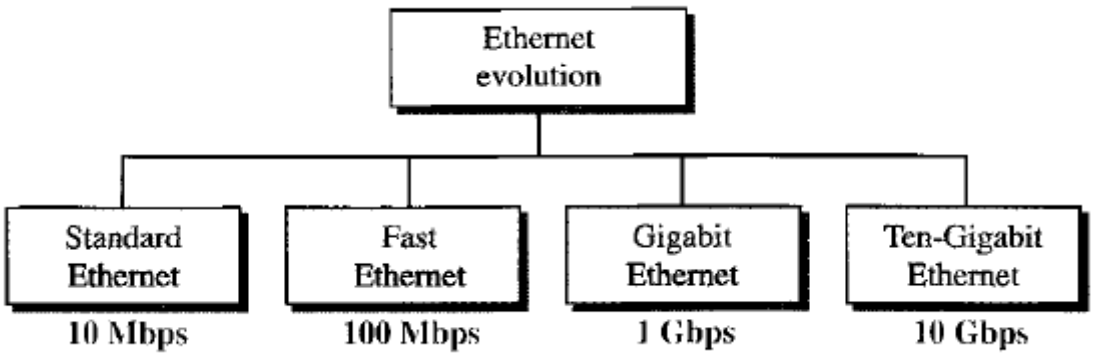


Figure 5.2 : Generations of Ethernet

Ethernet defines its own format for its frames. IEEE 802.3 is the standard from IEEE for it. This format remains same irrespective of different types of ethernet. Different fields of frame are as given in Figure 5.2.

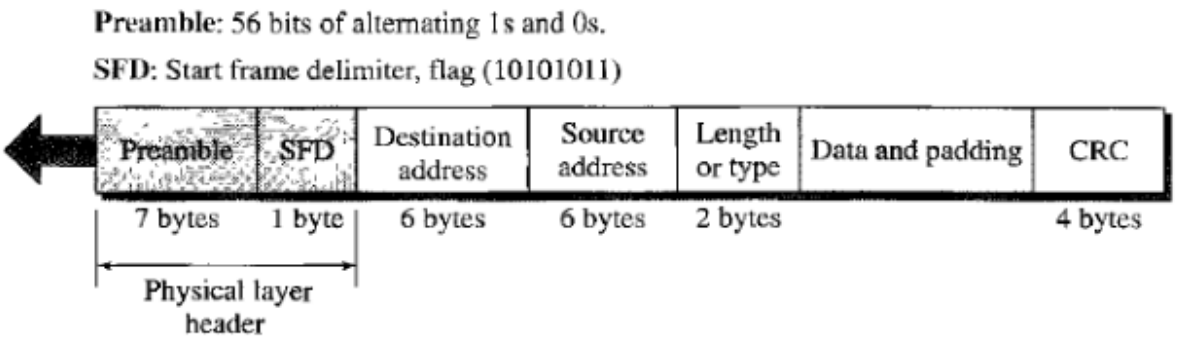


Figure 5.3 : Ethernet Frame Format

Preamble : Seven bytes of alternate 1s and 0s form this field. 56 bits simply provide timings to the receiver and its for synchronization of receiver with sender timings.

SFD : Start Frame Delimiter, One byte in length, is 10101011. Last bit, instead of 0, is 1 to indicate to receiver that whatever follows this bit is the begining of destination address and next byte of header. No more message to synchronization after this bit.

Destination Address : This field is 48 bit physical address of receiver. It should be included in

frame so that intermediate systems farward it to right path so that receive receives it.

Source Address : This field contains the physical-address of the sender-station.

Type or Length : This field is defined as a i) type field or ii) length field.

i) In original Ethernet, this field is used as the type field. Type field defines the upper-layer protocol using the MAC frame.

ii) In IEEE standard, this field is used as the length field. Length field defines the number of bytes in the data-field.

Data & Padding : This field carries data encapsulated from the upper-layer protocols.

Minimum data size = 46 bytes. Maximum data size = 1500 bytes.

CRC : This field contains error detection information such as a CRC-32.

Frame Length

Ethernet has imposed restrictions on both minimum & maximum lengths of a frame (Figure 5.4). The minimum length restriction is required for the correct operation of CSMA/CD.

• Minimum length of frame = 64 bytes.

1) Minimum data size = 46 bytes.

2) Header size + Trailer size = 14 + 4 = 18 bytes.

(i.e. 18 bytes □ 6 bytes source-address + 6 bytes dest-address + 2 bytes length + 4 bytes CRC).

• The minimum length of data from the upper layer = 46 bytes.

• If the upper-layer packet is less than 46 bytes, padding is added to make up the difference.

• Maximum length of frame =1518 bytes.

1) Maximum data size = 1500 bytes.

2) Header size + trailer size = 14 + 4 = 18 bytes.

• The maximum length restriction has 2 reasons:

1) Memory was very expensive when Ethernet was designed.

A maximum length restriction helped to reduce the size of the buffer.

2) This restriction prevents one station from

→ monopolizing the shared medium

→ blocking other stations that have data to send.

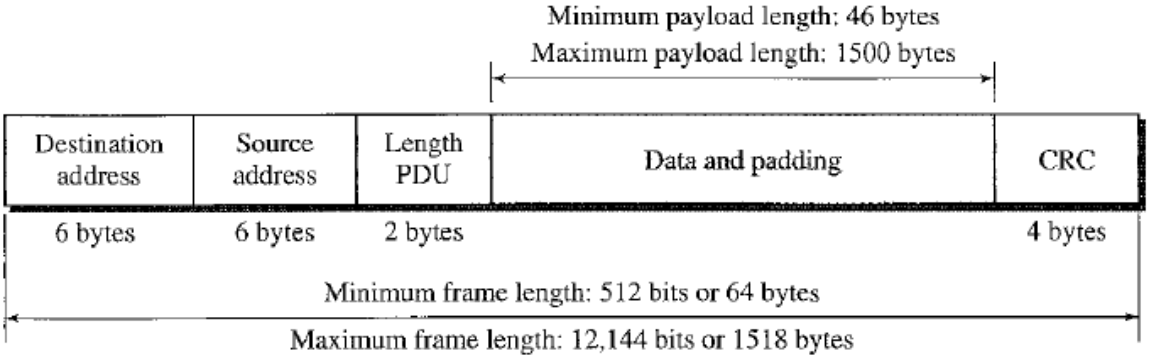


Figure 5.4 : Ethernet Frame fields and length

Addressing

In an Ethernet-network, each station has its own NIC (6-byte=48 bits). The NIC provides the station with a 6-byte physical-address (or Ethernet-address). For example, Ethernet MAC address looks something like this -> 06 : 01 : 02 : 01 : 2C : 4B

Unicast, Multicast, and Broadcast Addresses

A source-address is always a unicast address i.e. the frame comes from only one station.

However, the destination-address can be 1) Unicast 2) Multicast or 3) Broadcast.

• As shown in Figure 5.5, if LSB of first byte in a destination-address is 0,

Then, the address is unicast;

Otherwise, the address is multicast.

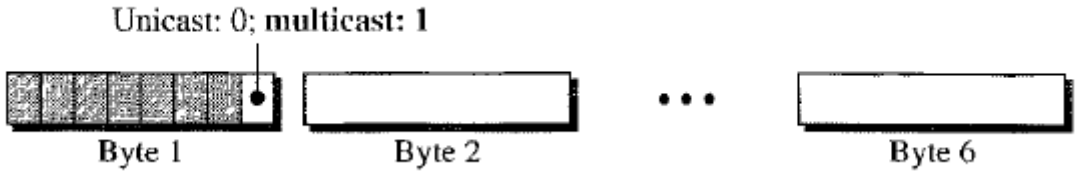


Figure 5.5 : Types of Physical Address.

1) A unicast destination-address defines only one recipient.

¤ The relationship between the sender and the receiver is one-to-one.

2) A multicast destination-address defines a group of addresses.

¤ The relationship between the sender and the receivers is one-to-many.

3) The broadcast address is a special case of the multicast address.

¤ The recipients are all the stations on the LAN.

¤ A broadcast destination-address is 48 1s (6-byte □ 48 bits).

5.1.1 Standard Ethernet :

The original Ethernet technology with data-rate of 10 Mbps are referred to as the Standard Ethernet. Standard-Ethernet uses 1-persistent CSMA/CD.

1) Slot Time

Slot time = round-trip time + time required to send the jam sequence.

• The RTT means time required for a frame to travel from one end of a maximum-length network to the other end (RTT □ round-trip time).

• The slot time is defined in bits.

• The slot time is the time required for a station to send 512 bits.

• The actual slot time depends on the data-rate.

For example: 10-Mbps Ethernet has slot time of 51.2 μs.

2) Slot Time and Collision

• The choice of a 512-bit slot time was not accidental.

• It was chosen to allow the proper functioning of CSMA/CD.

3) Slot Time and Maximum Network Length

• There is a relationship between

→ slot time and

→ maximum length of the network (collision domain).

