UNIT 1 BASICS OF DATA COMMUNICATION

Structure **Page No.**

1.0 Introduction 5
1.1 Objectives 5
1.2 Concept of Communication System 6
1.3 Analog and Digital Communication 8
1.4 Data Communication Modes 10
1.4.1 Synchronous and Asynchronous Transmission
1.4.2 Simplex, Half-Duplex, Full Duplex Communication
1.5 Networking Protocols and Standards 13
1.5.1 Layering
1.5.2 OSI Reference Model
1.5.3 Encapsulation
1.5.4 End-to-End Argument
1.5.5 Protocol Design Issues
1.6 Applications of Computer Networking 20
1.7 Summary 21
1.8 References/Further Reading 21
1.9 Solutions/Answers 21

1.0 INTRODUCTION

This is the first unit of our course on Fundamentals of Computer Networks. It will introduce you to some of the basic concepts of data communication and computer networking. In other words, though this unit we would like to explain the "What, Why, When, How, Where" of data communication. In the beginning, you will be introduced with the concept of communication and "communication system". Once you understand the communication system and its components, we think other areas will be simpler for you. Different forms of data communication are further introduced to you in this unit. Next, we will discuss about various modes of communications. This unit will eventually cover an introduction to computer networking, networking protocols and standards, those are necessary for any effective communication. In the end of our unit, we will discuss various applications of data communication and computer networking.

1.1 OBJECTIVES

After going through this unit, you should be able to:

- Know the concept of communication system
- Understand the communication system and its components
- Differentiate between analog and digital Communication
- Know data communication modes
- Differentiate between synchronous and asynchronous transmission
- Differentiate among Simplex, half-duplex, full duplex communication
- Understand the need of protocols and standards
- Know the functions of OSI layers
- Understand the concepts of encapsulation and End-to-end argument
- Know the different protocol design issues

- Know the applications of computer network

1.2 CONCEPT OF COMMUNICATION SYSTEM

Before we discuss about "communication system" and its components, let us understand "communication". Can you define it, what definition will come first in your mind? When we asked some students, answers were like:

- Delivery of message
- Proper way of passing a signal to the intended user
- Right message, to right person, at right time through right way.

So many "rights"! and all seems to be 'Right'. Let me inform you about some definitions of communication:

- " Communication is transfer of information from one person to another, whether or not it elicits confidence. But the information transferred must be understandable to the receiver – G.G. Brown.
- The imparting or exchanging of information by speaking, writing, or using some other medium.-Oxford Dictionary

After going though these definitions, I am sure now we can list the components required for some communication:

- **Sender**: who is trying to send a message to the receiver?
- **Message or Signal**: the message is the actual content for communication
- **Communication Medium**: The medium is what the message is transmitted on. The phone system, it is wire. Television and radio use air
- **Receiver**: The receiver is the target of the message.

There is something missing in this list, can you guess? That is encoding and decoding. Try to conceptualize a discussion with your friend. While talking with your friend you encode your message in a speaking language and on the other side your friends (receiver) decodes your language and understand the message. In the same way, if you are talking to your friend over telephone, it is not possible to actually transmit voice across the wire for any distance. The telephone set converts the sound into electrical pulses, which can be transmitted by wires. The decoder takes the encoded message and converts it to a form the receiver understands, in continuation of our previous example phone system convert electrical pulses into voice.

Let's see a block diagram for a communication system as depicted in the Figure 1.

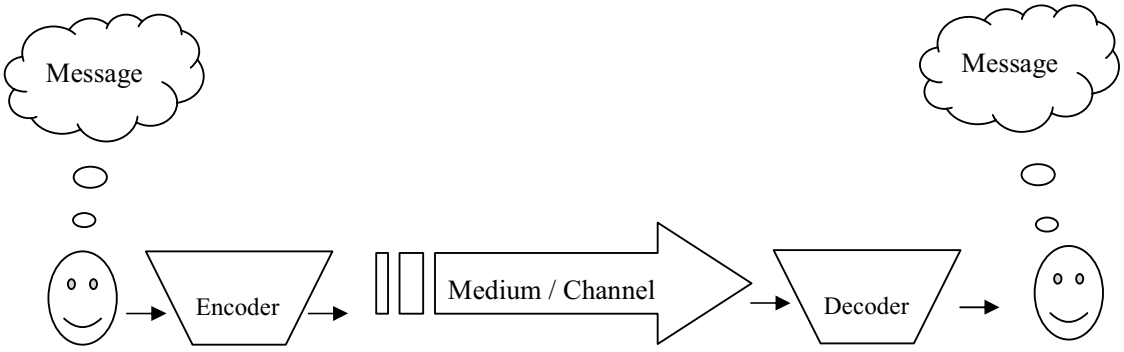


Figure 1: Block Diagram of Communication System

Now, can you try to explore some communication systems around us? Ok, let's list some:

- Human communications operate through speech, signs, gestures, body language, etc. Note that the communication mediums are air and light.

- Telephone system are a kind of communications system which we use in day to day basis, it is a collection of individuals, telephone handsets for transforming voice into electrical signals), wires (communication medium), some controlling and call management devices, Telephone exchange, etc. Remember that the components of a communications system serve a common purpose, are technically compatible, use common procedures, respond to controls, and operate in unity.

- A radio or television communication system is composed of several communications subsystems that give exterior communications capabilities. These are also known as public broadcasting systems, because they broadcast the messages/signals in the air and any one in the coverage area with a receiver can receive the signals. Such systems comprises of a large transmitting station for converting and transmitting the audio/video into the air, and on the other side if signals are public can be decoded and converted again into the same audio/video.

By now, you must be curious to know the mechanism used for converting a message (text, video, audio, etc) into some electrical signals. How does it function in a system?

Let us understand the communication from the technical or mathematical point of view, C.E. Shannon [Claude Elwood Shannon (April 30, 1916 – February 24, 2001) was an American mathematician, electronic engineer, and cryptographer known as "the father of information theory".] had worked on some of the fundamentals in the communication, like:

- How the symbols of communication are transmitted between sender and receiver?

- How the meaning is conveyed through the transmitted symbols?

- What is the effect of the received meaning?

According to Shannon, following are the essential elements of communication also shown in the Figure 2 below:

1. Information source: Source that produces a message

2. Transmitter: An element that functions on the message to generate a signal which can be delivered through a medium/channel

3. Communication Channel: that is a medium over which the signal (carrying the information that composes the message) is sent.

4. Receiver: An element that intercept the signal and converts it back into the message

5. Destination: It can be a person/machine, for whom / which the message is intended.

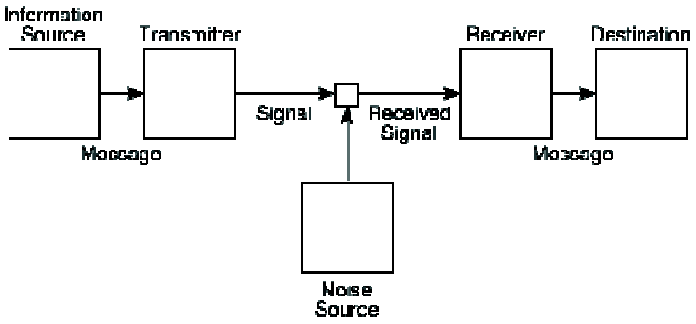


Figure 2: Shannon's diagram of a general communications system

Here, the noise is considered as an error or undesired disturbance that occurs during the transmission (before receiver and after transmitter), from natural and sometimes man-made sources.

1.3 ANALOG AND DIGITAL COMMUNICATION

We all have heard these terms several times, like analog signal, digital signal, digital TV, analog radios, etc. In this section, we will explore the basic definitions and differences of analog and digital communication. Two main types of signals encountered in practice are analog and digital. The figure 3 shows analog, discreet and digital signals, digital signals outcome from approximating an analog signal by its values at particular time moments. Digital signals are discrete and quantized, while analog signals possess neither property.

Technically, if we observe the elements and processes of any communication system, you may notice that all the components and processes of a communication system should be aligned, compatible and work as a unit. Try to remember our example where telephone system is converting voice into electrical signal, in this case receiver instrument must be compatible and convert the electrical signal or voice signal in the similar way, otherwise your message will never be delivered.

Here, we have to address one important point that is how does the message is being converted? We have two options one is analog and another is digital.

- As you may know that an analog is something continuous, which is having time varying feature (variable). Analog signal is a representation of some time varying quantity. For example, Human voice can be considered as an analog signal. In analog signals data are represented by continuously variable, measurable, physical quantities, such as current, voltage, or pressure.
- A digital signal is a physical signal that is a representation of a sequence of discrete values, for example of a bit stream. In digital technology, generally a signal is converted into a bit form represented by a series of "1"s and "0"s. Please note here that "1"s and "0"s are nothing but two states usually represented by some measurement of an electrical property: Voltage is the most common, but current is used in some logic families. A threshold is designed for each logic family.

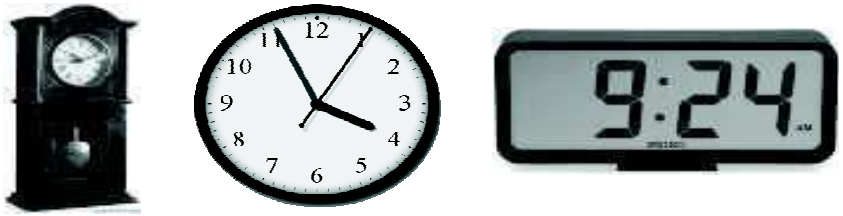


Figure 3: Analog, Discrete and Digital Clocks

In a communication system, data signals are propagated from one point to another by means of electrical signals. An analog signal (Figure 4a) is a continuously varying voltage signal that may be propagated over a variety of media. A digital signal (Figure 4b) is a sequence of discrete values for example any bit stream.

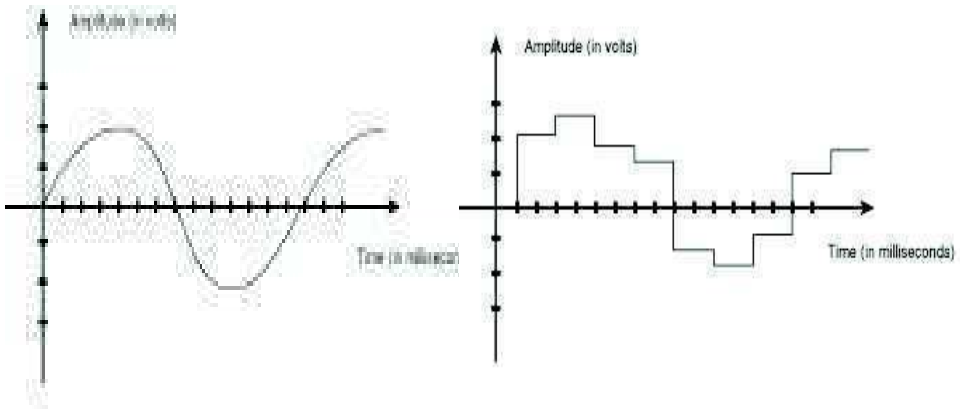


Figure 4 a): Analog Signal **Figure 4 b): Digital Signal**

Analog and **digital** signals are used to transmit information, usually through electric signals. In both these systems, the information, such as any text, audio or video, is transformed into electric signals. Let us see some of the differences between analog and digital systems below in table 1.

Table 1: Comparison between Analog and Digital system

Analog	Digital
Signals are records waveforms as they are. Signal occupies the same order of spectrum as the analog data.	Converts analog waveforms into set of numbers and records them. The numbers are converted into voltage stream for representation. In case of binary it is converted in 1's and 0's.
In analog systems electronic circuits are used for transformation of signals.	In this transformation is done using logic circuits.
About Noise analog signals are more likely to get affected and results in reducing accuracy	Digital signals are less affected, because noise response are analog in nature
Analog signal is a continuous signal which transmits information as a response to changes in physical phenomenon.	Digital signals are discrete time signals generated by digital modulation.
Data transmission is not of high quality	Data transmission has high quality.
Analog devices are not very precise.	Digital systems are very precise.

Can you explore the reasons why digital signals are seems to be better? Let us see why digital communication having high quality? Because, digital devices decode and reconstruct data, due to which loss of quality of data as compared to analog devices is much higher. But analog signal are affected by noise. While amplifying the signal noise also gets amplified. Therefore it becomes difficult to filter out noise from the signal and the message gets corrupted. Digital signal are least affected by noise. And further

computer advancement has enabled use of error detection and error correction techniques to remove disturbances artificially from digital signals and improve quality. Now days, digital signals has been most proficient in cellular phone industry. Analog phones have become superfluous even though sound clarity and quality was better.

☞ Check Your Progress 1

1. List the essential elements of communication system. Also, draw and explain the Shannon model of communication system.

   ……

   ……

   ……

   ……

2. Write any four differences between analog and digital communication.

   ……

   ……

   ……

   ……

1.4 DATA COMMUNICATION MODES

In this section, we will learn about some modes of data communication used in computer networking. Because we are going to study computer networking, we assume all data communication is digital. Digital communications is the physical transfer of data/bits over a communication channel. As you may know, data are represented as an electromagnetic signal, such as an electrical voltage, radio-wave, microwave, or infrared signal. The channel or medium could be air (for wireless/mobile communication), copper wires, or optical-fibers. Remember, the data transmitted can be pure digital messages generated from a digital-data source, like a computer or a keyboard. However, it may also be an analog signal such as a human voice over phone call, which could be digitalized.

Serial and parallel transmission

In digital communication, serial transmission of data refers to sequential transmission of bits, where a group of bits over a single channel represents a character. It requires less processing and fewer chances for error. The start and stop of a communication is specified by LSB (lowest significant bit) and MSB (most significant bit) as shown in the Figure 5.

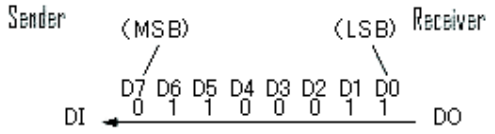


Figure 5: Serial Communication

Parallel transmission refers to simultaneous transmission of the bits over two or more separate channels. Here, we can transmit multiple bits simultaneously as given in Figure 6, which allows for higher data transfer rates than that can be achieved with serial

transmission. For example, for internal data communication in a computer system this method of parallel transmission is used. Parallel data transmission is less reliable for long distances because error correction is not very simple and economical in this case.

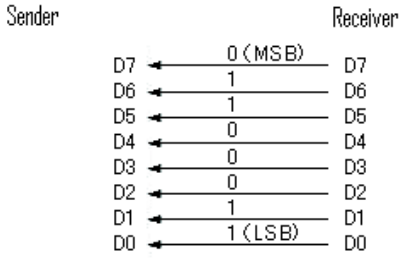


Figure 6: Parallel data transmission

In serial transmission, the byte plus the parity bit are transmitted one bit after another in a continuous line. In parallel transmission, 8 bits (a byte) plus a parity bit are transmitted at the same time over nine separate paths. Thus, parallel transmission is generally faster than serial transmission.

1.4.1 Synchronous and Asynchronous Transmission

Synchronous transmission means both receiver and sender has an agreement (or aware) about timing for the sending data, so that both sender and receiver can coordinate (synchronize) their data signals. Asynchronous means "not synchronous", or no coordination between sender and receiver before transmission. Can you try to explore some examples of Synchronous and asynchronous communication that occurs in your day to day life?

The asynchronous transmission uses start and stop bits to signify the beginning bit. For example, if sender wants to send some data "**11100001**", it will be appended with the start and stop bit and look like "1 **11100001** 0". Where, we have assumed that '0' is start bit and '1' is stop bit. Asynchronous transmission works well where the characters are transferred at irregular intervals e.g. data entry from the keyboard.

Asynchronous transmission has some advantages and disadvantages, like:

- Each individual character is complete unit, hence if there is an error in a character, other sequence of characters are not affected. However, Error in start and stop bit(s) may cause serious problems in data transfer.
- Doesn't require synchronization of both communication sides.
- It is cost effective
- The speed of transmission is limited.
- Large relative overheads, a high proportion of the transmitted bits are uniquely used for control purposes

In case of synchronous transmission, we do not use any start and stop bits, but instead of that clock signal end (clock is built into each end of transmission) is being used for synchronizing the data transmission at both the receiving and sending. A constant stream of bits is sent between the sender and receiver. As clock synchronization may disturbed the possibility of error increases in synchronous transmission. Synchronous transmission has following advantages and disadvantages:

- In comparison to asynchronous communication it has higher speeds, because the system has lesser possibility of error. But, if an error takes place, the complete set of data is lost instead of a single character.
- Serial synchronous transmission is principally used for high-speed communication between computers but is unsuitable where the characters are transferred at irregular intervals.
- It is gives lower overheads and thus, greater throughput.
- Process is more complex
- It is not very cost effective as hardware are more expensive

1.4.2 Simplex, Half-Duplex, Full Duplex Communication

The data transmission mode on the channel, can be classified into three ways simplex, half-duplex and full-duplex as given below in Figure 7.