• This relationship is dependent on the propagation-speed of the signal in the particular medium.

i) In most transmission media, the signal propagates at 2x10 8 m/s (two-thirds of the rate for propagation in air).

ii) For traditional Ethernet, we calculate

MaxLength = Propagation Speed X (Slot Time/2)

MaxLength= (2 x 10 8) x (51.2 X 10 -6)/2= 5120m

The efficiency is defined as the ratio of the time used by a station to send data to the time the medium is occupied by this station. The practical efficiency of standard Ethernet has been measured to be

$$\text{Efficiency} = 1 / (1 + 6.4 \text{ X } a)$$

where a = number of frames that can fit on the medium.

a = (propagation delay)/(transmission delay)

• As the value of parameter a decreases, the efficiency increases.

• If the length of the media is shorter or the frame size longer, the efficiency increases.

• In the ideal case, a = 0 and the efficiency is 1.

The Standard-Ethernet defines several physical-layer implementations as given in Figure 5.6

Characteristics	*10Base5*	*10Base2*	*10Base-T*	*10Base-F*
Media	Thick coaxial cable	Thin coaxial cable	2 UTP	2 Fiber
Maximum length	500 m	185 m	100 m	2000 m
Line encoding	Manchester	Manchester	Manchester	Manchester

Figure 5.6 : Summary of Standard Ethernet implementations.

All standard implementations use digital-signaling (baseband) at 10 Mbps.

1) At the sender, data are converted to a digital-signal using the Manchester scheme as shown in Figure 5.7.

2) At the receiver, the received-signal is

→ interpreted as Manchester and

→ decoded into data.

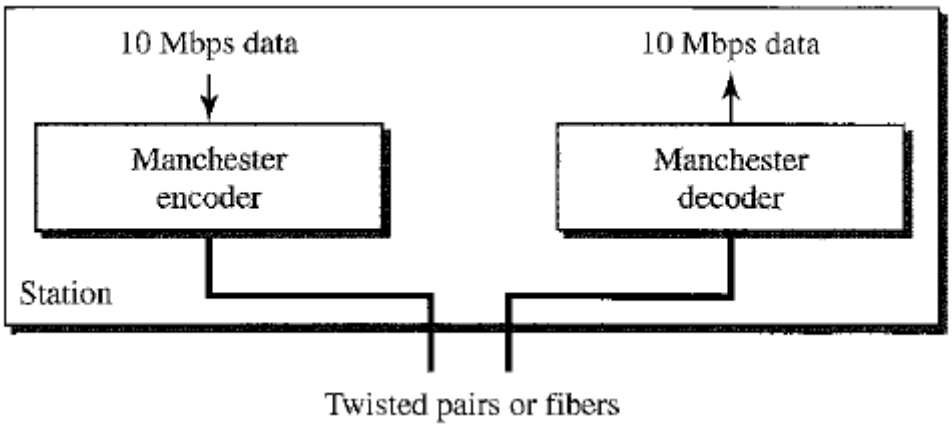


Figure 5.7 : Encoding and Decoding in Standard Ethernet

5.1.1.1 10Base5: Thick Ethernet

- 10Base5 uses a bus topology (Figure 5.8).
- A external transceiver is connected to a thick coaxial-cable.

  (transceiver transmitter/receiver)
- The transceiver is responsible for
  - → transmitting
  - → receiving and
  - → detecting collisions.
- The transceiver is connected to the station via a coaxial-cable.

The cable provides separate paths for sending and receiving.

The collision can only happen in the coaxial cable.

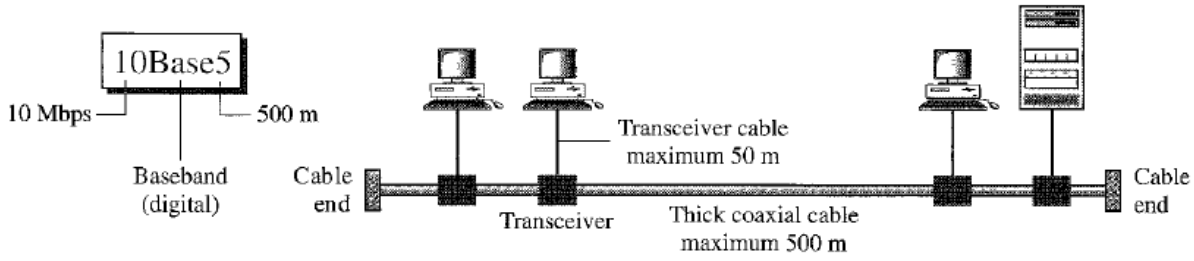


Figure 5.8 : Thick Ethernet

The maximum-length of the cable must not exceed 500m. If maximum-length is exceeded, then there will be excessive degradation of the signal. If a cable-length of more than 500 m is needed, the total cable-length can be divided into up to 5 segments. Each segment of maximum length 500-meter, can be connected using repeaters.

5.1.1.2 10Base2: Thin Ethernet

 10Base2 uses a bus topology (Figure 5.9).

 The cable is much thinner and more flexible than 10Base5.

 Flexible means the cable can be bent to pass very close to the stations.

 The transceiver is part of the NIC, which is installed inside the station.

 The collision can only happen in the coaxial cable.

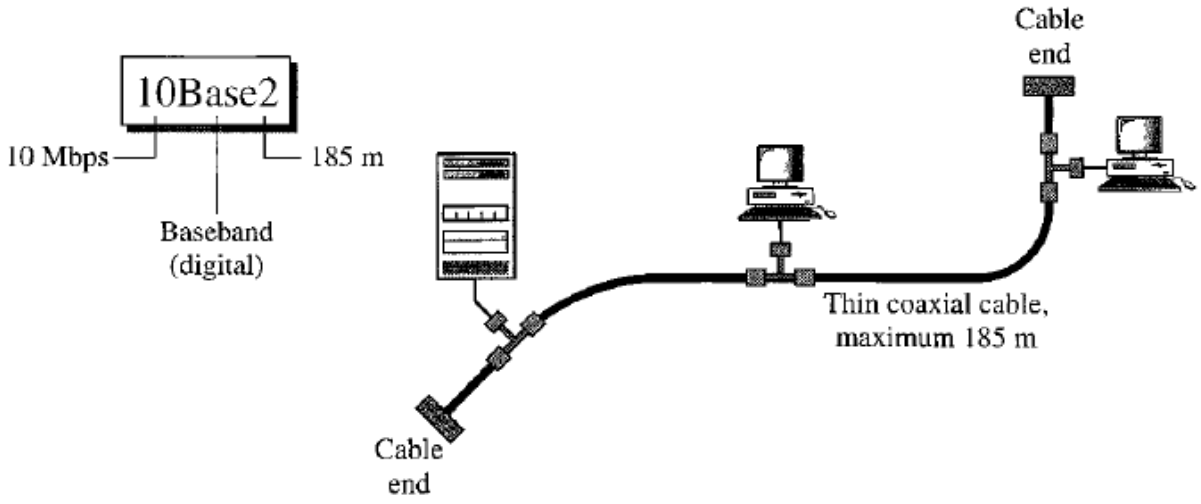


Figure 5.9 : Thin Ethernet

Advantages:

1) Thin coaxial-cable is less expensive than thick coaxial-cable.

2) Tee connections are much cheaper than taps.

3) Installation is simpler because the thin coaxial cable is very flexible.

Disadvantage: Length of each segment cannot exceed 185m due to the high attenuation in the cable.

5.1.1.3 10Base- T: Twisted Pair Ethernet

 10Base-T uses a star topology to connect stations to a hub (Figure 5.10).

 The stations are connected to a hub using two pairs of twisted-cable.

Two pairs of twisted cable create two paths between the station and the hub.

1) First path for sending.

2) Second path for receiving.

 The collision can happen in the hub.

 The maximum length of the cable is 100 m. This minimizes the effect of attenuation in the cable.

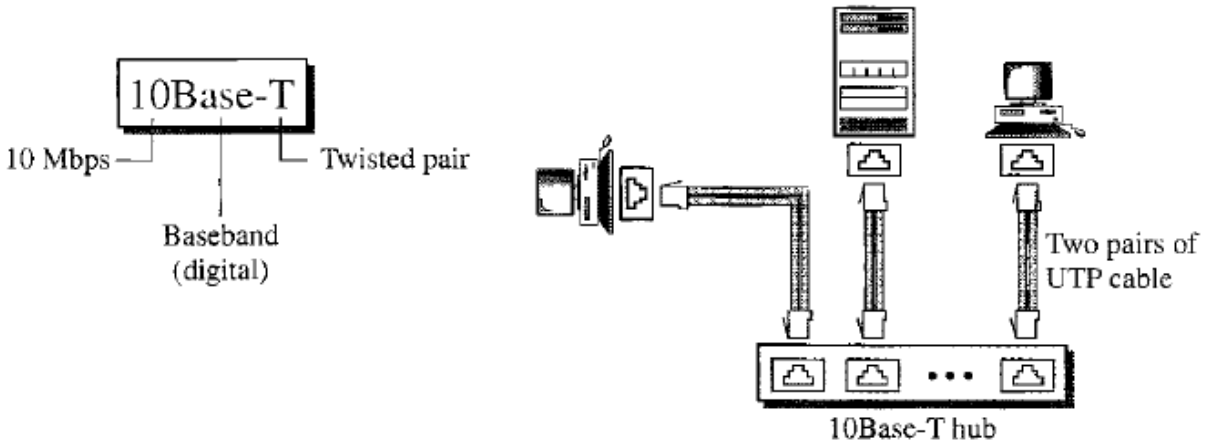


Figure 5.10 : 10Base- T - Twisted Pair Ethernet

5.1.1.4 10Base-F: Fiber Ethernet

- 10Base-F uses a star topology to connect stations to a hub (Figure 5.11).
- The stations are connected to the hub using two fiber-optic cables.

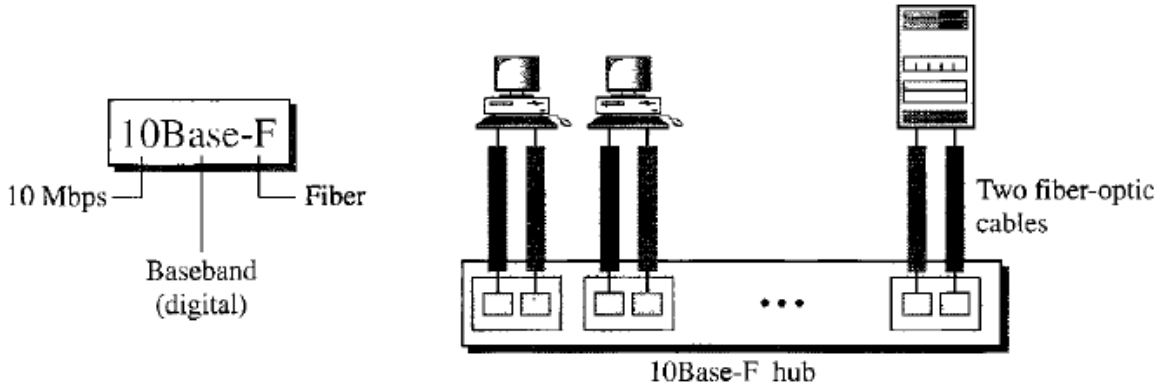


Figure 5.11 : 10Base- F - Fiber Ethernet

5.1.2 Changes in the Standard

Certain changes were suggested to basic version of standard Ethernet as given below that lead to lot of improvements such as increase in bandwidth, reduction in collision and complexities etc.

5.1.2.1 Bridged Ethernet

• Bridges have two effects on an Ethernet LAN:

i) They raise the bandwidth &

- A bridge divides the network into two or more networks.
- Bandwidth-wise, each network is independent.

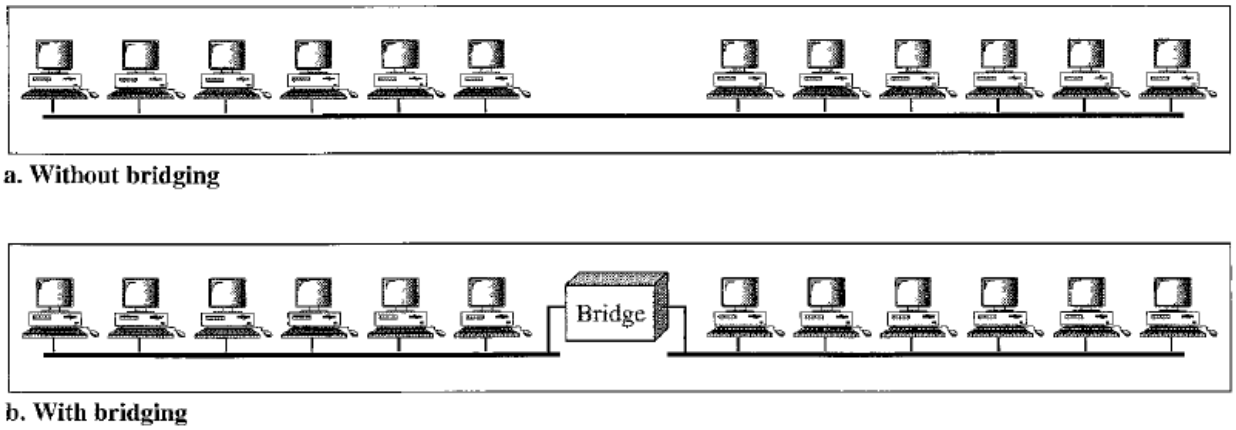
For example (Figure 5.12):

- A network with 12 stations is divided into two networks, each with 6 stations.

 Now each network has a capacity of 10 Mbps.

 The 10-Mbps capacity in each segment is now shared between 6 stations (actually 7 because the bridge acts as a station in each segment), not 12 stations.

 In a network with a heavy load, each station theoretically is offered 10/6 Mbps instead of 10/12 Mbps.



a. Without bridging

b. With bridging

a. Without bridging

b. With bridging

Figure 5.12 : Brdiged Ethernet

ii) Separating Collision Domains

• Another advantage of a bridge is the separation of the collision domain.

• Figure 5.12 shows the collision domains for an un-bridged and a bridged network.

• You can see that the collision domain becomes much smaller and the probability of collision is reduced tremendously.

5.1.2.2 Switched Ethernet

• The idea of a bridged LAN can be extended to a switched LAN (Figure 5.13).

• If we can have a multiple-port bridge, we can have an N-port switch.

• In this way, the bridgeid is shared only between the station and the switch.

• A layer-2 switch is an N-port bridge with additional sophistication that allows faster handling of the packets.

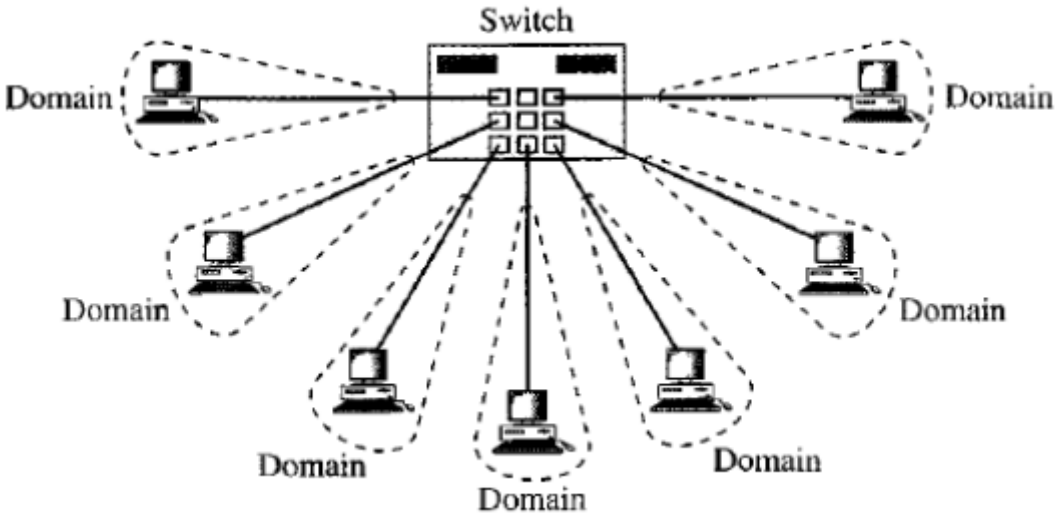


Figure 5.13 : Switched Ethernet

5.1.2.3

Full-Duplex Ethernet

• The full-duplex mode increases the capacity of each domain from 10 to 20 Mbps.

• Instead of using one link between the station and the switch, the configuration uses two links: one to transmit and one to receive. Hence,

No Need for CSMA/CD

 In full-duplex switched Ethernet,

 There is no need for the CSMA/CD method.

 Each station is connected to the switch via two separate links.

 Each station or switch can send and receive independently without worrying about collision.

 Each link is a point-to-point dedicated path between the station and the switch.

 There is no longer a need for carrier sensing; there is no longer a need for collision-detection.

 The job of the MAC layer becomes much easier.

 Carrier sensing and collision-detection functionalities of the MAC sublayer can be turned off.

MAC Control Layer

 To provide for flow and error control in full-duplex switched Ethernet, a new sublayer, called the MAC control, is added between the LLC sublayer and the MAC sublayer.

5.1.3 FAST ETHERNET (100 MBPS)

IEEE created Fast-Ethernet under the name 802.3u. Fast-Ethernet was designed to compete with LAN protocols such as FDDI or Fiber Channel. Features of Fast-Ethernet are,

1) Upgrade the data-rate to 100 Mbps.
2) Make it compatible with Standard-Ethernet.
3) Keep the same 48-bit address.
4) Keep the same frame format.
5) Keep the same minimum and maximum frame-lengths.

Access Method

• Access method is same as in Standard-Ethernet.

• Only the star topology is used.

• For the star topology, there are 2 choices:

1) In the half-duplex approach, the stations are connected via a hub. CSMA/CD was used as access-method.
2) In the full-duplex approach, the connection is made via a switch with buffers at each port.

There is no need for CSMA/CD.

Autonegotiation

 A new feature added to Fast-Ethernet is called autonegotiation.