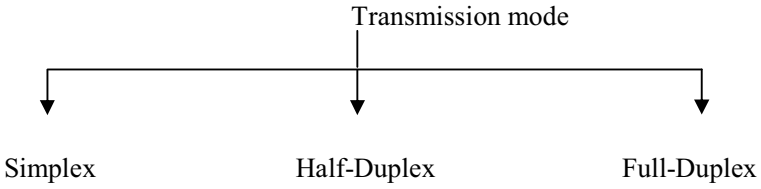


Figure 7: Transmission mode

In simplex transmission *(*Figure 8a), signals are transmitted in only one direction; one station is a transmitter and the other is the receiver. In the half-duplex operation (Figure 8b), both stations may transmit, but only one at a time. In full-duplex operation (Figure 8c), both stations may transmit simultaneously. In the latter case, the medium is carries signals in both directions at same time.

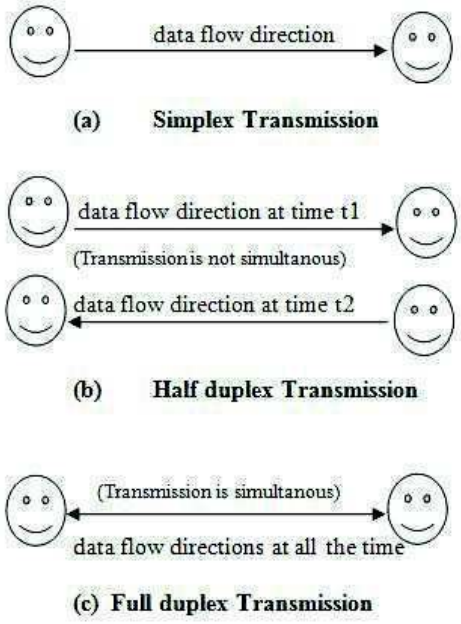


Figure 8(a) (b) (c): Directions of data transmission in Simplex 8(a), half-duplex 8(b), full duplex 8(c) communication

Simplex Transmission

Simplex transmission is one-way transmission. As the name implies, is simple in term of process and hardware. It is also called unidirectional because the signal travels in **only**

one direction. For example, Radio or TV broadcasting system, which are always in one direction from Radio/TV station to our radio or TV sets.

Half-Duplex Transmission

In half-duplex transmission data transmission can be take place in both directions, but not at the same time. This means that only one side can transmit at a time. For example, walky-talky devices used by security agencies are half-duplex as only one person can talk at one time.

Full-Duplex Transmission

Full-duplex (also known as Duplex) transmission can take place in both directions at the same time. For example, telephone or mobile conversation is an example of full-duplex communication, where both sender and receiver can hear each other at the same time.

☞ **Check Your Progress 2**

1. Differentiate between Synchronous and asynchronous transmission.

   ……

   ……

2. Give an example of each communication system based on:
   - Simplex communication,
   - half-duplex communication,
   - full duplex communication

   ……

   ……

   ……

1.5 NETWORKING PROTOCOLS AND STANDARDS

All modes of communication described above follow some 'set of rules' or protocol. Protocol is set of rules that governs communication between the entities engaged in conversation, for example in railways, if a green colour flag is shown a train can start, and if red colour flag is shown train will stop, this is a set of rule and we can say it is a protocol. When we write a letter or talk to someone we follow protocol(s). In case of computer communication also both sender computer and receiver computer should agree on some set of rules like communication language/syntax, scheme of acknowledgement, rules for data control, error control, and other mechanism. Thus, we can say that the conversation is governed by some set of rules known to both the parties. This set of rules is called protocol and it necessary for proper and disciplined conversation/communication.

Problems in Computer Communication

When protocols are implemented for computer communication, we encounter some challenges due to the infrastructure and machines used in computer network may not be compatible and aligned with one another. The concept of Internetworking though, highly desirable, is not easily achievable. Let us see one simple example to understand the compatibility problem, any two networks, cannot directly communicate by connecting a wire between the networks. For example, one network could represent a binary 0 by-5 volts, another by +5 volts. Similarly, one could use a packet size of 128 bytes, whereas other could use 256 byte packets. The method of acknowledgement or error detection

could be different. There could be many such differences. The incompatibility issues are handled at two levels:

i) **Hardware Issues**

At the hardware level, an additional component called router is used to connect physically distinct networks. A router connects to the network in the same way as any other computer. Any computer connected to the network has a Network Interface Card (NIC), which has the address (network id+ host id), hard coded into it. A router is a device with more than one NICs. Router can connect incompatible networks as it has the necessary hardware (NIC) and protocols.

ii) **Software Issues**

The routers must agree about the way information would be transmitted to the destination computer on a different network, since the information is likely to travel through different routers, there must be a predefined standard to which routers must confirm. Packet formats and addressing mechanism used by the networks may differ. One approach could be to perform conversion and reconversion corresponding to different networks. But this approach is difficult and cumbersome. Therefore, the Internet communication follows one protocol suite, the TCP/IP. The basic idea is that it defines a packet size, routing algorithms, error control, flow control methods universally.

It would be unwise to club all these features in a single piece of software ─ it would make it very bulky. Therefore, all these features are logically sub-grouped and then the sub-groups are further grouped into groups called layers. Each layer has an interface with the adjacent layers, and performs specific functions.

1.5.1 Layering

Since it is difficult to deal with complex set of rules, and functions required for computer networking, these rules and functions are divided into logical groups called layers. Each layer can be implemented interdependently with an interface to other layers providing with services to it or taking its services like data, connection and error control functions are grouped together and make a layer. A. Speech in telephone conversation is translated, with electrical segments and vice-versa. Similarly in computer system the data or pattern are converted into signals before transmitting and receiving. These function and rules are grouped together and form a layer.

1.5.2 OSI Reference Model

The OSI model is based on a proposal developed by the International Standards Organization as a first step towards international standardization of the various communication functions and services. Here, communication functions are grouped into logical layers. The model is called the ISO - OSI (International Standard Organisation - Open Systems Interconnection) Reference Model because it deals with connecting open systems — that is, systems that follow the standard are open for communication with other systems, irrespective of a manufacturer. Its main objectives were to allow manufacturers of different systems to interconnect equipment through standard interfaces globally. Allow software and hardware to integrate well and be portable on different systems. The OSI model has seven layers shown in Figure 9. The principles

that were applied to arrive at the seven layers are as follows:

1. Each layer should perform a well-defined function.

2. The function of each layer should be chosen with an eye toward defining internationally standardized protocols.

3. The layer boundaries should be chosen to minimize the information flow across the interfaces.

4. A layer serves the layer above it and is served by the layer below it.

The set of rules for communication between entities in a layer is called protocol for that layer. The seven layers of ISO OSI reference model as shown in the Figure 9 are following:

i) Physical Layer

ii) Data Link Layer

iii) Network Layer

iv) Transport Layer

v) Session Layer

vi) Presentation Layer

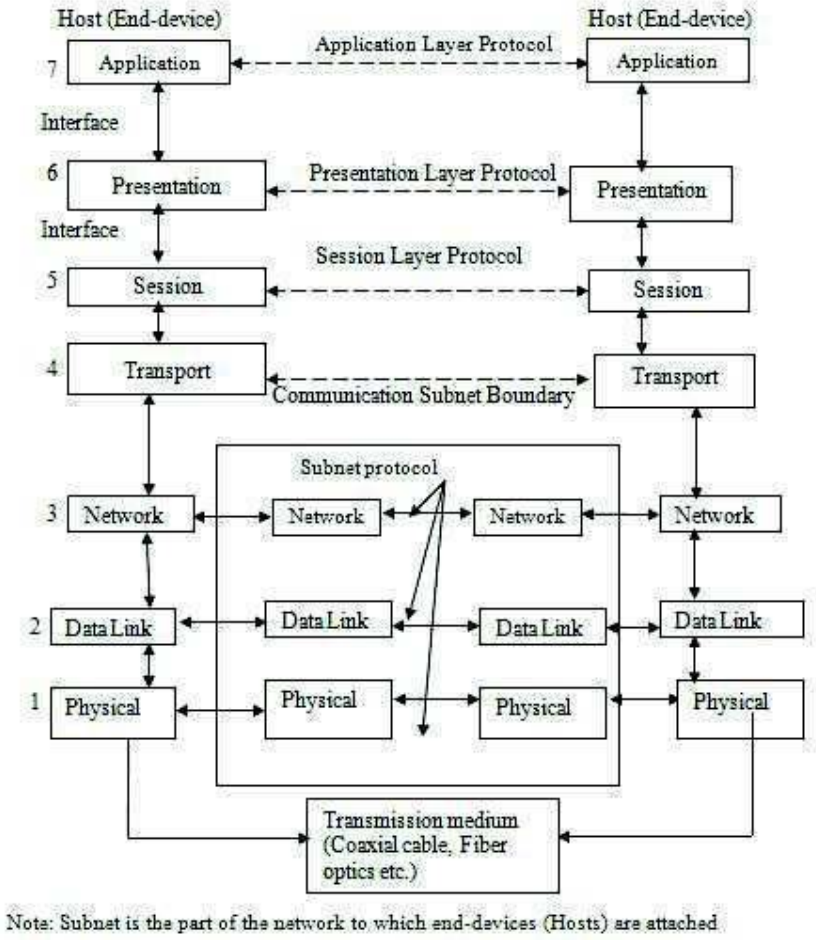


Figure 9: OSI Reference Model

a) **The Physical Layer**

Physical Layer defines functional, electrical and mechanical specifications of signaling, cables, and connectors options that physically link two nodes on a network.

b) **The Data Link Layer**

The main task of data link layer is to provide error free transmission. It accomplishes this task by having the sender configure input data into data frames, transmit the frames sequentially, between network devices and process the acknowledgement frames sent back by the intermediate receiver. The data link layer creates and recognises frame boundaries. This can be accomplished by attaching special bit patterns to the beginning and end of the frame. Since these bit patterns can accidentally occur in the data, special care is taken to make sure these patterns are not incorrectly interpreted as frame boundaries.

c) **The Network Layer**

The network layer ensures that each packet travels from its sources to destination (both in different networks) successfully and efficiently. 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. Finally, they can be highly dynamic, being determined anew for each packet, to reflect the current network load. When a packet has to travel from one network to another to reach its destination, many problems can arise. The addressing mechanism is used by the second network may be different from the first one. The second network 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.

d) **The Transport Layer**

The basic function of the transport layer is to accept data from the session layer, 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.

Transport Layer provides location and media independent end-to-end data transfer service to session layer.

e) **The Session Layer**

The main tasks of the session layer are to provide:

- Session Establishment
- Session Release – Orderly or abort
- Synchronization
- Data Exchange
- Expedited Data Exchange.

The session layer allows users on different machines to establish sessions between them. A session allows ordinary data transport, as does the transport layer, but it also provides enhanced services useful in some applications. A session might be used to allow a user to log into a remote timesharing system or to transfer a file between two machines.
One of the services of the session layer is to manage dialogue control. Sessions can allow traffic to go in both directions at the same time, or in only one direction at a time. If traffic can only go one way at a time (analogous to a single railroad track), the session layer can help in keeping track of whose turn it is.

A related session service is token management. For some protocols, it is essential that both sides do not attempt the same operation at the same time. To manage these activities, the session layer provides tokens that can be exchanged. Only the side holding the token may perform the desired operation.

Another session service is synchronization. Consider the problem that might occur when trying to do a 2 hour file transfer between two machines with a one hour mean time between crashes. After each transfer was aborted, the whole transfer would have to start over again and would probably fail again the next time as well. To eliminate this problem, the session layer provides a way to insert markers after the appropriate checkpoints.

f) **The Presentation Layer**

Unlike all the lower layers, which are just interested in moving bits reliably from here to there, the presentation layer is concerned with the syntax and semantics of the information transmitted.

A typical example of a presentation service is encoding data in a standard agreed upon format. Most user programs do not exchange random binary bit strings, they exchange things such as people's names, dates, amounts of money and invoices. These items are represented as character strings, integers, floating-point number, and data structures composed of several simpler items. Different computers have different codes for representing character strings (e.g., ASCII and Unicode), integers (e.g., one's complement and two's complement), and so on. In order to make it possible for computers with different representations to communicate, the data structure to be exchanged can be defined in an abstract way, along with a standard encoding to be used. The presentation layer manages these abstract data structure and converts from the representation used inside the computer to the network standard representation and back. It also performs the task of encryption and decryption.

g) **Application Layer**

Application Layer supports functions that control and supervise OSI application processes such as start/maintain/stop application, allocate/deallocate OSI resources, accounting, check point and recovering. It also supports remote job execution, file transfer protocol, message transfer and virtual terminal.

1.5.3 Encapsulation

Encapsulation is a technique of implementing layered architecture of a communication system. In OSI model we have separated all the communication functions/services into seven layers. We know that a layer serves the layer above it and is served by the layer below it, so to make it possible encapsulation techniques is followed for sending/receiving data between and through layers. In encapsulation we add some control information or "Header/Trailer" to a Data Unit by a communications protocol. This data along with header/trailer is known as Protocol Data Unit (PDU). This header/trailer actually creates an envelope for the PDU which has its address and addressee.

The Figure 10 shows the header associated with each of the N layers of some communication model. When a packet of data (we are saying it as PDU because packets is relevant to some protocol at some layer) is passed by any layer, attach a header (control information) of its layer and passes the packet (along with header) to the layer below. Each layer appends a new header to the PDU received from upper layer. Each layer considers the PDU of upper layer as data, and does not worry about headers in the PDU. This process continues until the packet reaches the lowest layer, which is the communication channel.

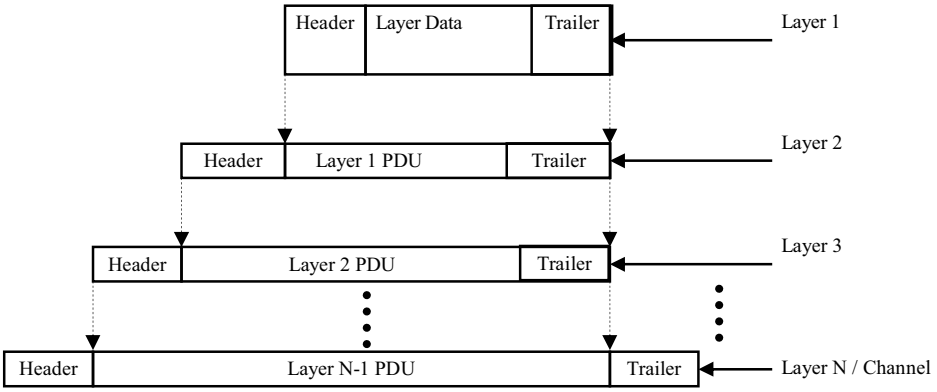


Figure 10: Encapsulation

This lowest layer is considered as physical layer by OSI model, which also add its control information. As you know that physical layer is final end or beginning of any side of communication. It converts PDU+control information into a series of bits and sends it across a cable or telecommunications circuit to the destination. Generally, due to this physical layer add control information at both ends of PDU for management point of view. At the receiver side, the Layer N of receiver reassembles the series of bits to form a packet and forwards the packet for processing by the upper (N-1) layer. This removes the N-1 layer header, and passes it to the next upper layer. The processing continues until finally the original packet data is sent to the program/application running at Layer 1.

1.5.4 End-To-End Argument

Assume we want to transfer some important file or information from a machine available on a network to a machine on other network. In case of a reliable communication we will establish a connection. If connection is available we send the data. Before accepting the data we will ensure the reliability of data at each step or each layer. But at the final stage at receiver (application layer) reliability check have to be performed. If we have to perform a final reliability check at application layer, can we say that we do not require reliability checks at lower layers? Is there any need to implement reliability at lower layers? Yes, it can be implemented but only for improving the performance in case the link quality is poor.

The end-to-end principle states that application-specific functions must be implemented in the end hosts of a network instead of intermediary nodes, provided these functions are “completely and correctly" implemented at the end hosts. The basic concept behind the end-to-end principle is that for two processes communicating with each other via some communication channel, the reliability obtained from that means cannot be expected to be absolutely associated with the reliability requirements of the processes. To be specific we can say, obtaining a very high ‘reliability’ requirements of communicating processes in a small network is more costly than obtaining that ‘reliability’ by end-to-end acknowledgements and retransmissions.

A system should consider only functions that can be completely and correctly implemented within it. We needs to be careful before implementing a functionality that we believe that is useful to an application at a lower layer. If the application can implement a functionality correctly, implement it a lower layer only as a performance enhancement. If implementation of function in higher levels is not possible due to technological/economic reasons then it may be placed at lower levels.