 It provides a station/hub with a range of capabilities.

 It was used for the following purposes:

1) To allow 2 devices to negotiate the mode or data-rate of operation.
2) To allow incompatible devices to connect to one another.

   For example: a device with a maximum capacity of 10 Mbps can communicate with a device with a 100 Mbps capacity.
3) To allow one device to have multiple capabilities.
4) To allow a station to check a hub's capabilities.

Implementation

• Fast-Ethernet can be classified as either a two-wire or a four-wire implementation (Figure 5.14).

1) The 2-wire implementations use

→ Category 5 UTP (100Base-TX) or

→ Fiber-optic cable (100Base-FX)

2) The 4-wire implementations use category 3 UTP (100Base-T4).

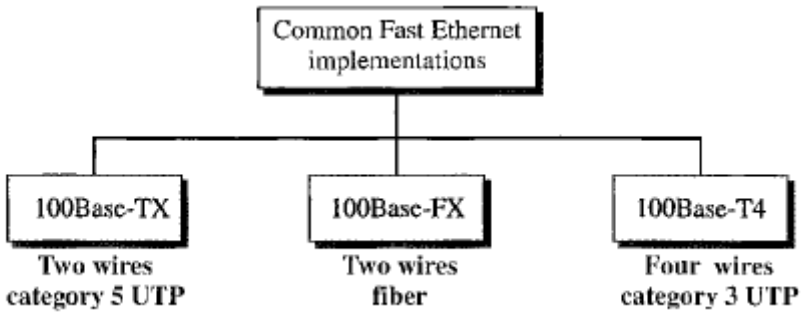


Figure 5.14 : Implementations of Fast Ethernet

Physical Layer

• The physical-layer in Fast-Ethernet is more complicated than the one in Standard-Ethernet.

• Some of the features of this layer are as follows. 1) Topology 2) Implementation and 3) Encoding.

Topology

• Fast-Ethernet is used to connect two or more stations together (Figure 5.15).

1) If there are only 2 stations, they can be connected in point-to-point.

2) If there are 3 or more stations, they can be connected in star topology with a hub at the center.

Characteristics	*100Base-TX*	*100Base-FX*	*100Base-T4*
Media	Cat 5 UTP or STP	Fiber	Cat 4 UTP
Number of wires	2	2	4
Maximum length	100 m	100 m	100 m
Block encoding	4B/5B	4B/5B	
Line encoding	MLT-3	NRZ-I	8B/6T

Figure 5.15 : Summary of Fast Ethernet types

Encoding

• There are 3 different encoding schemes.

i) 100Base-TX

 This uses 2 pairs of twisted-pair cable (either category 5 UTP or STP) (Figure 5.16a).

 The MLT-3 encoding scheme is used for implementation.

This is because MLT-3 has good bandwidth performance.

 However, 4B/5B block-coding is used to provide bit synchronization.

This is because MLT-3 is not a self-synchronous line coding scheme.

 4B/5B coding creates a data-rate of 125 Mbps, which is fed into MLT-3 for encoding.

ii) 100Base-FX

 This uses 2 pairs of fiber-optic cables (Figure 5.16b).

 Optical fiber can easily handle high bandwidth requirements.

 The NRZ-I encoding scheme is used for implementation.

 However, 4B/5B block-coding is used to provide bit synchronization.

This is because NRZ-I is not a self-synchronous line coding scheme.

 4B/5B encoding increases the bit rate from 100 to 125 Mbps, which can easily be handled by fiber-optic cable.

iii) 100Base-T4

 This uses 4 pairs of UTP for transmitting 100 Mbps (Figure 5.16c).

 Each UTP cannot easily handle more than 25 Mbaud.

 One pair switches between sending and receiving.

 Three pairs of UTP can handle only 75 Mbaud (25 Mbaud) each.

 Encoding/decoding is more complicated.

 We need an encoding scheme that converts 100 Mbps to a 75 Mbaud signal. This requirement is satisfied by 8B/6T.

 The 8B/6T encoding scheme is used for implementation.

a) 8 data elements are encoded as 6 signal elements.

b) This means that 100 Mbps uses only (6/8) x 100 Mbps, or 75 Mbaud.

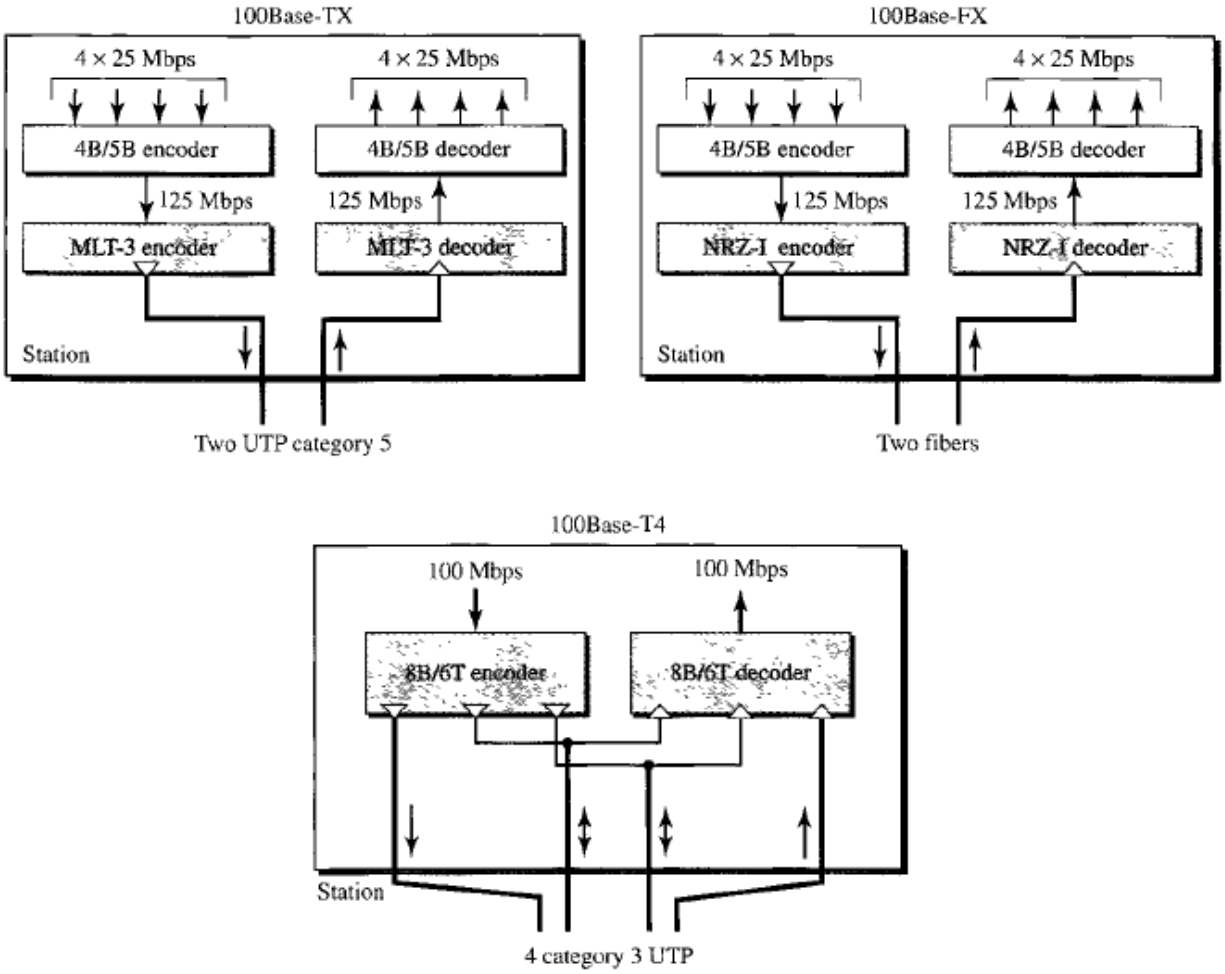


Figure 5.16 : Diffrent Implementations of Fast Ethernet

5.1.4 GIGABIT ETHERNET

IEEE created Gigabit-Ethernet under the name 802.3z. Features of Gigabit-Ethernet are,

1) Upgrade the data-rate to 1 Gbps.

2) Make it compatible with Standard or Fast-Ethernet.

3) Use the same 48-bit address.

4) Use the same frame format.

5) Keep the same minimum and maximum frame-lengths.

6) To support auto-negotiation as defined in Fast-Ethernet.

MAC Sublayer

Gigabit-Ethernet has two distinctive approaches for medium access: half-duplex and full-duplex. Almost all implementations of Gigabit-Ethernet follow the full-duplex approach.

1) Full Duplex Mode

There is a central switch connected to all computers or other switches. Each switch has buffers for each input-port in which data are stored until they are transmitted. There is no collision. This means that CSMA/CD is not used.

Lack of collision implies that the maximum length of the cable is determined

→ by the signal attenuation in the cable &

→ not by the collision-detection process.

2) Half Duplex Mode

A switch is replaced by a hub, which acts as the common cable in which a collision might occur.CSMA/CD is used. The maximum length of the network is totally dependent on the minimum frame size. Three methods have been defined: traditional, carder extension, and frame bursting.

i) Traditional

- Like traditional Ethernet, the minimum length of a frame is 512 bits.
- However, because the length of a bit is 1/100 shorter,

Slot time is 512 bits x 1/1000 gs which is equal to 0.512 gs.

- The reduced slot time means that collision is detected 100 times earlier.
- The maximum length of the network is 25 m.
- This length may be suitable if all the stations are in one room.

ii) Carrier Extension

- To allow for a longer network, we increase the minimum frame-length.
- Minimum length of frame is 512 bytes (4096 bits). Thus, minimum length is 8 times longer.
- A station adds extension bits (padding) to any frame that is less than 4096 bits.
- The maximum length of the network is 200 m.
- A length from the hub to the station is 100 m.

iii) Frame Bursting

- Carrier extension is very inefficient if
  - → we have a series of short frames to send
  - → each frame carries redundant data.
- To improve efficiency, frame bursting was proposed.
- Instead of adding an extension to each frame, multiple frames are sent.
- However, to make these multiple frames look like one frame, padding is added between the frames. Thus, the channel is not idle Physical Layer. The physical-layer in Gigabit-Ethernet is more complicated than that in Standard or Fast-Ethernet.

Some of the features of this layer are as follows. 1) Topology 2) Implementation and 3) Encoding.

Topology

Gigabit-Ethernet is used to connect two or more stations together.

1) If there are only 2 stations, they can be connected in point-to-point.

2) If there are 3 or more stations, they can be connected in star topology with a hub at center.

Implementation

Gigabit-Ethernet can be classified as either a two-wire or a four-wire implementation (Figure 5.17).

1) The 2-wire implementations use

→ Fiber-optic cable (1000Base-SX, short-wave, or 1000Base-LX, long-wave) or

→ STP (1000Base-CX)

2) The 4-wire implementations use category 5 twisted-pair cable (1000Base-T).

Characteristics	*1000Base-SX*	*1000Base-LX*	*1000Base-CX*	*1000Base-T*
Media	Fiber short-wave	Fiber long-wave	STP	Cat 5 UTP
Number of wires	2	2	2	4
Maximum length	550 m	5000 m	25 m	100 m
Block encoding	8B/10B	8B/10B	8B/10B	
Line encoding	NRZ	NRZ	NRZ	4D-PAM5

Figure 5.17 : Summary of Gigabit Ethernet Implementations

Encoding

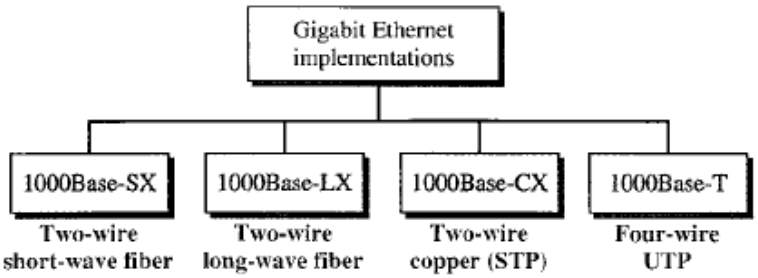
1) Two Wire Implementation

• The NRZ encoding scheme is used for two-wire implementation (Figure 5.18). However, 8B/10B block-coding is used to provide bit synchronization. This is because NRZ is not a self-synchronous line coding scheme. 8B/10B coding creates a data-rate of 1.25 Gbps. One wire (fiber or STP) is used for sending. Another wire is used for receiving.

2) Four Wire Implementation

In this, it is not possible to have 2 wires for input and 2 for output (Figure 5.18). This is '.' each wire would need to carry 500 Mbps, which exceeds the capacity for category 5 UTP. As a solution,

4D-PAM5 encoding is used to reduce the bandwidth. Thus, all four wires are involved in both input and output. Each wire carries 250 Mbps, which is in the range for category 5 UTP cable.



Encoding

Figure 13.24 shows the encoding/decoding schemes for the four implementations.

Figure 13.24 *Encoding in Gigabit Ethernet implementations*

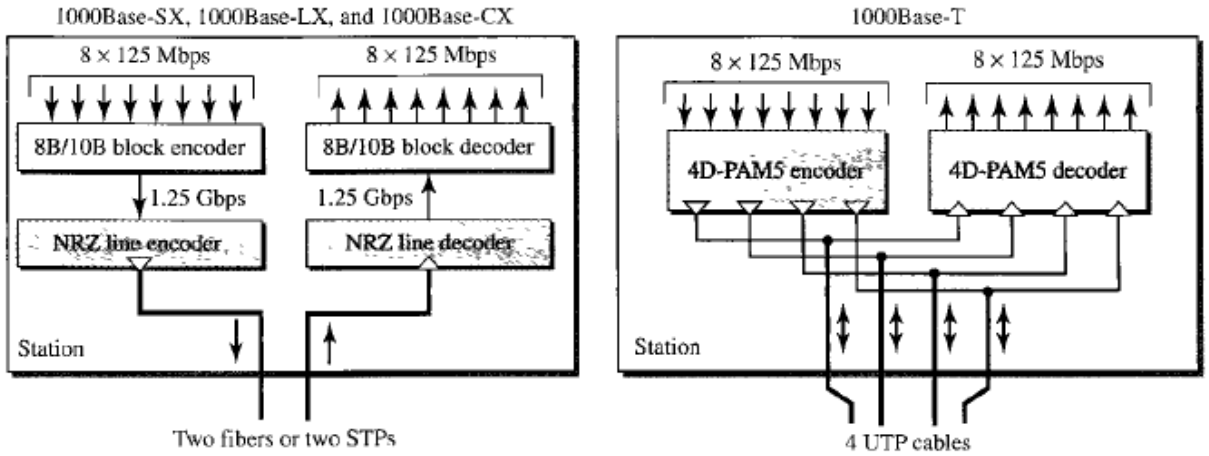


Figure 5.18 : Implememntations of Gigabit Ethernet

5.1.5 TEN GIGABIT ETHERNET

IEEE created Ten-Gigabit-Ethernet under the name 802.3ae. Goals of the Gigabit-Ethernet are,

1) Upgrade the data-rate to 10 Gbps.
2) Make it compatible with Standard, Fast, and Gigabit-Ethernet.
3) Use the same 48-bit address.
4) Use the same frame format.
5) Keep the same minimum and maximum frame-lengths.
6) Allow the interconnection of existing LANs into a MAN or a WAN .
7) Make Ethernet compatible with technologies such as Frame Relay and ATM.

Implementation

Ten-Gigabit-Ethernet operates only in full duplex mode.This means there is no need for contention; CSMA/CD is not used. Four implementations are the most common (Figure 5.19):

1) 10GBase-SR
2) 10GBase-LR
3) 10GBase-EW and
4) 10GBase-X4

Characteristics	*10GBase-S*	*10GBase-L*	*10GBase-E*
Media	Short-wave 850-nm multimode	Long-wave 1310-nm single mode	Extended 1550-mm single mode
Maximum length	300 m	10 km	40 km

Figure 5.19 : Summary of Ten Gigabit Ethernet

5.2 INTRODUCTION OF WIRELESS LANS

Ethernet can be used as wireless. Its very cumbersome to use wires now a days in office/homes etc. People prefer wireless everywhere and thats why technolgy has evolved to blue tooth keyboards, blue tooth mouse etc. In a wired LAN, we use wires to connect hosts. In a switched LAN, with a link-layer switch, the communication between the hosts is point-to-point and full-duplex (bidirectional). In a wireless LAN, the medium is air, the signal is generally broadcast. When hosts in a wireless LAN communicate with each other, they are sharing the same medium (multiple access).

In a wired LAN, a host is always connected to its network at a point with a fixed link layer address related to its network interface card (NIC). Of course, a host can move from one point in the Internet to another point. In this case, its link-layer address remains the same, but its network-layer address will change. In a wireless LAN, a host is not physically connected to the network; it can move freely and can use the services provided by the network. Therefore, mobility in a wired network and wireless network are totally different issues.

A wired isolated LAN is a set of hosts connected via a link-layer switch (Figure 5.20). A wireless isolated LAN, called an ad hoc network in wireless LAN terminology, is a set of hosts that communicate freely with each other. The concept of a link-layer switch does not exist in wireless LANs.