1.5.5 Protocol Design Issues

For communication to take place, protocols have to be agreed upon. Data are sent and received on communicating systems to establish communications. Protocols should therefore specify rules governing the transmission. In general, the following issues should be addressed for designing these protocols:

- Data formats: The format of data should be well defined, how the bit strings are divided in fields and in which format. Here, the packet size and format, PDU format, header size and format should be defines properly for proper communication. Let us assume a postal system, in which we specify, where the address of sender/ receiver should be written. Different kind of letters are represented by different methods like speed post, telegraph, registered post, book post, post card, etc.
- Address formats: Addresses are used to recognize both the sender and the proposed receiver. Mostly, addresses (also a bit string) are stuffed in the header field of the packet, to find whether the packet/data are intended for someone or not. The rules explaining the purpose of the address value are called an addressing scheme. For example, in the postal system, the method and sequence of writing an address is well formulated like name, father's name, house number, street, city, country, pin code, etc.
- Address mapping. Sometimes protocols need to map addresses of one scheme on addresses of another scheme. When the address formats are different than mapping is needed. For example, physical address of a computer need to be mapped with network address of a computer.
- Routing. When systems are not directly connected, intermediary systems along the route to the intended receiver(s) need to forward messages on behalf of the sender. In the postal system, we can see the post offices are selecting and sorting the letter according to the given addresses.
- Acknowledgements Scheme: In connection-oriented communication (communication systems where connection is not established before communication like email or SMS), acknowledgement of correct reception of packets is required. Acknowledgements are sent from receivers back to their respective senders, in the same way of registered posts. connection-oriented communication ensure the reliability by acknowledgement.
- Data Loss and damage: There is a possibility that data is lost or get corrupted (changed from 0 to 1 or vice versa). To address the data loss, protocols may implement acknowledgement scheme. Protocols may use timeout mechanism, in which if data is not received within a time frame sender is requested to retransmit the data. If data is corrupted, different error correction and detection mechanisms can be used.
- Sequence control: In this we wants to ensure that the packets (chunk of bits) are received in a correct sequence or not. The packets are sent on the network individually, so some packets may get lost or delayed or take different routes to their destination on some types of networks. As a result pieces may arrive out of sequence. necessary scheme should be implemented for retransmissions and reassemble the packets in right order to get the original message.
- Flow control is needed when the sender transmits faster than the receiver can process the transmissions. Flow control can be implemented by various schemes, which you will study further in the course.

1.6 APPLICATIONS OF COMPUTER NETWORKING

The main reason is that each computer network is designed with a specific purpose. Due to advancement in Computer Networks field we are now moving from personalized

computing to network computing. Therefore, its application is increasing every day. For example, a computer network in an office is used to connect computers in a smaller area, and it provides fast communication between the office persons/machines. The following is the list of some general application of computer network:

Resource sharing

Using networks we can share any resource, CPU processing power, peripherals like printers, scanners, etc, information like files and data and even software. This sharing is done by communicating the machine through whom we want to share.

Personal communication

There are many examples available with us for personal communication through computer networks, like email, chatting, audio/video conferencing, etc

Information Broadcasting and Search

This is also a mostly used application like website, blogs, social networking website, search engines, etc. Computer network provide us tremendous opportunity for information broadcasting, display, searching and information retrieval. Apart from these commonly used applications of computer networks we have following specific applications of computer networking.

Some Specific end applications

- Campus-wide computing and resources sharing
- Collaborative research and development
- Integrated system for design + manufacturing + inventory
- Electronic commerce, publishing and digital libraries
- Multimedia communication (tele-training, etc.)
- Health-care delivery (remote diagnosis, telemedicine)
- Video-on-demand.
- On-line learning.

☞ Check Your Progress 3

1. Explain the need of layering in the data communication protocols stack.

……

……

……

……

2. List and explain any two functions of each OSI layer.

……

……

……

……

1.7 SUMMARY

We hope you must have understood the concept of communication and communication system. As we discussed communication system is comprised of Information Source, Transmitter, Communication Channel, Receiver and the Destination. The information could be sent in the various forms like analog and digital. The concept of analog and digital transmission deals with form in which information is available and the way it is transmitted. Analog data is represented by continuous signals. The other type of signal is digital, which uses a bit steam. In this unit we have studied various modes and mechanism of communication like synchronous and asynchronous communication, simplex, half and full duplex communication. We studied that in simplex the data/signals are transmitted in one direction by a station i.e., by the sender, in half duplex the transmission can be done in one direction at a time whereas in full duplex the transmission can take place in both directions simultaneously. Further, in the unit we have explored the computer networking systems, its difficulties in data communication and the need of protocols and standards for these systems. We have also studied the details of OSI reference model and functions of OSI layers. In the end of this unit we have discussed different protocol design issues and listed some of the applications of computer networks. In the next unit, you will be introduced with various modulation techniques and there advantages. These modulation techniques are used to convert the message signal into a different form(s) so that it can be communicated through computer networks.

1.8 REFERENCES/FURTHER READING

1. *Computer Networks*, A. S. *Tanenbaum* 4th Edition, Practice Hall of India, New Delhi. 2003.
2. *Introduction to Data Communication & Networking,* 3rd Edition, Behrouz Forouzan, Tata McGraw Hill.
3. *Computer Networking*, J.F. Kurose & K.W. Ross, A Top Down Approach Featuring the Internet, Pearson Edition, 2003.
4. *Communications Networks*, Leon Garcia, and Widjaja, Tata McGraw Hill, 2000.
5. *Data and Computer Communications,* Willian Stallings, 6th Edition, Pearson Education, New Delhi.
6. www.wikipedia.org
7. Larry L. Peterson, *Computer Networks*: A Systems Approach, 3rd Edition (The Morgan Kaufmann Series in Networking).

1.9 SOLUTIONS/ANSWERS

☞ **Check Your Progress 1**

1. Following are the essential elements of communication system.
   a) Information source: Source that produces a message
   b) Transmitter: An element that functions on the message to generate a signal which can be delivered through a medium/channel
   c) Communication Channel: that is the medium over which the signal (carrying the information that composes the message) is sent.
   d) Receiver: An element converts the signal back into the intended message.

e) Destination: It can be a person/machine, for whom / which the message is intended.

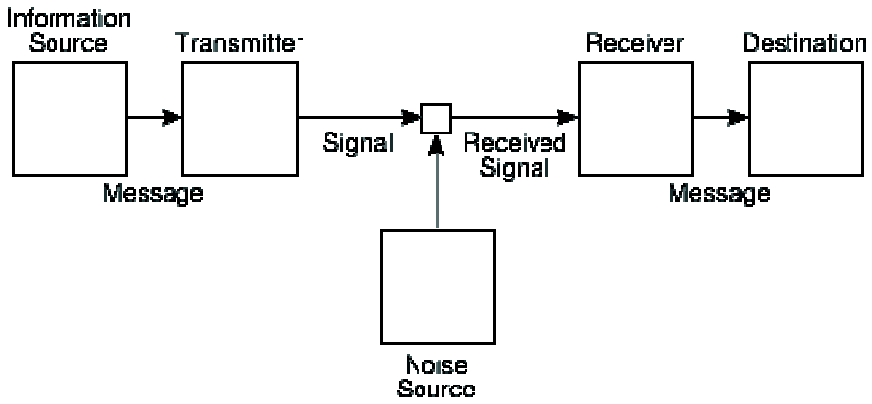


Diagram: Shannon's diagram of a general communications system

Here, the noise is considered as an error or undesired disturbance occurs during the transmission (before receiver and after transmitter), from natural and sometimes man-made sources.

2. Following are some differences between analog and digital communication:

Analog	Digital
Signals are records waveforms as they are. Signal occupies the same order of spectrum as the analog data.	Converts analog waveforms into set of numbers and records them. The numbers are converted into voltage stream for representation. In case of binary it is converted in 1's and 0's.
In analog systems electronic circuits are used for transformation of signals.	In this transformation is done using logic circuits.
About Noise analog signals are more likely to get affected and results in reducing accuracy	Digital signals are less affected, because noise response are analog in nature
Data transmission is not of high quality	Data transmission has high quality.

☞ Check Your Progress 2

1. Following are the main differences between Synchronous and asynchronous transmission.

Asynchronous transmission has following advantages and disadvantages:

- Each individual character is complete unit, hence if there is an error in a character, other sequence of characters are not affected. However, Error in start and stop bit(s) may cause serious problems in data transfer.
- Doesn't require synchronization of both communication sides.
- It is cost effective
- The speed of transmission is limited.
- Large relative overhead, a high proportion of the transmitted bits are uniquely for control purposes

Synchronous transmission has following advantages and disadvantages:

- In comparison to asynchronous communication it has higher speeds, because the system has lesser possibility of error. But, if an error takes place, the complete set of data is lost instead of a single character.
- Serial synchronous transmission is principally used for high-speed communication between computers but is unsuitable where the characters are transferred at irregular intervals.
- Lower overhead and thus, greater throughput.
- Process is more complex
- It is not very cost effective as hardware are more expensive

2. Following are the example for each:
   - Simplex communication: Radio/ Television Broadcasting System
   - half-duplex communication: walky-talky System
   - full duplex communication: Mobile or telephone system

☞ Check Your Progress 3

1. Explain the need of layering in the data communication protocol stack.

   The data communication follows protocols or protocols stack like OSI reference model. Since it is difficult to deal with complex set of rules, and functions required for computer networking, these rules and functions are divided with logical groups called layers. Each layer can be implemented interdependently with an interface to other layers providing with services to it or taking its services like data, connection and error control functions are grouped together into a layer. Speech in telephone conversation is translated, with electrical segments and vice-versa. Similarly in computer system the data or pattern are converted into signals before transmitting and receiving. These function and rules are grouped together into a layer.

2. List and explain any two functions of each OSI layer.

   The seven layers of ISO OSI reference model are:

   i) Physical Layer
   ii) Data Link Layer
   iii) Network Layer
   iv) Transport Layer
   v) Session Layer
   vi) Presentation Layer
   vii) Application Layer.

   a) **The Physical Layer**
      - Physical Layer defines electrical and mechanical specifications of cables and connectors.
      - Specify signaling options for sending control information between two nodes on a network.

   b) **The Data Link Layer**
      - The main task of the data link layer is to provide error free transmission.

- The data link layer creates and recognises frame boundaries. This can be accomplished by attaching special bit patterns to the beginning and end of the frame.

c) **The Network Layer**

- The network layer ensures that each packet travels from its sources to destination successfully and efficiently. It determining how packets are routed from source to destination.
- Addressing is another important task of this layer. The addressing used by the second network may be different from the first one. The second network 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.

d) **The Transport Layer**

- The basic function of the transport layer is to accept data from the session layer, split it up into smaller units if need be, pass these to the network layer, and to ensure that the pieces all arrive correctly at the other end.
- Transport Layer provides location and media independent end-to-end data transfer service to session and upper layers.

e) **The Session Layer**

- Session Establishment and Session Release – Orderly or abort
- Synchronization, Data Exchange and Expedited Data Exchange.

f) **The Presentation Layer**

- Presentation layer is concerned with the syntax and semantics of the information transmitted.
- The presentation layer manages these abstract data structure and converts from the representation used inside the computer to the network standard representation and back.

g) **Application Layer**

- Application Layer supports functions that control and supervise OSI application processes, such as start/maintain/stop application, allocate/deallocate OSI resources, accounting, check point and recovering.
- It also supports remote job execution, file transfer protocol, message transfer and virtual terminal.

UNIT 2: MODULATION AND ENCODING

Structures **Page No.**

2.0 Introduction 25
2.1 Objectives 25
2.2 Need for Modulation 25
2.3 Modulation 26
2.4 Amplitude Modulation 27
2.5 Frequency Modulation 29
2.6 Phase Modulation 31
2.7 Digital Communication 33
2.8 Digital Modulation Techniques 35
2.9 Amplitude Shift Keying (ASK) 35
2.10 Frequency Shift Keying 36
2.11 Phase Shift Keying 37
2.12 Summary 39
2.13 References/Further Reading 39
2.14 Solutions/Answers 39

2.0 INTRODUCTION

Electronic communication has become an analogy for the communication in the present era of electronic gadgets. As a general concept, we can say that transfer of information from one place to another is communication. A significant point about communication is that it involves a sender (transmitter) and a receiver. Only a sender or a receiver can not complete the process of communication. Therefore dual process of "transmitting and receiving" or "coding and decoding" information can be called as communication making it a two way process. In this unit we will discuss about different modulation and encoding techniques. In this unit both analog and digital modulation will be discussed. Further, this unit we will explore how analog signal are converted into digital system and vice-versa.

2.1 OBJECTIVES

After going through this unit, you should be able to:

- Know the concept of modulation
- Understand the different Analog Modulation techniques
- Differentiate between analog and digital modulation
- Know process of analog to digital signal conversion
- Understand the sampling and quantization process
- Know the digital to analog signal conversion process
- Understand the Digital Modulation techniques

2.2 NEED FOR MODULATION

Normal communication signals loose strength as they travel to the large distances. Hence, we often transmit the signals through electromagnetic waves and we use antennas to recover them at a remote point. To send transmitting message signals effectively for long distances, we use Modulation. At the receiver end, after receiving the signal, we need to "move" them back to the original frequency band (baseband) through demodulation. Therefore, we can see the modulation task as "giving wings"

to the information message. However, the original information is retrieved at the receiver end.

2.3 MODULATION

Often, the message being communicated is itself a signal, e.g., an audio signal, and to produce a signal that is suitable for transmission through the channel, we effect some transformation on the message signal. Modulation is the Process by which a property or a parameter of a signal is varied in proportion to another signal. The original signal is normally referred as the modulating signal and the high frequency signal, whose properties are changed, is referred as the carrier signal. The resulting signal is finally referred as the modulated signal.

For example in case of the amplitude modulation, the amplitude of the carrier wave is varied in accordance with the amplitude of the message signal, whereas in the angle modulation, phase angle of the carrier is varied with respect to the message signal.

Benefits of Modulation

1. Modulation can shift the frequency spectrum of a message signal into a band which is better suited to the channel. Antennas only efficiently radiate and admit signals, whose wavelength is similar to their physical aperture. Hence, to transmit and receive, say, voice, by radio we need to shift the voice signal to a much higher frequency band.
2. Modulation permits the use of multiplexing. Multiplexing means allowing simultaneous communication by multiple users on the same channel. For instance, the radio frequency spectrum must be shared and modulation allows users to separate themselves into bands.
3. Modulation can provide some control over noise and interference. For example the effect of noise can be controlled to a large extent by frequency modulation.

Modulation can be classified into two categories Analog Modulation and Digital Modulation. Let's see what are these Analog Modulation and Digital Modulation in detail.

Analog Modulation

Analog modulation is the simplest form of the modulation. In analog modulation, the modulation is applied continuously in response to the analog information signal. The process of the analog modulation has been shown in the Figure 1, below. Here the original signal at the baseband frequency has been shifted to the broadband frequency (F_c) .

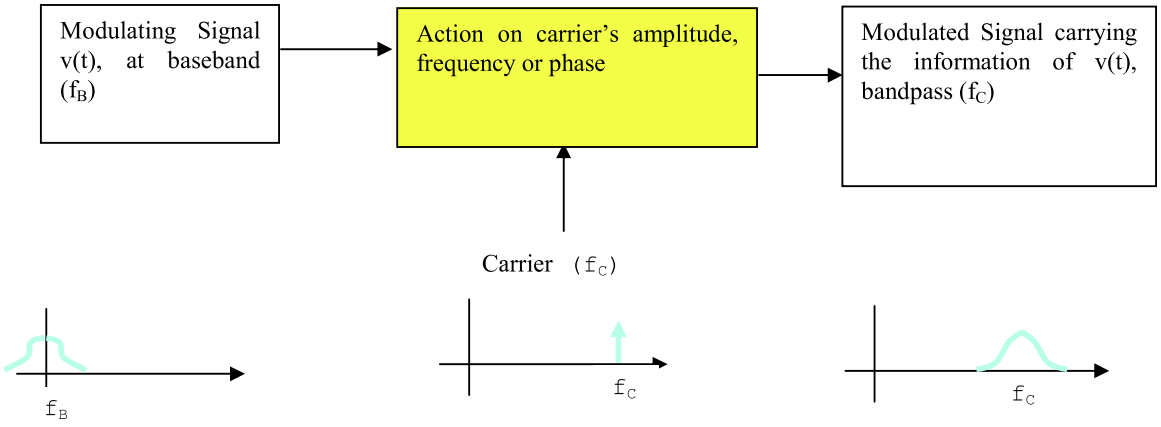


Figure 1: Process of the Analog Modulation

Common analog modulation techniques are:

1. Amplitude Modulation (AM): Here the amplitude of the carrier signal is varied in accordance to the instantaneous amplitude of the modulating signal.

2. Angle Modulation: Here the frequency or phase of the carrier signal is varied in accordance with the strength of the modulating signal. Consequently, the Analog Modulation has two forms:

   i) Frequency Modulation (FM): In this case, the frequency of the carrier signal is varied in accordance to the instantaneous frequency of the modulating signal)

   ii) Phase Modulation (PM): In this case, the phase of the carrier signal is varied in accordance to the instantaneous phase of the modulating signal)

2.4 AMPLITUDE MODULATION

Amplitude modulation (AM) is a technique used in electronic communication, most commonly for transmitting information via a high frequency carrier wave. AM works by varying the strength of the transmitted signal in relation to the information being sent. For example, changes in signal strength may be used to specify the sounds to be reproduced by a load speaker, or the light intensity of television pixels. "Undulatory currents" are the initial implementations of the Amplitude modulation. These were the first method to successfully produce quality audio over telephone lines in 1870's. The Figure 2 illustrates the process of modulation, by showing the modulating, carrier and modulated signals. Figure 3, further illustrates the Amplitude modulation process by varying the amplitudes of the modulating (input signal) and plotting the corresponding modulated signal.

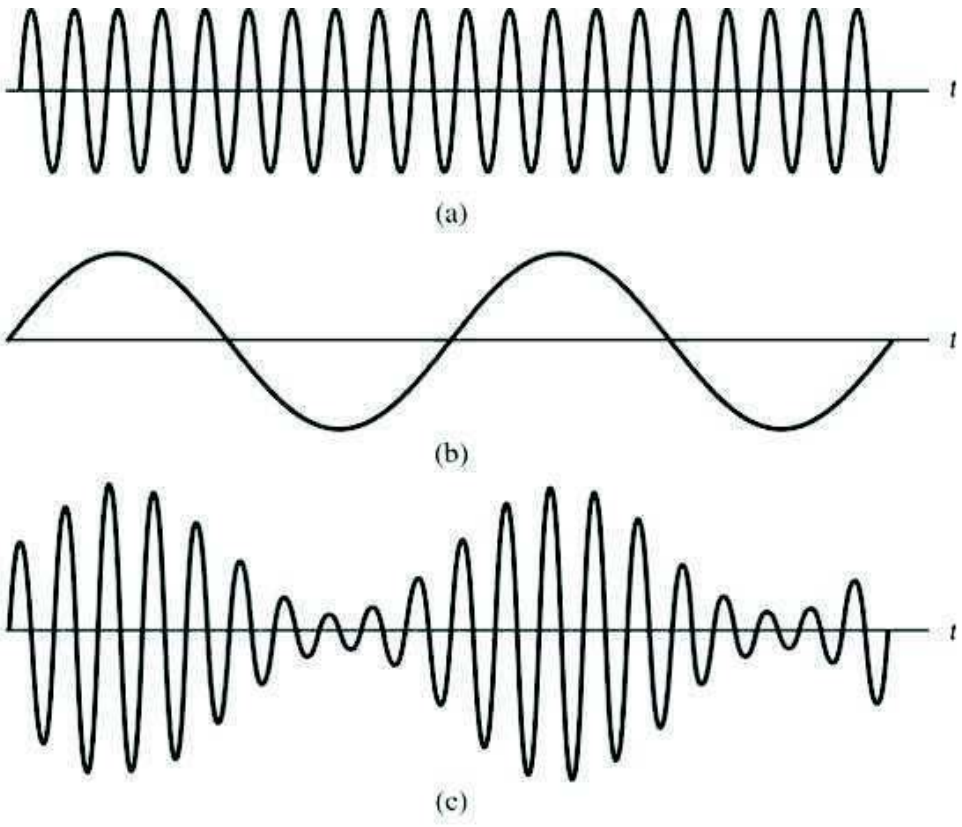


Figure 2: a) Carrier Signal, b) Modulating Signal, c) Modulated Signal

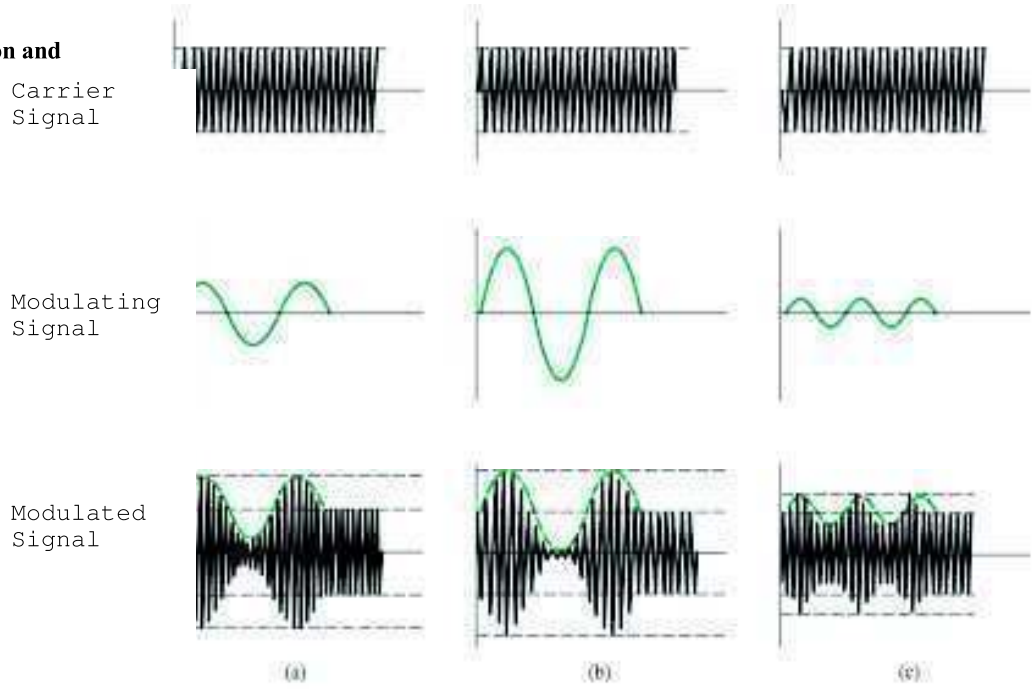


Figure 3: AM Modulation with varying amplitudes of the input signal

Advantages and Disadvantages

Advantages of Amplitude Modulation

1. Coverage area of AM Receiver is wider than FM because atmospheric propagation
2. AM is long distance propagation because of its high power.
3. AM Circuit is the cheapest and least complex,
4. AM can be easily demodulated using a Diode Detector.

Disadvantages of Amplitude Modulation

1. Amplitude modulation is very much sensitive to noise and hence the performance is very weak.
2. Signal of AM is not stronger than FM when it propagates through and obstacle.
3. Only one sideband of AM transmits Information Signal, so it loses power on other sideband and Carrier. Hence the power efficiency of the Amplitude Modulation is very poor.
4. Noise mixes AM Signal in amplitude when it propagates in free space that it makes it difficulty to recover the original signal at receiver in a highly noisy environment.

☞ Check Your Progress 1

1. What is the need for modulation?

   ……

   ……

   ……

   ……

2. What are Analog modulation techniques?

……

……

……

3. Define amplitude modulation.

……

……

……

4. What are the limitations of amplitude modulation?

……

……

……

2.5 FREQUENCY MODULATION

Frequency modulation, FM is widely used for a variety of radio communications applications. FM broadcasts on the VHF bands still provide exceptionally high quality audio, and FM is also used for a variety of forms of two way radio communications, and it is especially useful for mobile radio communications, being used in taxis, and many other forms of vehicle. in view of its widespread use, frequency modulation, FM, is an important form of modulation, despite many forms of digital transmission being used these days. Since its first introduction the use of frequency modulation, FM has grown enormously. Now wideband FM is still regarded as a very high quality transmission medium for high quality broadcasting. FM, is also widely used for communications where it is resilient to variations in signal strength.

Frequency Modulation is the technique in which, the frequency of the carrier wave is changed in accordance with the Amplitude of the modulating signal. The process is shown in the Figure 4 below.

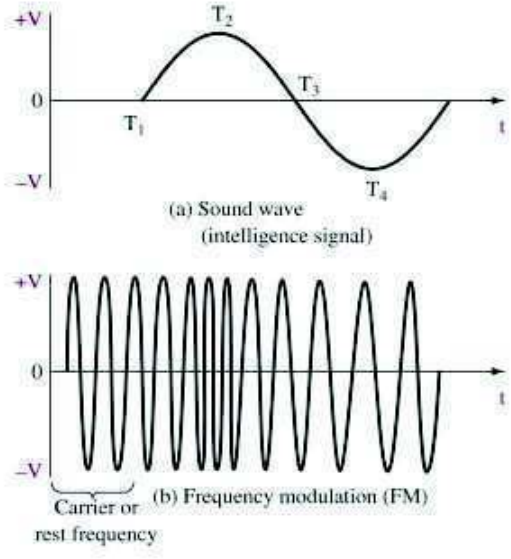


Figure 4: FM representation.

Advantages of Frequency Modulation

There are many advantages of using frequency modulation. These have been widely used for many years, and will remain in use for many years.

- Resilient to noise: One of the main advantages of frequency modulation is that it has been utilised by the broadcasting industry to take care of noise. As most noise is amplitude based, this can be removed by running the signal through a limiter so that only frequency variations appear. This is provided that the signal level is sufficiently high to allow the signal to be limited.

- Resilient to signal strength variations: Signal variations are reduced since noise effect is eliminated. This means that one of the advantages of frequency modulation is that it does not suffer amplitude variations as the signal level varies, and it makes FM ideal for use in mobile applications where signal levels constantly vary. This is provided that the signal level is sufficiently high to allow the signal to be limited.

- Does not require linear amplifiers in the transmitter: As only frequency changes are required to be carried, any amplifiers in the transmitter do not need to be linear.

- Enables greater efficiency than many other modes: The use of non-linear amplifiers, e.g. class C, etc. means that transmitter efficiency levels will be higher - linear amplifiers are inherently inefficient.

Disadvantages of Frequency Modulation

There are a number of dis-advantages to the use of frequency modulation. Some are can be overcome quite easily, but others may mean that another modulation format is more suitable.

- Requires more complicated demodulator: One of the minor dis-advantages of frequency modulation is that the demodulator is a little more complicated, and hence slightly more expensive than the very simple diode detectors used for AM. Also requiring a tuned circuit adds cost. However this is only an issue for the very low cost broadcast receiver market.

- Some other modes have higher data spectral efficiency: Some phase modulation and quadrature amplitude modulation formats have a higher spectral efficiency for data transmission that frequency shift keying, a form of frequency modulation. As a result, most data transmission system uses the digital transmission techniques such as PSK and QAM.

- Sidebands extend to infinity either side: The sidebands for an FM transmission theoretically extend out to infinity. To limit the bandwidth of the transmission, filters are used, and these introduce some distortion of the signal.

Comparison of FM and AM Transmission

Both the Amplitude and Frequency Modulation have their own advantages and disadvantages. However a comparison of the general performance is shown in the table 1 below:

Table 1: Comparison of AM and FM

S. No.	AM Broadcasting	FM Broadcasting
1.	It requires smaller transmission bandwidth	It requires larger bandwidth.
2.	It can be operated in low, medium and high frequency bands.	It needs to be operted in very high and high frequency bands.
3.	It has wider coverage.	Its range is restricted to 50 km.
4.	The demodulation is simple.	The process of demodulation is complex.
5.	The stereophonic transmission is not possible.	In this, stereophonic transmission is possible.
6.	The system has poor noise performance.	It has an improved noise performance.
7.	The AM signal reception does not have any threshold in the useful range of signal noise ratio (SNR).	The FM signal recepition exhibits a three the useful range of signal noise ratio (SM, SNR value should be higher than the ????

2.6 PHASE MODULATION

Frequency Modulation and the Phase Modulation are the two forms of the angle modulation. The main characteristic of the angle modulation is that the amplitude of the carrier frequency is maintained constant, whereas the frequency or phase is changed. In the phase modulation, the phase of the carrier wave is shifted in accordance with the amplitude of the modulating frequency. Phase modulation is a form of modulation that can be used for radio signals used for a variety of radio communications applications. As it will be seen later that phase modulation and frequency modulation are closely linked together and it is often used in many transmitters and receivers used for a variety of radio communication applications from two way radio communications links, mobile radio communications and even maritime mobile radio communications. Phase modulation is also the basis for many forms of digital modulation based around phase shift keying, PSK which is a form of phase modulation. As various forms of phase shift keying are the favored form of modulation for digital or data transmissions, this makes phase modulation particularly important.

Before looking at phase modulation it is first necessary to look at phase itself. A radio frequency signal consists of an oscillating carrier in the form of a sine wave is the basis of the signal. The instantaneous amplitude follows this curve moving positive and then negative, returning to the start point after one complete cycle - it follows the curve of the sine wave. This can also be represented by the movement of a point around a circle, the phase at any given point being the angle between the start point and the point on the wave.

Phase modulation works by modulating the phase of the signal, i.e. changing the rate at which the point moves around the circle. This changes the phase of the signal from what it would have been if no modulation was applied. In other words the speed of rotation around the circle is modulated about the mean value. To achieve this, it is necessary to change the frequency of the signal for a short time. In other words when phase modulation is applied to a signal there are frequency changes and vice versa. Phase and frequency are inseparably linked as phase is the integral of frequency. Frequency modulation can be changed to phase modulation. The information

regarding sidebands, bandwidth and the like also hold true for phase modulation as they do for frequency modulation, bearing in mind their relationship.

Unlike its more popular counterpart, i.e. frequency modulation (FM), PM is not very widely used for radio transmissions. This is because it tends to require more complex receiving hardware and there can be ambiguity problems in determining whether, for example, the signal has changed phase by +180° or -180°. PM is used, however, in digital music synthesizers such as the Casio CZ synthesizers, or to implement FM Synthesis in digital synthesizers such as the Yamaha DX7. The Phase modulation signals have been illustrated in the Figure 5 and Figure 6 below.

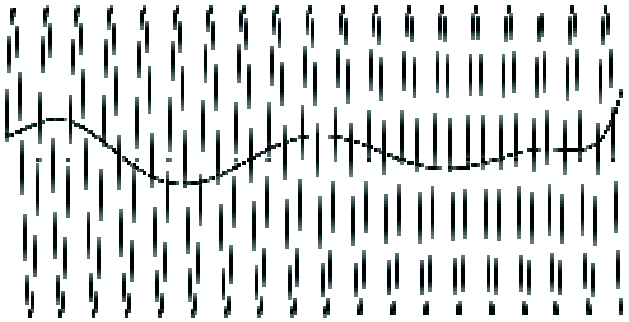


Figure 5: Modulating Signal and the Carrier Wave

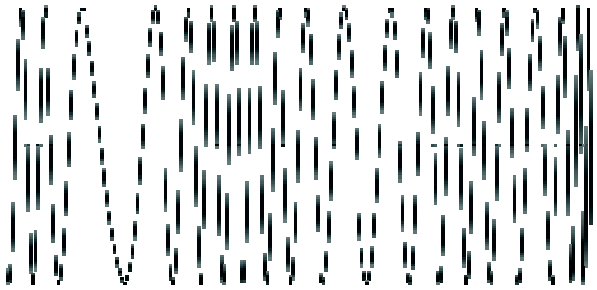


Figure 6: Modulated Wave

☞ Check Your Progress 2

1. Define Frequency Modulation.

……

……

……

……

2. What do you know by angle modulation?

……

……

……

……

3. What are limitations of AM?

……

……

……

……

2.7 DIGITAL COMMUNICATION

Digital communication is the process of communication in which, the signals are transferred in the form of discrete formats rather than the continuous analog forms. Digital communication is very common in the present day communication systems and the signals are normally transmitted in binary formats. It is always easy to process the digital information as compared to the analog signals, because of their discrete nature and hence they have become more popular in the electronic communication. However, the voice based communication is Analog in nature, the signals needs to be converted into the digital formats to process in through the digital communication systems. The opposite process happens, while reconstructing the voice signal at the receiver end. A device called Modem (Modulator + demodulator) in the process. A modem (modulator-demodulator) is a device that modulates an analog carrier signal to encode digital information, and also demodulates such a carrier signal to decode the transmitted information. The goal is to produce a signal that can be transmitted easily and decoded to reproduce the original digital data. The basic process is depicted in the Figure 7 below.