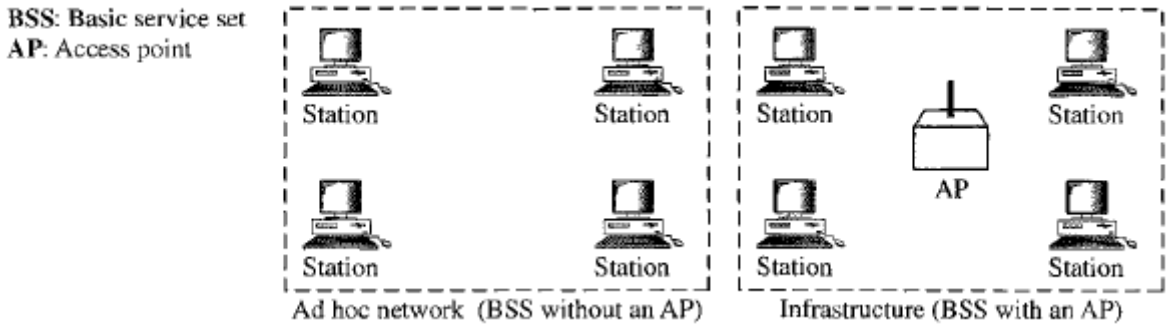


Figure 5.20 : Two types of wireless networks

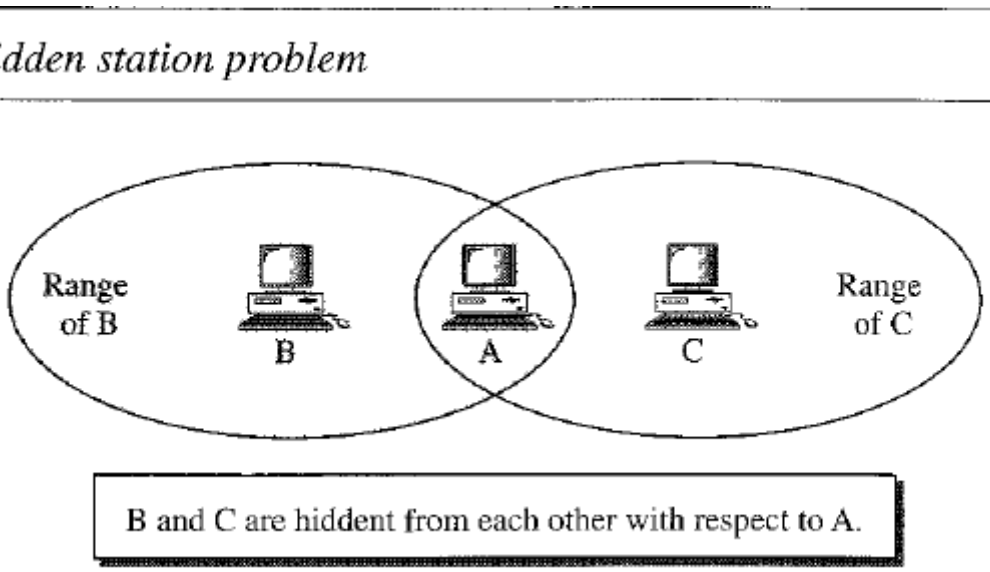
Connection to Other Networks

A wired LAN can be connected to another network or the Internet using a router. A wireless LAN may be connected to a wired infrastructure network, to a wireless infrastructure network, or to another wireless LAN (Figure 5.20). In this case, the wireless LAN is referred to as an infrastructure network, and the connection to the wired infrastructure, such as the Internet, is done via a device called an access point (AP). An access point is gluing two different environments together: one wired and one wireless.

1) Communication between the AP and the wireless host occurs in a wireless environment.
2) Communication between the AP and the infrastructure occurs in a wired environment.

Two types of problems in wireless networks are Hidden Station Problem and Exposed Station Problem.

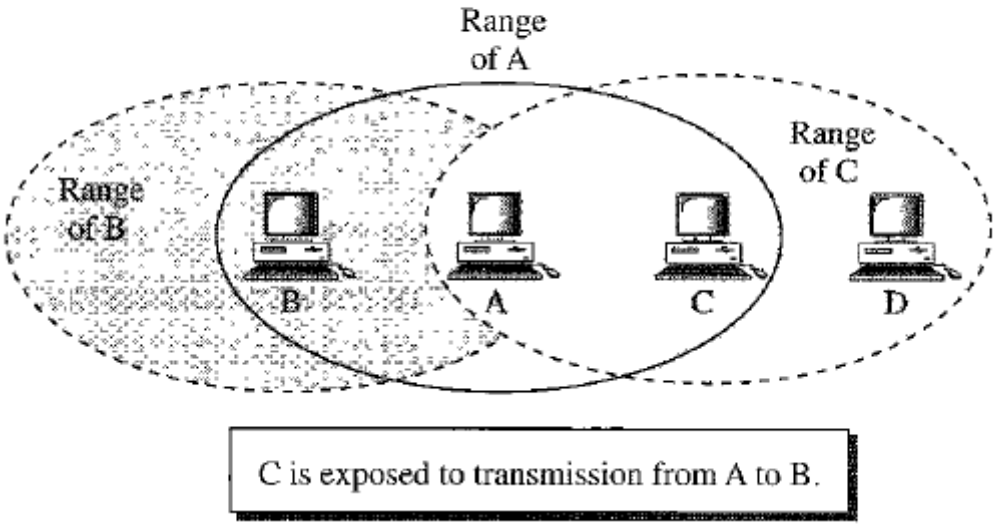
Figure 5.21 : Hidden Station Problem



Hidden Station Problem

Figure 5.21 shows an example of the hidden station problem. Every station in transmission range of Station B can hear any signal transmitted by station B. Every station in transmission range of Station C can hear any signal transmitted by station C. Station C is outside the transmission range of B. Likewise, station B is outside the transmission range of C. However, Station A is in the area covered by both B and C; Therefore, Station A can hear any signal transmitted by B or C.

Figure 5.21 : Exposed Station Problem



Exposed Station Problem

In this problem, a station refrains from using a channel even when the channel is available for use.

In the figure 5.21, station A is transmitting to station B. Station C has some data to send to station D, which can be sent without interfering with the transmission from A to B. However, station C is exposed to transmission from A i.e. station C hears what A is sending and thus refrains from sending. In other words, C is too conservative and wastes the capacity of the channel. The handshaking messages RTS and CTS cannot help in this case. Station C hears the RTS from A, but does not hear the CTS from B. Station C, after hearing the RTS from A, can wait for a time so that the CTS from B reaches A; it then sends an RTS to D to show that it needs to communicate with D. Both stations B and A may hear this RTS, but station A is in the sending state, not the receiving state.However, Station B responds with a CTS. The problem is here (Figure 5.21). If station A has started sending its data, station C cannot hear the CTS from station D because of the collision; it cannot send its data to D. It remains exposed until A finishes sending its data as the figure shows.

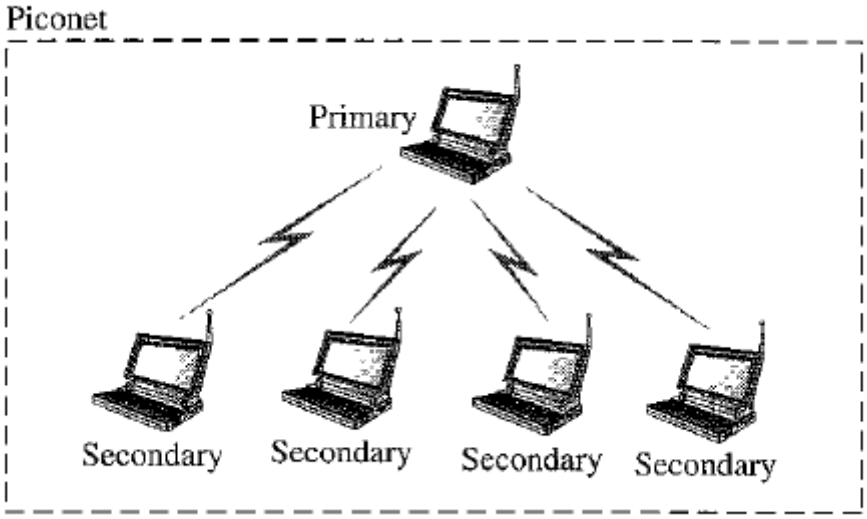
5.3 BLUETOOTH

Bluetooth is a wireless-LAN technology designed to connect devices of different functions such as telephones, notebooks, computers, cameras, printers, coffee makers, and so on. A Bluetooth LAN is an ad hoc network. This means the network is formed spontaneously. The devices find each other and make a network called a piconet (Usually, devices are called gadgets). A Bluetooth LAN can even be connected to the Internet if one of the devices has this capability. By nature, a Bluetooth LAN cannot be large. If there are many devices that try to connect, there is confusion. Bluetooth technology has several applications. Peripheral devices such as a wireless mouse/keyboard can communicate with the computer. In a small health care center, monitoring-devices can communicate with sensor-devices. Home security devices can connect different sensors to the main security controller. Conference attendees can synchronize their laptop computers at a conference. Today, Bluetooth technology is the implementation of a protocol defined by the IEEE 802.15 standard. The standard defines a wireless PAN operable in an area the size of a room or a hall. (PAN personal-area network).

Architecture

Bluetooth defines 2 types of networks: 1) Piconet and 2) Scatternet.

Figure 5.22 : Piconet



Piconets

A Bluetooth network is called a piconet, or a small net. (Figure 5.22). A piconet can have up to 8 stations. A piconet can have only one primary station. The communication between the primary and the secondary can be one-to-one or one-to-many.

i) One of station is called the primary.

ii) The remaining stations are called secondaries.