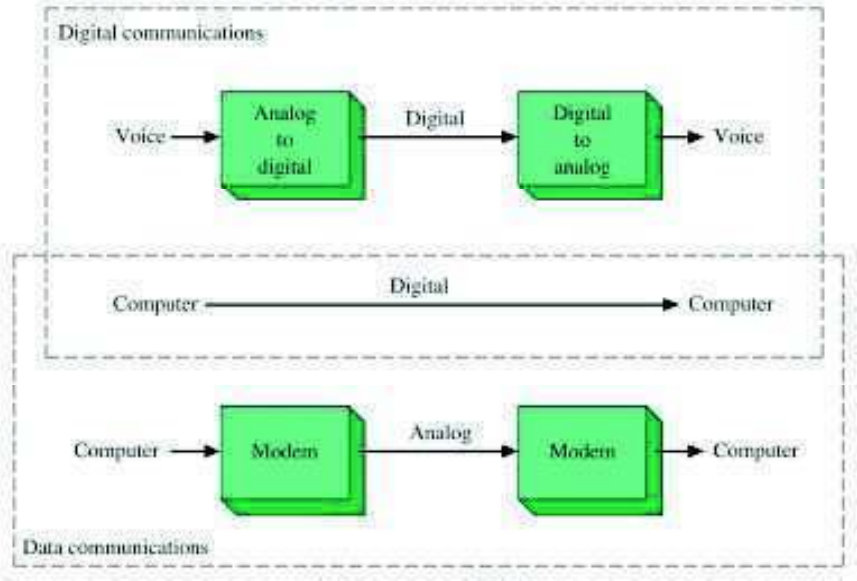


Figure 7: Digital Communication System

Advantages of Digital Communication

1. Reliable communication
2. Less sensitivity to changes in environmental conditions (temperature, etc.)
3. Easy multiplexing
4. Easy signaling
5. Voice and data integration
6. Easy processing like encryption and compression

7. Easy system performance monitoring
8. Quality of Service monitoring
9. Better Signal to Noise Ratio
10. Easy Regeneration of signals

Disadvantages

1. Increased bandwidth requirement for the communication channels.
2. Need for precision timings (Bit, character, frame synchronization needed)
3. Need for the Analog to Digital and Digital to Analog conversions
4. Higher complexity of the system design

Sampling

Digital communication uses the discrete signals; hence the natural analog signals needs to be converted to the discrete signals, in order to process them digitally. For this, purpose a technique known as sampling is employed. In electronic signals, sampling is the reduction of a continuous signal to a discrete signal. A sample refers to a value or set of values at a point in time and/or space. The process is illustrated in the Figure 8 below.

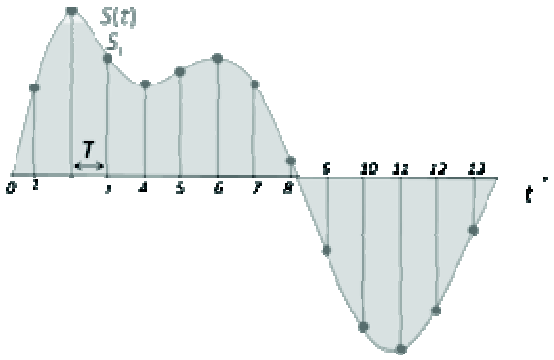


Figure 8: Sampling from the Analog Signals

Sampling is the first step towards the digitization. However, in order to codify these samples, the flat top sampling is most widely used. The block diagram of the natural and flat top sampling has been shown in the Figure 9 below.

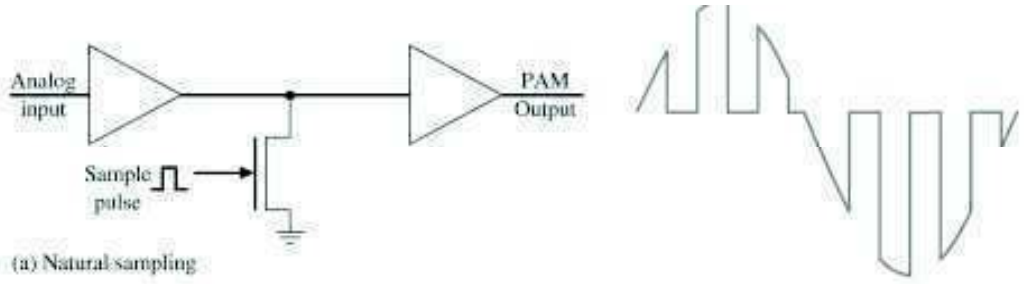


Figure 9: a) Natural sampling

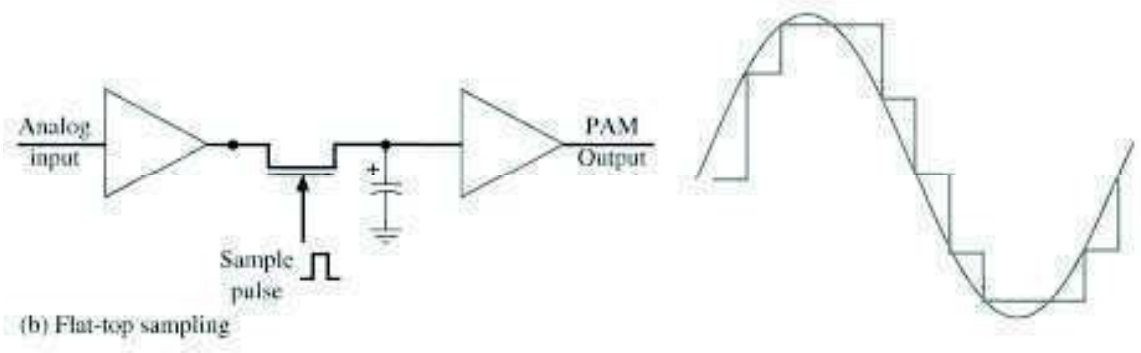


Figure 9: b) flat-top sampling

Once we get the samples, these samples are then quantized as per the voltage levels and finally converted to the binary codes to process them digitally. This process along with corresponding voltage levels and binary codes are shown in the Figure 9 b above.

Analog to Digital Conversion

An analog-to-digital converter (ADC, A/D) is a device that converts the input continuous physical quantity to a digital number that represents the quantity's amplitude. Instead of doing a single conversion, an ADC often performs the conversions ("samples" the input) periodically. The result is a sequence of digital values that have converted a continuous-time and continuous-amplitude analog signal to a discrete-time and discrete-amplitude digital signals. The most commonly employed A/D converter is the Ramp based circuit. It uses a comparator to compare the voltage levels

Digital to Analog Conversion

Digital to analog converter is the electronic circuit, which takes digital input and converts this into an analog waveform. A common use of digital-to-analog converters is generation of audio signals from digital information in music players. Digital video signals are converted to analog in television and cell phones to display colors and shades

2.8 DIGITAL MODULATION TECHNIQUES

There are three major classes of digital modulation techniques used for transmission of digital data:

- Amplitude Shift Keying
- Frequency Shift Keying(FSK)
- Phase-shift keying (PSK)

All of these processes convey the data by changing some aspect of a carrier wave, in response to a data signal.

2.9 AMPLITUDE SHIFT KEYING (ASK)

Amplitude-shift keying (ASK) is a form of modulation that represents digital data as variations in the amplitude of a carrier wave.

Any digital modulation scheme uses a finite number of distinct signals to represent digital data. ASK uses a finite number of amplitudes, each assigned a unique pattern of binary. Usually, each amplitude encodes an equal number of bits. Each pattern of bits forms the symbol that is represented by the particular amplitude. The demodulator, which is designed specifically for the symbol-set used by the modulator,

determines the amplitude of the received signal and maps it back to the symbol it represents, thus recovering the original data. Frequency and Phase of the carrier are kept constant.

Like Amplitude Modulation, ASK is also linear and sensitive to atmospheric noise, distortions, propagation conditions on different routes in PSTN (Public Switched Telephone Network) etc. Both ASK modulation and demodulation processes are relatively inexpensive. The ASK technique is also commonly used to transmit digital data over optical fiber. For LED transmitters, binary 1 is represented by a short pulse of light and binary 0 by the absence of light. Laser transmitters normally have a fixed "bias" current that causes the device to emit a low light level. This low level represents binary 0, while a higher-amplitude light wave represents binary 1. The simplest and most common form of ASK operates as a switch, using the presence of a carrier wave to indicate a binary one and its absence to indicate a binary zero. This type of modulation is called on-off keying, and is used at radio frequencies to transmit Morse code.

More sophisticated encoding schemes have been developed which represent data in groups using additional amplitude levels. For instance, a four-level encoding scheme can represent two bits with each shift in amplitude; an eight-level scheme can represent three bits; and so on. These forms of amplitude-shift keying require a high signal-to-noise ratio for their recovery, as by their nature much of the signal is transmitted at reduced power.

2.10 FREQUENCY SHIFT KEYING

Frequency-shift keying (FSK) is a frequency modulation scheme in which digital information is transmitted through discrete frequency changes of a carrier wave. The simplest FSK is a binary FSK (BFSK). BFSK uses a pair of discrete frequencies to transmit binary (0s and 1s) information. With this scheme, the "1" is called the mark frequency and the "0" is called the space frequency.

FSK Transmitter

The block diagram of the FSK modulator is shown below in Figure 10. The modulating signal and the carrier frequency are fed to the frequency modulator circuitry and correspondingly the output is transmitted in the form of a signal with varied frequency.

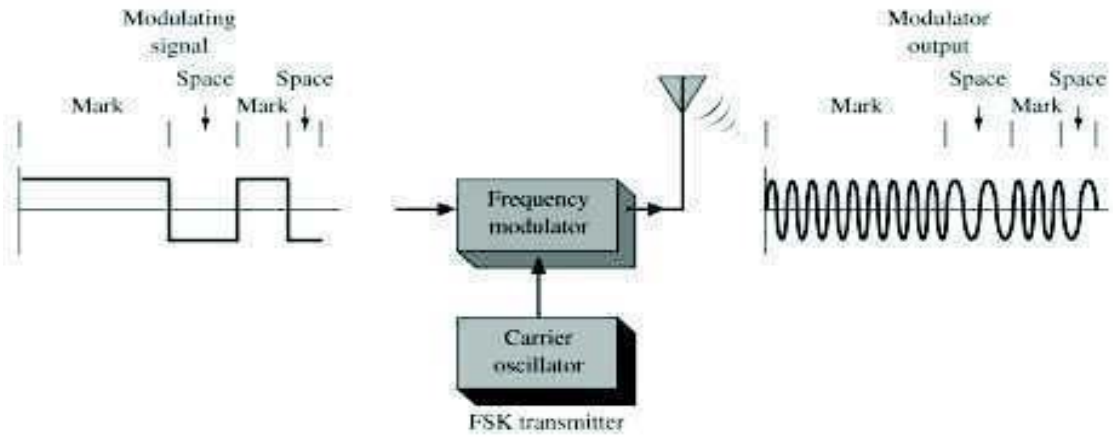


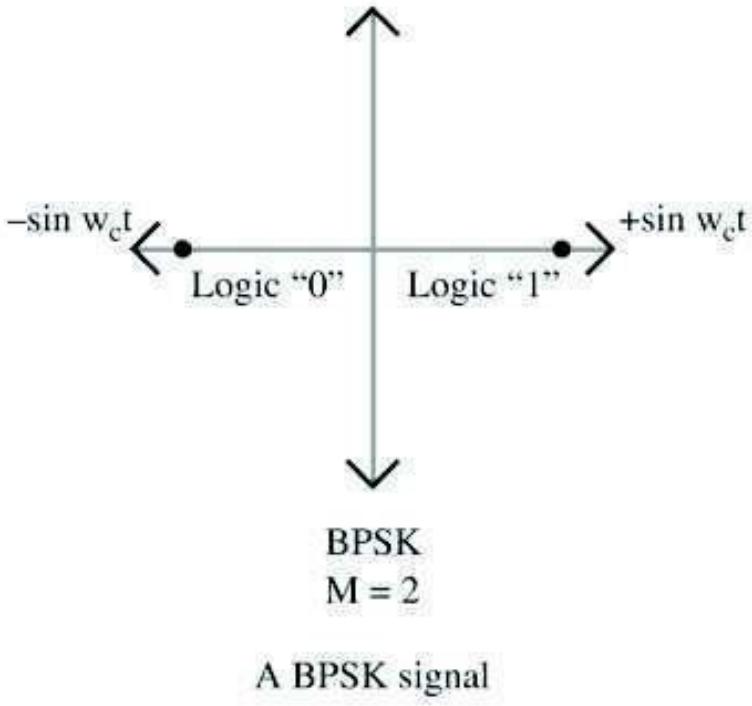
Figure 10: FSK Modulator

2.11 PHASE SHIFT KEYING

Phase-shift keying (PSK) is a digital modulation scheme that communicates the data by changing, or modulating, the phase of the carrier wave. Any digital modulation scheme uses a finite number of distinct signals to represent digital data. PSK uses a finite number of phases, each assigned a unique pattern in the form of a binary code. Each pattern of bits forms the symbol that is represented by the particular phase. On the other hand, the demodulator is designed specifically for the symbol-set used by the modulator. It determines the phase of the received signal and maps it back to the symbol it represents, thus recovering the original data. This requires the receiver to be able to compare the phase of the received signal to a reference signal. Another simple way of operation is that instead of operating with respect to a constant reference wave, the broadcast can operate with respect to itself. Changes in phase of a single broadcast waveform can be considered the significant items. In this system, the demodulator determines the changes in the phase of the received signal rather than the phase (relative to a reference carrier wave) itself. Since this scheme depends on the difference between successive phases, it is termed differential phase-shift keying (DPSK). DPSK can be significantly simpler to implement than ordinary PSK since there is no need for the demodulator to have a copy of the reference signal to determine the exact phase of the received signal.

Like any form of shift keying, there are defined states or points that are used for signaling the data bits. The basic form of binary phase shift keying is known as Binary Phase Shift Keying (BPSK) or it is occasionally called Phase Reversal Keying (PRK). A digital signal alternating between +1 and -1 (or 1 and 0) will create phase reversals, i.e. 180 degree phase shifts as the data shifts state. This has been illustrated in the Figure 11 below.

The problem with phase shift keying is that the receiver cannot know the exact phase of the transmitted signal to determine whether it is in a mark or space condition. This would not be possible even if the transmitter and receiver clocks were accurately linked because the path length would determine the exact phase of the received signal. To overcome this problem PSK systems use a differential method for encoding the data onto the carrier. This is accomplished, for example, by making a change in phase equal to a one, and no phase change equal to a zero. Further improvements can be made upon this basic system and a number of other types of phase shift keying have been developed. One simple improvement can be made by making a change in phase by 90 degrees in one direction for a one, and 90 degrees the other way for a zero. This retains the 180 degree phase reversal between one and zero states, but gives a distinct change for a zero. In a basic system not using this process it may be possible to loose synchronization if a long series of zeros are sent. This is because the phase will not change state for this occurrence. There are many variations on the basic idea of phase shift keying. Each one has its own advantages and disadvantages enabling system designers to choose the one most applicable for any given circumstances. Other common forms include QPSK (Quadrature phase shift keying) where four phase states are used, each at 90 degrees to the other, 8-PSK where there are eight states used and so forth. For an example the output of a BPSK modulator circuit for a 1010101 input is shown in figure 11.



A BPSK signal

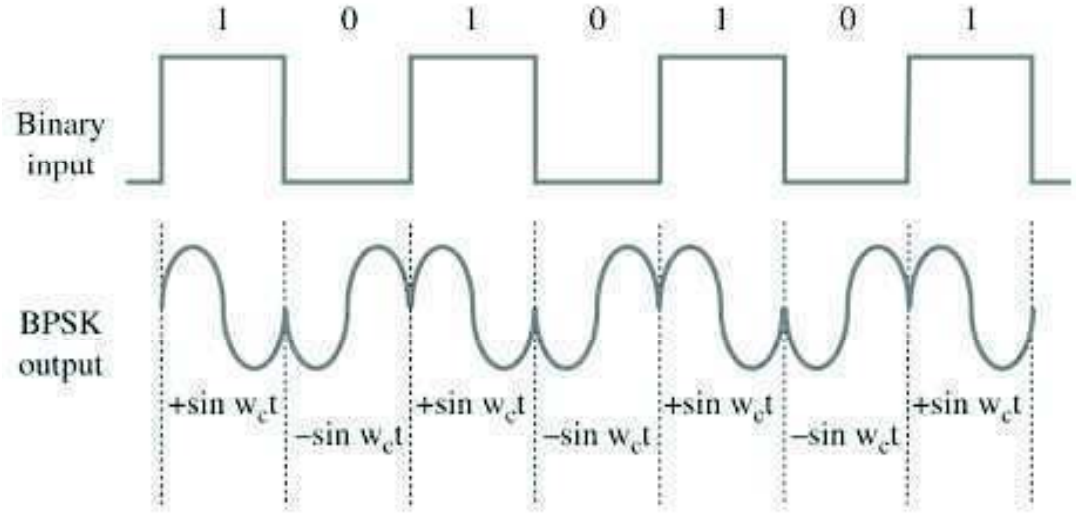


Figure 11: The output of a BPSK modulator circuit for a 1010101 input.

☞ Check Your Progress 3

1. What are the different digital modulation techniques?

……

……

2. How many different phase states are used in BPSK and QPSK?

……

……

……

……

3. Why digital modulation is better than the Analog Modulation?

……

……

……

……

2.12 SUMMARY

After completing this unit we are sure that you have understood the term modulation. Why modulation is need in out communication systems. In this unit we have studied about different modulation techniques both analog and digital modulation type. We have also discussed different techniques for converting the analog signals into digital system and vice-versa.

2.13 REFERENCES/FURTHER READING

1. *Computer Networks*, A. S. *Tanenbaum* 4th Edition, Practice Hall of India, New Delhi. 2003.
2. *Introduction to Data Communication & Networking,* 3rd Edition, Behrouz Forouzan, Tata McGraw Hill.
3. *Computer Networking*, J.F. Kurose & K.W. Ross, A Top Down Approach Featuring the Internet, Pearson Edition, 2003.
4. *Communications Networks*, Leon Garcia, and Widjaja, Tata McGraw Hill, 2000.
5. *Data and Computer Communications,* Willian Stallings, 6th Edition, Pearson Education, New Delhi.
6. www.wikipedia.org
7. Larry L. Peterson, *Computer Networks*: A Systems Approach, 3rd Edition (The Morgan Kaufmann Series in Networking).

2.14 SOLUTIONS/ANSWERS

☞ **Check Your Progress 1**

1. To send transmitting message signals effectively for long distances, we use Modulation.

2. The Analog modulation techniques are:
   a) Amplitude modulation
   b) Angle modulation
      i) Frequency modulation
      ii) Phase modulation.

3. This is defined as the modulation, in which amplitude of carrier is changed in accordance to the amplitude of the modulating signal.

4. The amplitude modulation suffers from the following limitations

i) The useful power is contained in the sidebands and even at 100% modulation the and contain only 33% of the total power and hence the modulation efficiency is poor.

ii) Due to poor efficiency, the transmitters employing amplitude modulation have very poor range.

iii) The reception in this modulation is noisy. The radio receiver picks up all the surrounding noise along with the signal.

☞ Check Your Progress 2

1. Frequency Modulation is the technique in which, the frequency of the carrier wave is changed in accordance with the Amplitude of the modulating signal.

2. It is possible to convey or transmit information by varying its frequency as well as angle of phase. These are known as frequency and phase modulations respectively and both collectively are known as "Angle Modulation". The frequency and phase modulation systems have similar characteristics with minor differences.

   Briefing we can say angle modulations of two types:

   i) Frequency modulation

   ii) Phase modulation

3. Limitations of AM:

   i) Power of carrier and of one side band is useless.

   ii) The AM reception is noisy.

   iii) The BW is much less.

   iv) Only two S.Bs are available.

☞ Check Your Progress 3

1. There are three major classes of digital modulation techniques used for transmission of digitally represented data:

   - Amplitude Shift Keying
   - Frequency Shift Keying(FSK)
   - Phase-shift keying (PSK)

2. BPSK uses two different phase states and each one differs by 180° , whereas the QPSK uses four different phases and each one differs by 90°.

3. i) It is easy to process the digital information.

   ii) Digital systems are less prone to noise.

   iii) Digital signals can be easily re-transmitted.

`

UNIT 3: MULTIPLEXING AND SWITCHING

Structure **Page No.**

3.0 Introduction 41
3.1 Objectives 41
3.2 Multiplexing concept 42
3.3 Frequency-Division Multiplexing 43
3.4 Time-Division Multiplexing 45
3.5 Code Division Multiplexing 46
3.6 Space Division Multiplexing 47
3.7 Switching 48
3.8 Message Switching 50
3.9 Circuit Switching 51
3.10 Packet Switching 52
3.10.1 Connection Less Packet Switching
3.10.2 Connection Oriented Packet Switching
3.11 Summary 57
3.12 References/Further Reading 57
3.13 Solutions/Answers 57

3.0 INTRODUCTION

The most fundamental need of any communication system design is to cater to large number of users. But this requires a large number of resources and large bandwidths supporting multiple channels. Requirement for large number of resources can be met if the resources are available, but this makes it cost ineffective. Therefore, the aim is always to use minimum number of resources and make their utilisation to their fullest potential. Bandwidth always remains a critical resource due to its limited availability and therefore, communication systems try to harness its fullest potential. Networks always require us to accommodate multiple signals utilizing a single piece of cabling to make it cost effective and reduce complexity. This need is seen throughout networking whether we are talking about local area networks or wide area ones. Modern telephone systems must place a large number of calls over a limited amount of bandwidth (i.e. a trunk). Broadband LANs must have several different types of data on a single wire at once. For these applications, we need to share the resources and in particular the bandwidth. Multiplexing and Switching are the two most important techniques being employed for this purpose in the present day communication systems and have been discussed in the present unit.

3.1 OBJECTIVES

After going through this unit, you should be able to:

- Know the concept of Multiplexing and Switching in computer networks
- Understand the basic multiplexing techniques like FDM, TDM, CDM and SDM
- Differentiate between different types of multiplexing techniques
- Know the switching mechanisms
- Differentiate between packet, circuit and message switching
- Understand the different packet switching mechanisms

3.2 MULTIPLEXING CONCEPT

In general, a medium can carry only one signal at any moment in time. For multiple signals to share one medium, the medium must somehow be divided, giving each signal a portion of the total bandwidth. Multiplexing (also known as MUXing) is a method by which multiple analog message signals or digital data streams are combined into one signal over a shared medium. The basic aim of the Multiplexing is to share an expensive resource by putting-up multiple signals on the same channel. For example, in telecommunications, several telephone calls may be carried using one wire. Multiplexing originated in telegraphy in the 1870s, and is now widely applied in different streams of communications. When several communication channels are needed between the same two points, significant economies may be realized by sending all the messages on one transmission facility – called multiplexing. As shown in Figure 1, n number of signals from the low speed channels have been combined to one sigh speed link using a n:1 multiplexer. Whereas the opposite process is carried out at the other end, where the signals are further separated into n number of low speed channels. This opposite process is referred as demultiplexing.

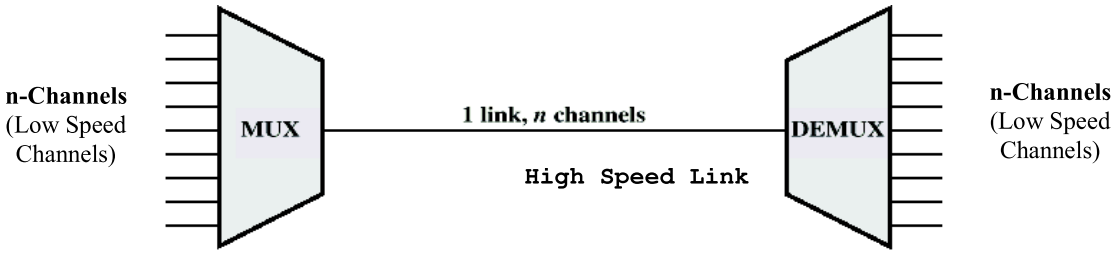


Figure 1: Multiplexing and De-Multiplexing

Thus, Multiplexing refers to the ability to transmit data coming from several pairs of equipment (transmitters and receivers) called *low-speed channels* on a single physical medium (called the *high-speed channel*). Whereas, a *multiplexer* is the multiplexing device that combines the signals from the different transmitters and sends them over the *high-speed channel*. A *demultiplexer* is the device which separates signal received from a *high-speed channel* into different signal and sends them to receivers.

There are four basic multiplexing techniques:

- Frequency division multiplexing (FDM)
- Time division Multiplexing (TDM)
- Code division Multiplexing (CDM)
- Space-division Multiplexing (SDM)
- Frequency division Multiplexing: Bandwidth is divided into different smaller frequency bands (range).
- Time division Multiplexing (TDM) (Time slots are allocated to message signals in an non overlapping manner in the time domain so that individual messages can be recovered from time synchronized switches)
- Quadrature Carrier/amplitude Multiplexing (QAM): Two message signals are transmitted in the same frequency band. The recovery is possible due to the carrier signals being orthogonal)

- Code division Multiplexing (CDM) users occupy the same frequency band but modulate their messages with different codes TDMA FDMA CDMA when used for multiple access TDMA, FDMA, e.g., GSM, FM, AM, Wireless networks

☞ **Check Your Progress 1**

1. Define multiplexing.

……..

……..

……..

……..

……..

2. State the importance of multiplexing.

……..

……..

……..

……..

……..

3. What are the multiplexing techniques?

……..

……..

……..

……..

……..

3.3 FREQUENCY-DIVISION MULTIPLEXING

Frequency division multiplexing (FDM) is the technique used to divide the available bandwidth into a number of smaller independent logical channels with each channel having a small bandwidth. The method of using a number of carrier frequencies each of which is modulated by an independent speech signal is in fact frequency division multiplexing. The following Figure 2 depicts the basic process of frequency division multiplexing, in which the total bandwidth has been divided into n-number of different channels and each one of them working with a specific bandwidth.

The following figure 2 depicts how three voice-grade telephone channels are multiplexed using FDM. When many channels are multiplexed together, 4000Hz is allocated to each channel to keep them well separated. First the voice channels are raised in frequency, each by a different amount. Then they can be combined, because no two channels can occupy the same portion of the spectrum. Notice that even though there are gaps (guard bands) between the channels, there is some overlap between adjacent channels, because the filters do not have sharp edges. This overlap means that a strong spike at the edge of one channel will be felt in the adjacent one as non-thermal noise.

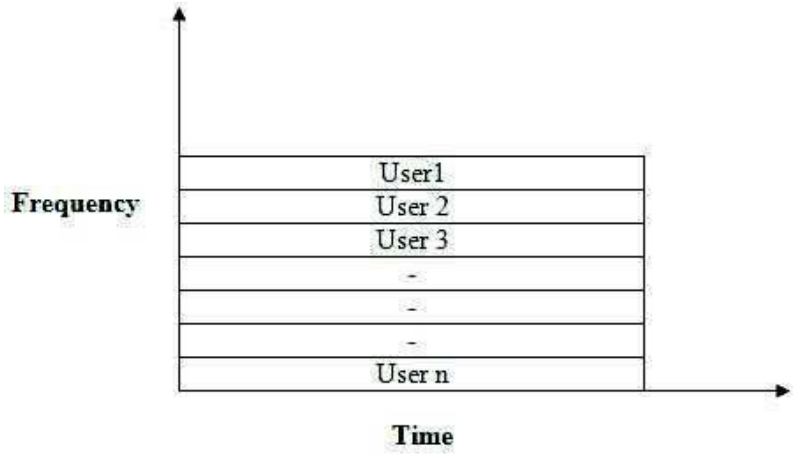


Figure 2: Frequency Division Multiplexing

In the telecommunication technology, the total bandwidth available in a communication medium is divided into a series of non-overlapping frequencies sub-bands using the frequency division multiplexing. Each one of these sub-bands then carries a separate signal. This allows a single transmission medium such as a cable or optical fiber to be shared by many signals. An example of a system using FDM is cable television, in which many television channels are carried simultaneously on a single cable. FDM is also used by telephone systems to transmit multiple telephone calls through high capacity trunk lines, communications satellites to transmit multiple channels of data on uplink and downlink radio beams, and broadband DSL modems to transmit large amounts of computer data through twisted pairs telephone lines, among many other uses.Frequency-division multiplexing works best with low-speed devices. The frequency division multiplexing schemes used around the world are very standardized. A wide spread standard is 12, 4000-Hz each voice channels (3000Hz for user, plus two guard bands of 500Hz each) multiplexed into the 60 to 108 KHz band. Many carriers offer a 48 to 56 kbps leased line service to customers, based on the group. The frequency band division has been illustrated in the Figure 3 taking some example frequencies.

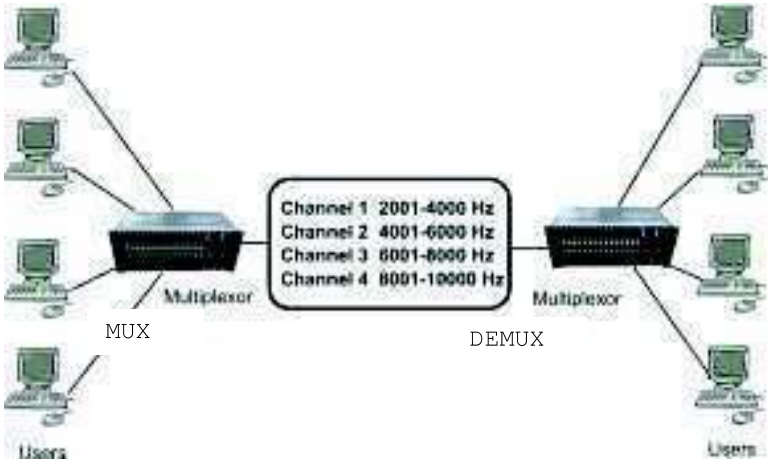


Figure 3: Illustration of FDM using four different channels

In Telephony, the most widely used method of modulation in FDM is single sideband modulation, which, in the case of voice signals, requires a bandwidth that is approximately equal to that of the original voice signal. Each voice input is usually assigned a bandwidth of 4 KHz. The bandpass filters following the modulators are used to restrict the band of each modulated signal to its prescribed range. The

resulting bandpass filter outputs are combined in parallel to form the input to the common channel. At the receiving terminal, a bank of band pass filters, with their inputs connected in parallel, is used to separate the message signals on a frequency-occupancy basis. The original message signals are recovered by individual demodulators

Frequency division multiplexing (FDM) is also referred as the Wavelength division multiplexing (WDM), where we are using the optical communications focusing on the wavelength rather than the frequency.

Advantages of FDM:

1. The users can be added to the system by simply adding another pair of transmitter modulator and receiver demodulators.
2. FDM system support full duplex information (Both side simultaneous Communication) flow which is required by most of application.

Disadvantages of FDM:

1. In FDM system, the initial cost is high. This may include the cable between the two ends and the associated connectors for the cable.
2. A problem with one user can sometimes affect the others.
3. Each user requires a precise carrier frequency for transmission of the signals.

3.4 TIME-DIVISION MULTIPLEXING

Time Division Multiplexing (TDM) is another popular method of utilizing the capacity of a physical channel effectively. Each user of the channel is allotted a small time interval during which it may transmit a message. Thus the total time available in the channel is divided and each user is allocated a time slot. Data from each user is multiplexed into a frame which is transmitted over the channel. In TDM, user's messages are buffered as they received and read from the buffer during its time slot to make a frame. Therefore each user can use the full channel bandwidth. The channel capacity is fully utilized in TDM by interleaving a number of messages belonging to different users into one long message. This message sent through the physical channel must be separated at the receiving end. Individual chunks of message sent by each user should be reassembled into a full message. The process of the Time division multiplexing has been shown in Figure 4.

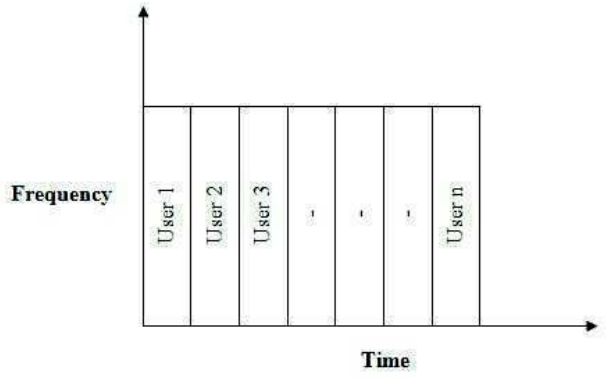


Figure 4: Time Division Multiplexing

Sharing of the signal is accomplished by dividing available transmission time on a medium among users. For example, in some countries, the individual stations have two logical sub channels: music and advertising. These two alternate in time on the same frequency first a burst of music, then a burst of advertising, then more music and

so on. This situation is time division multiplexing. Unfortunately, TDM can only be used for digital data multiplexing. Since local loops produce analog signals, a conversion is needed from analog to digital in the end office. Where all the individual local loops come together to be combined onto outgoing trucks. The TDM process is further illustrated in Figure 5 with the digital data stream.

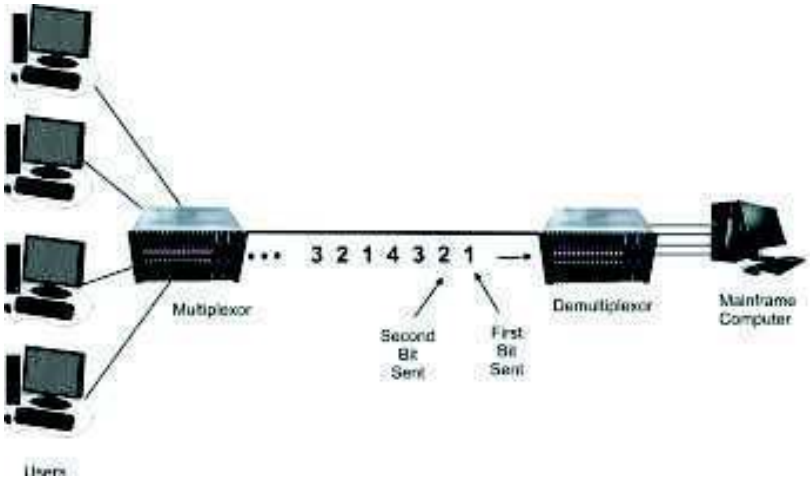


Figure 5: Digital Transmission using TDM

Applications of TDM

- The PDH (Plesiochronous Digital Hierarchy) system, also known as the PCM (Pulse Code Modulation) systems
- The synchronous digital hierarchy (SDH) / synchronous optical networking (SONET) network transmission standards.
- TDM can be further extended into the time division multiple Channel (TDMA) scheme, where several stations connected to the same physical medium, for example sharing the same frequency channel, can communicate. Application examples include the widely used GSM telephone system

Advantages of TDM

1. It uses a single link
2. It does not require precise carrier matching at both end of the links.
3. Use of the channel capacity is high.
4. Each to expand the number of users on a system at a low cost.
5. There is no need to include identification of the traffic stream on each packet.