All the secondary-stations synchronize their clocks and hopping sequence with the primary station. Although a piconet can have a maximum of 7 secondaries, an additional 8 secondaries can be in the parked state. A secondary in a parked state is synchronized with the primary, but cannot take part in

communication until it is moved from the parked state. Because only 8 stations can be active in a piconet, activating a station from the parked state means that an active station must go to the parked state.

Scatternet

Piconets can be combined to form a scatternet (Figure 5.23).

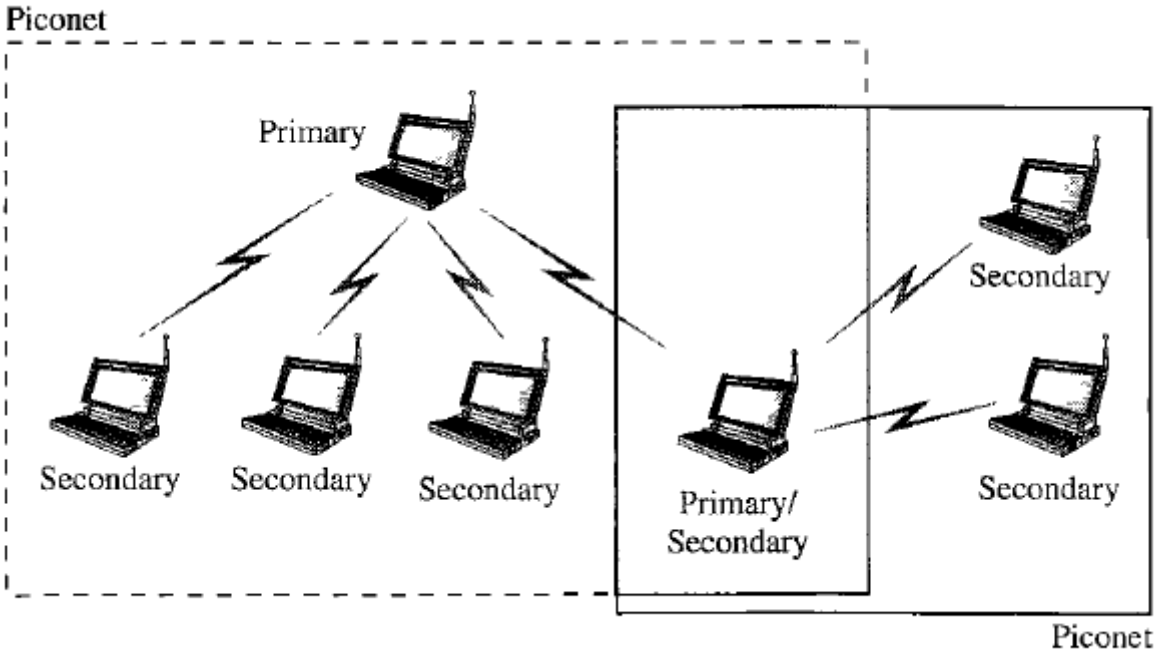


Figure 5.23 : Scatternet

A station can be a member of 2 piconets.A secondary station in one piconet can be the primary in another piconet. This is called mediator station. Acting as a secondary, mediator station can receive messages from the primary in the first piconet. Acting as a primary, mediator station can deliver the message to secondaries in the second piconet.

A Bluetooth device has a built-in short-range radio transmitter. The current data-rate is 1 Mbps with a 2.4-GHz bandwidth. This means that there is a possibility of interference between the IEEE 802.11b wireless-LANs and Bluetooth LANs.

5.4 Cellular Networks

Cellular telephony is designed to provide communications between two moving units called mobile-stations (MSs) or between one mobile-station and one stationary unit called a land unit (Figure 5.24).

A service-provider is responsible for

→ locating & tracking a caller

→ assigning a channel to the call and

→ transferring the channel from base-station to base-station as the caller moves out-of-range.

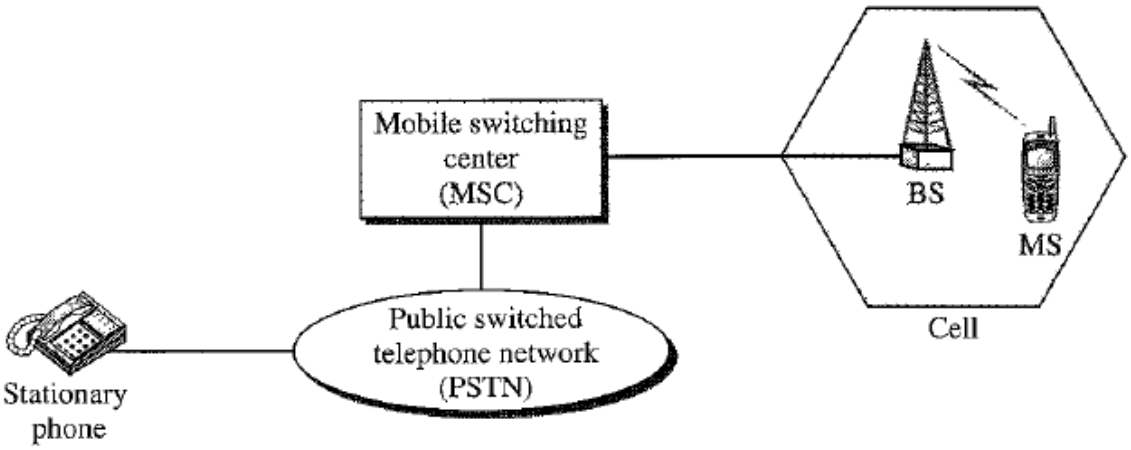


Figure 5.24 : Cellular Network Architecture

Each cellular service-area is divided into small regions called cells. Each cell contains an antenna. Each cell is controlled by AC powered network-station called the base-station (BS). Each base-station is controlled by a switching office called a mobile-switching-center (MSC). MSC coordinates communication between all the base-stations and the telephone central office. MSC is a computerized center that is responsible for

→ connecting calls

→ recording call information and

→ billing.

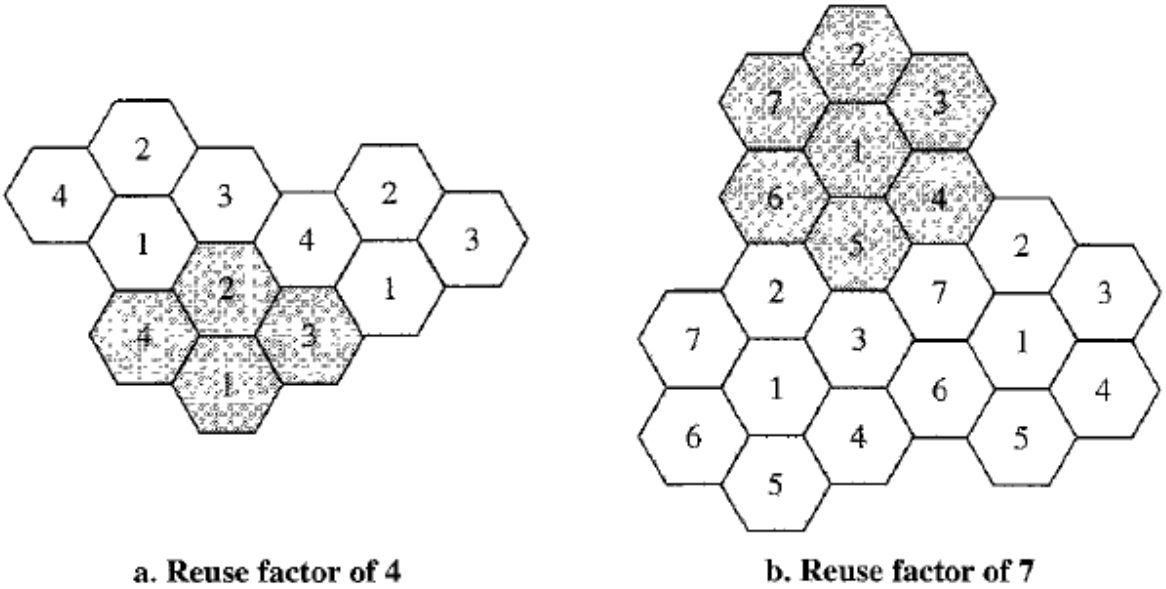
• Cell-size is not fixed; Cell-size can be increased or decreased depending on population of the area.

• Cell-radius = 1 to 12 mi. Compared to low-density areas, high-density areas require many smaller cells to meet traffic demands. Cell-size is optimized to prevent the interference of adjacent cell-signals.

Operation

Frequency Reuse Principle

In general, neighboring-cells cannot use the same set of frequencies for communication. Using same set of frequencies may create interference for the users located near the cell-boundaries. However, set of frequencies available is limited and frequencies need to be reused. A frequency reuse pattern is a configuration of N cells.



a. Reuse factor of 4 b. Reuse factor of 7

Figure 5.25 : Frequency reuse patterns in cellular system.

Where N = reuse factor

Each cell uses a unique set of frequencies.When the pattern is repeated, the frequencies can be reused.There are several different patterns (Figure 5.25). The numbers in the cells define the pattern. The cells with the same number in a pattern can use the same set of frequencies. These cells are called the reusing cells.