Disadvantages of TDM

1. The sensitivity to other user is very high and causes problems
2. Initial cost is high
3. Technical complexity is more

3.5 CODE DIVISION MULTIPLEXING

As you may know, the concept of multiple access where we can allow several transmitters to send information simultaneously over a single communication channel and it allows several users to share a band of frequencies (or you can say bandwidth).

CDMA uses spread-spectrum technology and a special coding scheme (where each transmitter is assigned a code generally pseudorandom code) to allow multiple users to be multiplexed over the same physical channel. By contrast, time division multiple access (TDMA) divides access by time, while frequency-division multiple access (FDMA) divides it by frequency. CDMA is a form of spread-spectrum signalling, since the modulated coded signal has a much higher data bandwidth than the data being communicated. This allows more users to communicate on the same network at one time than if each user was allotted a specific frequency range. Remember that CDMA is a digital technology, so analog signals must be digitized before being transmitted on the network.

3.6 SPACE DIVISION MULTIPLEXING

When we want to transmit multiple messages through any of the communication media, the ultimate goal is to maximize the use of the given resources (e.g. time and frequency in general). It involves grouping many separate wires into a common cable enclosure. A cable that has, for example, 50 twisted pairs inside it can support 50 channels. SDM has the unique advantage of not requiring any multiplexing equipment. It is usually combined with other multiplexing techniques to better utilize the individual physical channels. For example, if there are six persons in the office and all of them want to talk at the same time, this will give rise to interference between the conversations. To reduce the interference they may divide themselves into three groups of two, such that the conversation is between each pair of people. If the pairs continue talking whilst sitting next to each other, the interference would still be present. The best way for each pair to converse with minimal interference would be to sit a few feet away from the other pairs (within the same room) and converse. They would still be sharing the same medium for their conversations but the physical space in the room would be divided for each conversation. This is the simplest example of Space Division Multiplexing. The concept of SDM has been illustrated in Figure 6.

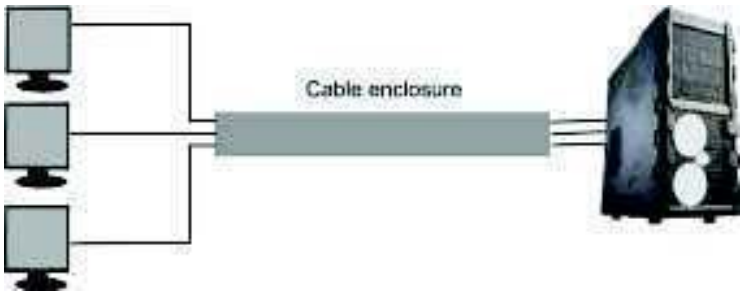


Figure 6: Space Division Multiplexing

Space Division Multiplexing is the multiplexing technique in which both the time and frequency can be reused by transmitting our information through a parallel set of channels.

In wired communication, space-division multiplexing simply implies different point-to-point wires for different channels. Examples include an analogue stereo audio cable, with one pair of wires for the left channel and another for the right channel, and a multipair telephone cable usually employed to provide PSTN connections in different homes. Another example is a switched star network such as the analog telephone access network (although inside the telephone exchange or between the exchanges, other multiplexing techniques are typically employed). In wireless communication, space-division multiplexing is achieved by multiple antenna elements forming a phased array antenna. Examples are multiple-input and multiple-output (MIMO), single-input and multiple-output (SIMO) and multiple-input and single-output (MISO) multiplexing.

3.7 SWITCHING

Switching forms a very important process in a communication system. A switch is used to connect the incoming link to the desired outgoing link and directs the incoming message to the appropriate outgoing link. Let us understand the concept of switching with the help of a simple illustrative example.

Consider a group of 8 people with telephones. If we were to use direct lines between all the people, we would need 28 duplex (wires that allow simultaneous two-way conversation) lines. The arithmetic is pretty simple - to connect **n** subscribers directly, we need **n(n-1)/2** lines. This is alright as long as the number of subscribers is less and the distances are also small. But in the present day electronic communication systems, we are talking about connecting the entire world - obviously direct connections are not the answer. We need to design a system, which can connect the people from anywhere. Now, if we were to use a switch instead, we could reduce the number of lines needed to just 4, because with 8 subscribers, there would at the most be just 4 conversations simultaneously. The switch would have 4 lines internally and it would use the each line to connect a pair of subscribers. This has been illustrated in Figure 7 below.

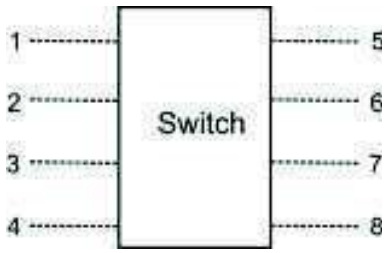


Figure 7: A simple switch with 4-input and 4-output lines

Let us assume the switch in the above diagram has 4 internal lines A, B, C and D. Say A is being used to connect 1 to 7 and B to connect 4 to 5. Now if 3 were to wish to get connected to 8, the switch would 'patch' the ends of C so that 3 and 8 are connected. Instead, if 6 had lifted the phone before the 3 and tried to get connected to 2, the switch would use C to 'patch' a connection between 6 and 2. We assume that the order in which the lines A, B, C and D are used is in accordance with their alphabetical order. This assumption is valid and any other order would not have any bearing upon the concept of switching. The fact remains that the lines A, B, C and D are not fixed. Their end-points change from time to time. *Thus they are* switched *circuits.*

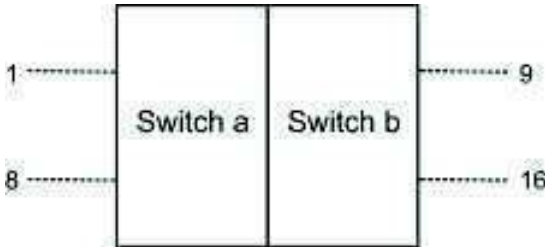


Figure 8: Two Switches with 8-subscribers

Consider 2 such 8-subscriber switches as shown in the Figure 8. They are seemingly connected by just one line. But this line is a multiplexed line, and is not switched. It is called a trunk. It is not switched because it always carries traffic from switch a switch b. Let us assume the multiplexer is capable of sending 4 simultaneous conversations over one line. Then the trunk could be carrying conversations between 1-15, 4-9, 8-11 and 5-12. The lines internal to switch a would connect 1,4,8 and 5 to the multiplexer /

demultiplexer (remember the line is duplex) and therefore the trunk. Similarly, the lines internal to switch b would connect 15, 9, 11 and 12 to the multiplexer/ demultiplexer. The switched circuits are inside the switches a and b. But the trunk between a and b is multiplexed with 4 conversations. So in a sense, the trunk is not switched. But if you had more than one trunk between switches a and b, then the trunks would also be switched. Why, because a call from 1 to 15 could go on either trunk 1 or trunk 2 (assuming there are two trunks each capable of carrying 4 conversations). Thus, the trunks are now switched, in addition to being multiplexed.

It is very important to understand the difference between switching and multiplexing. In simple terms, multiplexing is done to maximize the use of a communications channel. Whereas, the switching is the manipulation of the ends of the communications channel and is used to make the connections. The purpose of an electrical switch is to close /open a circuit to allow/stop flow of current. A communication switch is similarly used to allow/stop flow of message through the path connecting the receiver and the transmitter. Two users, one can be called sender and the other receiver, can be connected by a medium like a conducting wire over which messages in the form of electrical signals can be transmitted from one user to the other. A switch inserted in the electrical path between the two users facilitates connection/disconnection of the users as desired by controlling the switch. The path need not be on all the time. It needs to be switched on only when the users need to communicate. The role of such a switch becomes more important when there are a large number of users and a particular user at one time may want to communicate with another user and wants to communicate still another user at a different time. Thus the same user has to be connected to two different users at two different times. This can be done by a controlled switch. Thus in a set of say n users, different users may like to communicate with different users at different time. The simple 2 X 2 switch has been illustrated in the Figure 9 below.

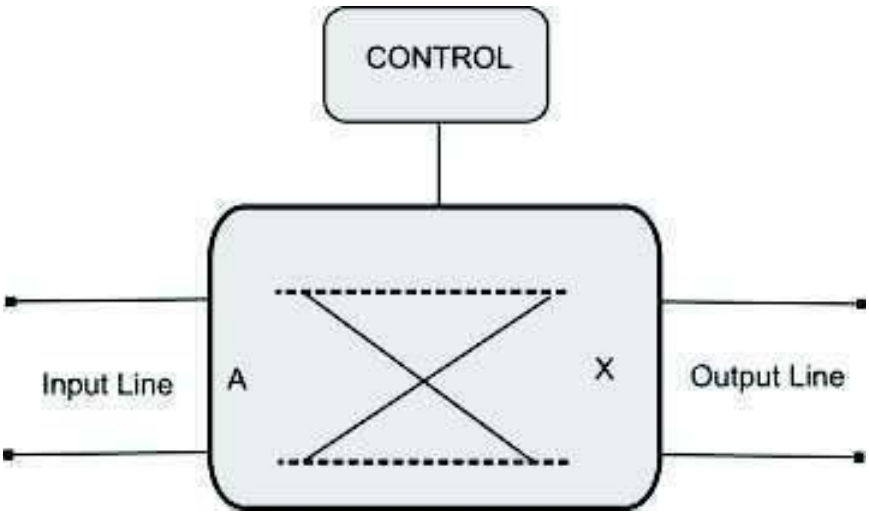


Figure 9: A Simple 2 X 2 Switch

Switching plays a very important role in telecommunication networks. It enables any two users to communicate with each other. Basically, there are three categories of Switching

a) Message Switching

b) Circuit Switching

c) Packet Switching

A circuit switch closes a circuit between the incoming and the outgoing paths so that the incoming message can go to the output link. The circuit between any two desired paths is closed by a control signal applied to the switch. In message and packet switching, the incoming message/packet to the node is stored in a memory location. Then the stored message/packet is transferred to another desired memory location, from where the message/packet can be delivered/forwarded to the next node or the receiver. The transfer from the incoming bin to the outgoing bin is done with a control/command signal.

3.8 MESSAGE SWITCHING

Message switching is one of the initial mode of switching, which helped a lot in the proliferation of the electrical communication. It is interesting to know that electrical communication in the form of Telegraph arrived earlier than the Telephone. Let us try to understand the working of the Telegraph system to build the concept of message switching. Consider the Figure 10 as a working model of the Telegraph Network. As an example of message switching: A, B, ….F are the message switching nodes/telegraph offices.

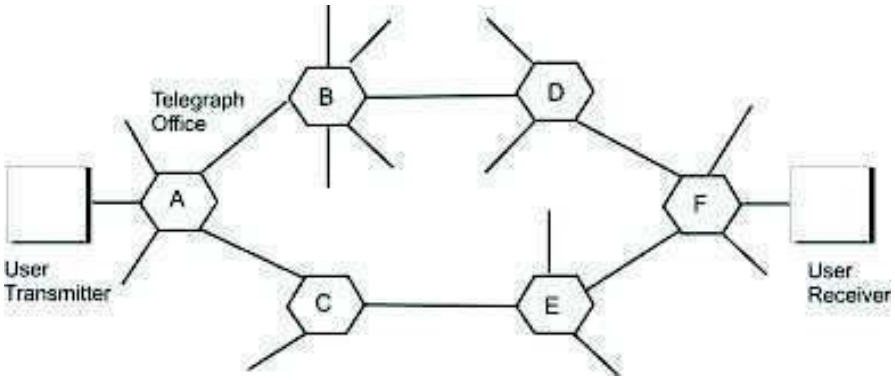


Figure 10: Working Model of the Telegraph Systems

The User who wants to send a telegraph comes to a Telegraph office with his message and hands it over to the counter operator. Now the following sequences of events occur:

- This message is sorted on the basis of the receiver's address and clubbed with other messages moving in the same direction, i.e., if in the Delhi's telegraph office the operator receives 10 messages for addresses in Mumbai, then they are bundled and are sent.
- The operator in this case does not bother if the entire path (to Mumbai) is available or not. He just forwards this message to the next node (Telegraph Office) in the path (generally predetermined).
- The operator at the next node receives all these messages, stores, sorts and forwards them.

In the olden days, the storage was done by manually. Human beings then did the sorting. Later on the storage process was automated using paper tapes. The advantage of using paper tapes is that the incoming signal is punched onto it automatically and the same tape can be directly fed into the telegraph machine for further transmission. In the Telegraph system, unlike telephones, no circuits are switched. Information is transmitted as discrete messages. So this method of switching is known as Message Switching. The important context is *'Store and Forward'*. At each node (telegraph office) the message that arrives from the previous node in the path is stored for some time, sorted, and depending on the availability of the path from this node to the next in the path, the message is forwarded.

There were central telegraph offices which acted like nodes of telegraph network and performed the task of message switching. as the teleprinters came, Morse code was replaced by machine telegraphy resulting in faster operations. Later computers were introduced to do the function of message switching. Computer based message switching is still used many organizations having many locations of working. However, if we compare the cost, the telegraph is less costly than the telephone due to the following reasons:

- Better utilization of transmission media
- The message switching is done over distributed time.
- Hogging (Capturing the entire path) does not occur in message switching. Only one of the links in the entire path may be busy at a given time.

However, message switching requires storage and this may raise-up the cost of the systems.

3.9 CIRCUIT SWITCHING

Circuit switching is defined as a mechanism applied in telecommunications (mainly in PSTN) whereby the user is allocated the full use of the communication channel for the duration of the call. That is if two parties wish to communicate, the calling party has to first dial the numbers of the called party. Once those numbers are dialed, the originating exchange will find a path to the terminating exchange, which will in turn find the called party. After the circuit or channel has been set up, then communication will take place, then once they are through the channel will be cleared. This mechanism is referred to as being connection-oriented.

Voice being a very vital medium of human communication, telephone was invented. It permitted long distance voice communication. The need of a user to talk to a desired person out of many persons on a real time basis leads to the concept of establishing a direct path between the caller and the called users. Circuit switching was conceived to be an appropriate technique for the purpose. Telephone systems use circuit switching largely to date because it serves the purpose very well. However, a major drawback of circuit switching is the requirement of a dedicated path between the calling and the called parties. This means reserving resources like the chain of switches and transmission media over the entire path. This is obviously a costly proposition. The circuit switching process has been illustrated in the Figure 11, for the telephone network. In which, the physical connections are made by the switching offices to connect the call of two users.

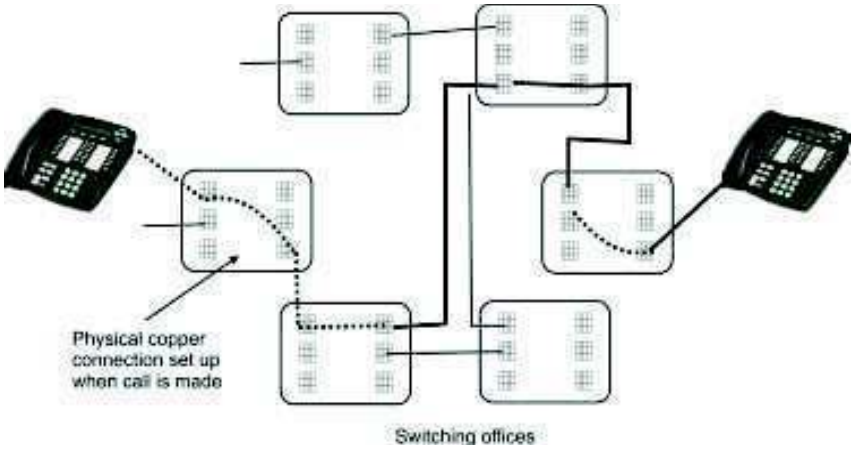


Figure 11: Circuit Switching in the Telephone Network

For each connection, physical switches are set in the telephone network to create a physical "circuit" – That's the job of the switching office Switches are set up at the beginning of the connection and maintained throughout the connection. Network resources reserved and dedicated from sender to receiver. However this is not a very efficient strategy as a connection "holds" a physical line even during "silence" periods (when there is nothing to transmit)

Advantages of Circuit Switching:

- Once the circuit has been set up, communication is fast and without error.
- It is highly reliable

Disadvantages:

- Involves a lot of overhead, during channel set up.
- Waists a lot of bandwidth, especial in speech whereby a user is sometimes listening, and not talking.
- Channel set up may take longer.