Transmitting

Procedure to place a call from a mobile-station:

1) The caller
   → enters a phone number and
   → presses the send button.
2) The mobile-station
   → scans the band to determine setup channel with a strong signal and
   → sends the data (phone number) to the closest base-station.
3) The base-station sends the data to the MSC.

4) The MSC sends the data on to the telephone central office.
5) If called party is available, a connection is made and the result is relayed back to the

MSC.

6) The MSC assigns an unused voice channel to the call, and a connection is established.

7) The mobile-station automatically adjusts its tuning to the new channel.

8) Finally, voice communication can begin.

Receiving

Procedure to receive a call from a mobile-station:

1) When a mobile phone is called, the telephone central office sends phone number to the MSC.

2) MSC searches for the location of the mobile-station by sending query-signals to each cell in a process. This is called paging.

3) When the mobile-station is found, the MSC transmits a ringing signal.

4) When the mobile-station answers, the MSC assigns a voice channel to the call.

5) Finally, voice communication can begin.

Handoff

During a conversation, the mobile-station may move from one cell to another. Problem: When the mobile-station goes to cell-boundary, the signal becomes weak. To solve this problem, the MSC monitors the level of the signal every few seconds. If signal-strength decreases, MSC determines a new cell to accommodate the communication. Then, MSC changes the channel carrying the call (hands signal off from old channel to a new one). Two types of Handoff: 1) Hard Handoff 2) Soft Handoff.

1) Hard Handoff

Early systems used a hard handoff. A mobile-station only communicates with one base-station. When the MS moves from one cell to another cell,

i) Firstly, communication must be broken with the old base-station.

ii) Then, communication can be established with the new base-station.

This may create a rough transition.

2) Soft Handoff

New systems use a soft handoff. A mobile-station can communicate with two base-stations at the same time. When the MS moves from one cell to another cell,

i) Firstly, communication must be broken with the old base-station.

ii) Then, the same communication may continue with the new base-station.

Roaming

Roaming means that the user can have access to communication or can be reached where there is coverage. Usually, a service-provider has limited coverage. Neighboring service-providers can provide extended coverage through a roaming contract.

First Generation (1G)

• The first generation was designed for voice communication using analog signals.

• The main system evolved in the first generation: AMPS (Advanced Mobile Phone System).

AMPS

• This system is a 1G analog cellular system.

• The system uses FDMA to separate channels in a link.

• Here we discuss, two issues:

1) Bands

The system operates in the ISM 800-MHz band.

The system uses 2 separate channels (Figure 16.8):

i) First channel is used for forward communication (base-station to mobile-station)

Band range: 869 to 894 MHz

ii) Second channel is used for reverse communication (mobile-station to base-station).

Band range: 824 to 849 Mhz

2) Transmission

The system uses FM and FSK for modulation.

i) Voice channels are modulated using FM.

ii) Control channels are modulated using FSK to create 30-kHz analog signals.

The system uses FDMA to divide each 25-MHz band into 30-kHz channels.

Second Generation (2G)

The second generation was designed for higher-quality voice communication using digital signals.

1G vs. 2G:

1) The first generation was designed for analog voice communication.

2) The second generation was mainly designed for digital voice communication.

Three major systems evolved in the second generation:

1) D-AMPS (digital AMPS)

2) GSM (Global System for Mobile communication) and

3) IS-95 (Interim Standard).

D-AMPS

• D-AMPS (Digital AMPS) was improved version of analog AMPS.

• D-AMPS was backward-compatible with AMPS.

• Thus, in a cell,

1) First telephone may use AMPS and

2) Second telephone may use D-AMPS.

• Here we discuss, two issues:

1) Band

The system uses the same bands and channels as AMPS.

2) Transmission

Each voice channel is digitized using a very complex PCM and compression technique.

GSM

• Aim of GSM: to replace a number of incompatible 1G technologies.

• Here we discuss, two issues:

1) Bands

2) Transmission

1) Bands

 The system uses two bands for duplex communication.

 Each band is 25 MHz in width.

 Each band is divided into 124 channels of 200 kHz.

2) Transmission

 Each voice channel is digitized and compressed to a 13-kbps digital signal.

 Each slot carries 156.25 bits.

 Eight slots share a frame (TDMA).

 26 frames also share a multiframe (TDMA).

IS-95

• The system is based on CDMA and DSSS.

• Here we discuss, following 6 issues:

1) Bands

2) Transmission

3) Synchronization

4) Two Data-rate Sets

5) Frequency-Reuse Factor

6) Soft Handoff

1) Bands

 The system uses two bands for duplex communication.

 The bands can be ISM 800-MHz band or ISM 1900-MHz band.

 Each band is divided into 20 channels of 1.228 MHz.

 Each service-provider is allotted 10 channels.

 IS-95 can be used in parallel with AMPS.

 Each IS-95 channel is equivalent to 41 AMPS channels (41 × 30 kHz = 1.23 MHz).

2) Transmission

 Two types of Transmission:

i) Forward Transmission (base to mobile)

¤ Communications between the base and all mobiles are synchronized.

¤ The base sends synchronized data to all mobiles.

ii) Reverse Transmission (mobile to base)

¤ The use of CDMA in the forward direction is possible because the pilot channel sends a continuous sequence of 1s to synchronize transmission.

¤ The synchronization is not used in the reverse direction because we need an entity to do that, which is not feasible.

¤ Instead of CDMA, the reverse channels use DSSS.

3) Synchronization

 All base channels need to be synchronized to use CDMA.

 To provide synchronization, bases use the services of a satellite system (GPS).

4) Two Data Rate Sets

 IS-95 defines two data-rate sets:

i) The first set defines 9600, 4800, 2400, and 1200 bps.

ii) The second set defines 14,400, 7200, 3600, and 1800 bps.

5) Frequency Reuse Factor

 The frequency-reuse factor is normally 1 because the interference from neighboring cells cannot affect CDMA or DSSS transmission.

6) Soft Handoff

 Every base-station continuously broadcasts signals using its pilot channel.

 Thus, a mobile-station can detect the pilot signal from its cell and neighboring cells.

 This enables a mobile-station to do a soft handoff.

Third Generation (3G)

• 3G cellular telephony provides both digital data and voice communication.

• For example: Using a Smartphone, A person can talk to anyone else in the world.

A person can download a movie, surf the Internet or play games.

• Interesting characteristics: the Smartphone is always connected; we do not need to dial a number to connect to the Internet. (IMT Internet Mobile Communication)

• Some objectives defined by the blueprint IMT-2000 (3G working group):

1) Voice quality comparable to that of the existing public telephone network.

2) Data-rate of

→ 144 kbps for access in a moving vehicle (car)

→ 384 kbps for access as the user walks (pedestrians) and

→ 2 Mbps for the stationary user (office or home).

3) Support for packet-switched and circuit-switched data services.

4) A band of 2 GHz.

5) Bandwidths of 2 MHz.

6) Interface to the Internet.

Fourth Generation (4G)

• 4G cellular telephony is expected to be a complete evolution in wireless communications.

• Some objectives defined by the 4G working group:

1) A spectrally efficient system.

2) High network capacity.

3) Data-rate of

→ 100 Mbps for access in a moving vehicle

→ 1 Gbps for stationary users and

→ 100 Mbps between any two points in the world.

4) Smooth handoff across heterogeneous networks.

5) Seamless connectivity and global roaming across multiple networks.

6) High quality of service for next generation multimedia support.

7) Interoperability with existing wireless standards.

8) All IP, packet-switched, networks.

4G is only packet-based networks. 4G supports Ipv6. 4G provides better multicast, security, and route optimization capabilities.Here we discuss, following issues:

1) Access Scheme

2) Modulation

4) Antenna

5) Applications

3) Radio System

1) Access Scheme

• To increase efficiency, i) capacity, ii) scalability & iii) new access techniques are being considered for 4G. For example:

i) OFDMA and IFDMA are being considered for the downlink & uplink of the next generation UMTS.

ii) MC-CDMA is proposed for the IEEE 802.20 standard.

2) Modulation

More efficient 64-QAM is being proposed for use with the LTE standards.

3) Radio System

• The 4G uses a SDR system.

• The components of an SDR are pieces of software and thus flexible.

• The SDR can change its program to shift its frequencies to mitigate frequency interference.

Antenna

• The MIMO and MU-MIMO antenna system is proposed for 4G.

• Using this antenna, 4G allows independent streams to be transmitted simultaneously from all the antennas to increase the data-rate.

• MIMO also allows the transmitter and receiver coordinates to move to an open frequency when interference occurs.

5) Applications

• At the present rates of 15-30 Mbps, 4G is capable of providing users with streaming high-definition television.

• At 100 Mbps, the content of a DVD-5 can be downloaded within about 5 minutes for offline access.

(OFDMA Orthogonal FDMA

IFDMA interleaved FDMA)

(LTE Long Term Evolution

SDR Software Defined Radio)

(MIMO multiple-input multiple-output MU-MIMO multiuser MIMO)

(UMTS Universal Mobile Telecommunications System)

(MC-CDMA multicarrier code division multiple access)