To overcome the disadvantages of circuit switching, packet switching was introduced, and instead of dedicating a channel to only two parties for the duration of the call it routes packets individually as they are available. This mechanism is referred to as being connectionless packet switching as discussed in the next section.

☞ Check Your Progress 2

1. Write differences between FDM and TDM.

   ……..

   ……..

   ……..

2. What is CDMA?

   ……..

   ……..

   ……..

   ……..

3. What is Circuit Switching?

   ……..

   ……..

   ……..

3.10 PACKET SWITCHING

Packet Switching is the backbone of the present day communication systems. The packet switching works on the principle that the long messages are fragmented into small size units, known as *packets*. It is these packets that are transmitted instead of the single long message. This method is slightly different from **Message switching** and is called **Packet switching**. Figure12 shows a message broken down into small sized packets P_1, P_2 …P5.

P1
P2
P3
P4
-
-
Pn

Figure 12: A Message broken into n number of packets

These packets are now transmitted over the network in the same manner as the messages in message switching. The model is just like Sharing by taking turn and is analogous to the conveyor belt in a warehouse. In this case, the Items are picked from the storage room and placed on the conveyor belt every time a customer makes an order. In this model, this is important that Different customers may request a different number of items and Different users' items may be interspersed on the conveyor belt (they are "multiplexed"). Similarly in the Packet Switching, packetizes the data to transfer and Multiplex it onto the wire. Thus packets from different connections share the same link

The packets are stored and forwarded at every node. Obviously every packet now has to have the source and destination addresses. Even in message switching repeated transmission of addresses at every node consumes network bandwidth. In packet switching the overhead/wastage is more because every packet is now required to carry the addresses on their head. Thus each packet is composed of the payload (the data we want to transmit) and a header. The header contains information useful for transmission, such as:

- Source (sender's) address
- Destination (recipient's) address
- Packet size
- Sequence number
- Error checking information

The header introduces overheads, that is, additional bits to be sent. Therefore, it is not wise to have packets that are too small. In the packet switching, each computer attached to a network is assigned a unique number (called address). A packet contains the address of the computer that sent it and the address of the computer to which it is sent. In general, packets need not be of the same size, The Internet Protocol specifies the maximum size in the form of Maximum transmission unit (MTU) and does not give the No minimum size. But, header size is fixed (e.g., 20 bytes for TCP/IP in the IP version 4). Packets are generated by the network hardware, however the application (e.g., email) does not know that the data to be transmitted is packetized. When packets are received, they are put together before the application accesses the data. The process is shown in the Figure 13 below, where A and B are the sender and C and D are the receiver.

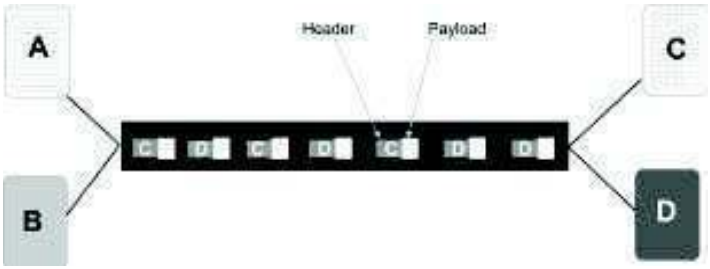


Figure13: Packet Switching Example

So with the user message in a packet with the header is to be transmitted also. From this point of view network bandwidth consumed is maximum in packet switching and minimum in circuit switching. Packets of the same message are launched into the network in parallel over different available forward links at a node. These packets would travel through different paths to arrive at the destination. This simultaneous transmission of packets over different paths results in further improvement of the link utilization compared to the message switching. Another advantage is that no link is engaged for a long time since the packets are of smaller size than the single message.

This permits better sharing of the links amongst multiple users. However the scheme just discussed has two major drawbacks. Firstly, the packets of the same message traveling through different paths may arrive at the destination at different times due to different delays encountered in different paths. Thus the packets may arrive out of order. In order to deliver them to the destination, they need to be ordered which requires extra processing and so more delay. They need to be given sequence numbers for reordering them. The sequence number increases the overhead and requires more network bandwidth. Secondly, some of the paths may not be very good and some packets may get lost. This worsens the quality. To improve quality, they require retransmission which in turn requires more processing time and more bandwidth. In spite of these drawbacks the packet switching is the most favored technique in the present day communication systems. The basic reasons behind this choice are:

a) Computer traffic being mostly text is non real time (in the beginning of the networking)

b) Computer data traffic is highly bursty in nature

Considering these features it becomes obvious that circuit switching was not the right kind of switching. Message switching can do the job but for better line utilization packet switching is preferable. Thus computer networks used packet switching. The difference between the packet switching and the circuit switching has been outlined in the Table 1.

Table 1: Difference between the packet switching and Circuit Switching

S.No.	Packet Switching	Circuit Switching
1	Bandwidth is allocated dynamically.	Fixed bandwidth allocation.
2	Packets has header, FCS.	Don't deal with data content and error-checking
3	Better buffering. System can be operated at different bit rate to inter-network.	Simple buffering
4	May be more economical as not needed dedicated circuit.	Costs more for hardware.
5	The packet needs to be re-transmitted every time when it gets lost, damaged before it is received in this method.	Once connection is established, communication is fast and almost errorless.
6.	Useful for bursty applications	Useful for delay sensitive applications

Categories of Packet Switching

The packet switching is basically, categorized in the following two categories:

a) Connection Less Packet Switching

b) Connection Oriented Packet Switching

3.10.1 Connection Less Packet Switching

In this mode of transmission, packets from a source machine to a destination machine are transmitted as per-packet basis, meaning that each packet is transmitted and routed independently from all other packets. So, even if the source and destination machines do not change, routers in the middle may decide to change the routes that different packets follow , resulting in the different packets reaching their destination in a different order from the sender because of the different transmission path length , difference in transmission rates, and the amount of congestion in the different paths. This is illustrated in the following Figure 14.

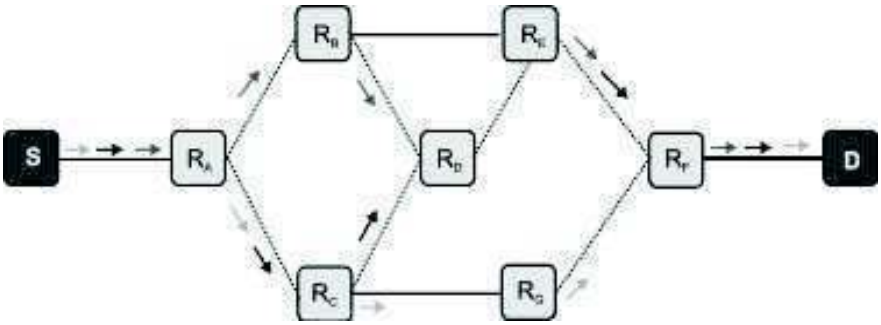


Figure 14: Connection less Packet Switching

In the figure 14, S denotes the source and D denotes the destination. R represents the router, whereas the packets have been shown by the arrow. Three packets are transmitted from the same source machine heading towards the same destination machine. Each route of the network shows the packets that have travelled over it. It is clear that the packets may arrive at the destination machine in an order different from the transmission order. Since the details of this routing table change with the movement of the packets, the routing of different packets often changes. The transmission process involves the following steps:

- Transmit Packet 1
- Transmit Packet 2
- ……
- ……..
- Transmit Packet N

Examples:

- POTS (Plane Old Telephone Systems)
- ATM (Asynchronous Transmission Mode)
- Frame Relay
- MPLS (Multi Protocol Label Switching)

Disadvantages of connectionless packet switching:

1. Extra processing power is required at the nodes for attaching source and destination addresses with every packet which also increases the required time of transmission.
2. Connectionless Packet switching requires overhead bits for indexing/numbering the packets.

3. Packets may arrive at the destination in a random manner. This requires that all the arriving packets are stored and rearranged.
4. Some packets may be lost in the network.

3.10.2 Connection Oriented Packet Switching

In this mode of transmission, packets from a source machine to a destination machine are moved as per the source destination pair basis, meaning that all packets from the same source going to the same destination are transmitted over the same routes and through the same routers. This results in having almost a constant delay of transmission for the different packets and the different packets reaching their destination in order

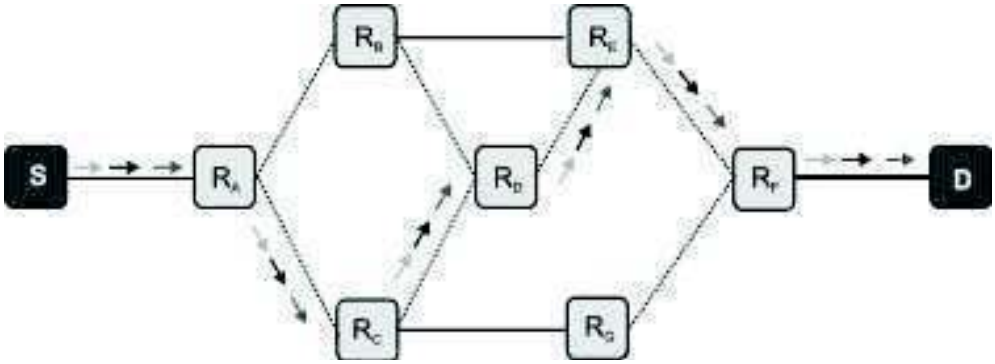


Figure 15: Connection Oriented Packet Switching

It is clear from the above Figure 15 that a circuit - like connection has been established. The process of transmission in the above case is also called Virtual - Circuit Packet Switching as it involves the establishment of a fixed path called Virtual Circuit or Virtual Connection between the source and destination prior to the transfer of packets. The transmission of packets involves the following steps:

1. Connection Request
2. Connection Confirm \
3. Transmit Packet 1
4. Transmit Packet 2
5. ……
6. ……
7. Transmit Packet N
8. Connection Release

Example:

ATM Networks

☞ Check Your Progress 3

1. Define the difference between switched and leased lines.

……..

……..

2. What are switched communications networks?

……..

……..

3. Discuss the advantages of packet switching over circuit switching.

……..

……..

3.11 SUMMARY

We hope you must have understood the concept of multiplexing and switching. As we discussed Multiplexing refers to the ability to transmit data coming from several pairs of equipment (transmitters and receivers) called *low-speed channels* on a single physical medium (called the *high-speed channel*). Whereas, A *multiplexer* is the multiplexing device that combines the signals from the different transmitters and sends them over the *high-speed channel*. Further in this unit you have studied four basic multiplexing techniques are frequency division multiplexing (FDM), Time division Multiplexing (TDM), Code division Multiplexing (CDM) and Space-division Multiplexing (SDM). As you have studied that Switching plays a very important role in telecommunication networks. It enables any two users to communicate with each other. Basically, there are three categories of Switching like Message Switching, Circuit Switching and Packet Switching.

3.12 REFERENCES/FURTHER READING

1. *Computer Networks*, A. S. *Tanenbaum* 4th Edition, Practice Hall of India, New Delhi. 2003.

2. *Introduction to Data Communication & Networking,* 3rd Edition, Behrouz Forouzan, Tata McGraw Hill.

3. *Computer Networking*, J.F. Kurose & K.W. Ross, A Top Down Approach Featuring the Internet, Pearson Edition, 2003.

4. *Communications Networks*, Leon Garcia, and Widjaja, Tata McGraw Hill, 2000.

5. www.wikipedia.org

6. *Data and Computer Communications,* Willian Stallings, 6th Edition, Pearson Education, New Delhi.

7. Larry L. Peterson, *Computer Networks*: A Systems Approach, 3rd Edition (The Morgan Kaufmann Series in Networking).

3.13 SOLUTIONS/ANSWERS

☞ Check Your Progress 1

1. Multiplexing is the set of techniques that allows the simultaneous transmission of multiple signals across a single data link.

2. To make efficient use of high speed telecommunications lines, some form of multiplexing is used. Multiplexing allows several transmission sources to share a larger transmission capacity.

A common application of multiplexing is in long-haul communications. Trunks on long-haul networks are high capacity fiber, coaxial or microwave links. These links can carry large numbers of voice and data transmission simultaneously using multiplexing.

3. Four basic multiplexing techniques are frequency division multiplexing (FDM), Time division Multiplexing (TDM), Code division Multiplexing (CDM) and Space-division Multiplexing (SDM).

☞ Check Your Progress 2

1. Frequency-Division Multiplexing (FDM) is a form of signal multiplexing where multiple baseband signals are modulated on different frequency carrier waves and added together to create a composite signal.

   Time-Division Multiplexing (TDM) is a type of digital multiplexing in which two or more signals or bit streams are combined into different slots of a frame. Transmission of frame carries simultaneously data from sub-channels in one communication channel, but are physically taking turns on the channel.

2. What is CDMA?

3. Circuit switching is defined as a mechanism applied in telecommunications hereby the user is allocated the full use of the communication channel for the duration of the call and hence a physical connection is set-up between the caller and the receiver.

☞ Check Your Progress 3

1. In switched line communications, a link that is established in a switched network, such as the international dial-up telephone system.

   A leased line is a symmetric dedicated service (the same upstream and downstream bandwidth) creating a permanent connection between your premises and the Internet.

2. In the switched communications networks data entering the network from a station are routed to the destination by being switched from node to node. For example in the Figure 16 data from station A intended for station F are send to node 4. They may then be routed via nodes 5 and 6 or nodes 7 and 6 to the destination. This is called switched communication networks.

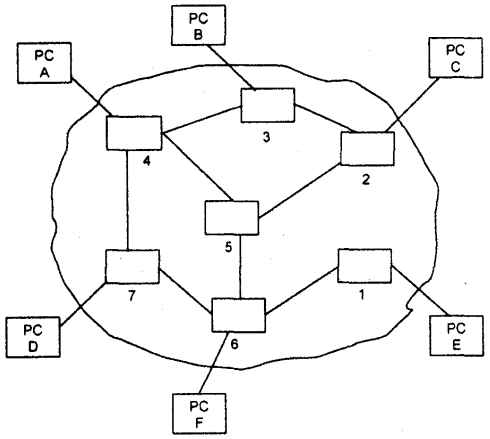


Figure 16: Simple Switching Network

3. i) Line efficiency is greater, because single node to node link can be dynamically shared by many packets over time. in other hand in circuit switching time on a node to node link is pre-allocated using synchronous time division multiplexing.

   ii) A packet switching network can perform data rate conversion.

   iii) When traffic becomes heavy on a circuit switching network, some caller are blocked, on the packet switching network, packets are still accepted, hut delivery delay increases.

   iv) Priorities can be used. Thus it can transmit higher priority packet first.

UNIT 4 COMMUNICATION MEDIUMS

Structure **Page No.**

4.0	Digital Data Transmission	60
4.1	Objectives	60
4.2	Serial and Parallel Transmission	61
4.3	Guided and Unguided Mediums	61
4.4	Twisted Pair	62
4.5	UTP Cable	63
4.6	STP Cable	63
4.7	Coaxial Cable	64
4.8	Fiber Optic Cables	65
4.9	Unguided Mediums	67
4.10	Connectors	69
4.11	Summary	71
4.12	References/Further Reading	71
4.13	Solutions/Answers	72

4.0 DIGITAL DATA TRANSMISSION

The term digital refers to the way it is conveyed: usually by a binary code consisting of a long string of 1s and 0s. Digital transmission or digital communications is a literal transfer of data over a point to point (or point to multipoint) link using transmission medium –such as copper wires, optical fibers, wireless communications media, or storage media. The data that is to be transferred is often represented as an electro-magnetic signal (such as a microwave). Digital transmission transfers messages discretely. These messages are represented by a sequence of pulses via a line code. Digital data transmission can occur in two basic modes: serial or parallel. The serial and parallel transmission is shown in Figure 1 below. Data within a computer system is transmitted via parallel mode on buses with the width of the parallel bus matched to the word size of the computer system. Data between computer systems is usually transmitted in bit serial mode. Consequently, it is necessary to make a parallel-to-serial conversion at a computer interface when sending data from a computer system into a network and a serial-to-parallel conversion at a computer interface when receiving information from a network. The type of transmission mode used may also depend upon distance and required data rate.

4.1 OBJECTIVES

After going through this unit, you should be able to:

- Know the concept of communication mediums
- Differentiate between Serial and Parallel Transmission
- Differentiate between Guided and Unguided Mediums
- Know the features and limitations of different wired mediums
- Understands the use of Twisted Pair, Coaxial and Fiber Optic Cables
- Know the functions of Unguided Mediums
- Understand the use of different connectors

4.2 SERIAL AND PARALLEL TRANSMISSION

Serial Transmission: In serial transmission, bits are sent sequentially on the same channel (wire) as shown in Figure 1, which reduces costs for wire but also slows the speed of transmission. Also, for serial transmission, some overhead time is needed since bits must be assembled and sent as a unit and then disassembled at the receiver. Serial transmission can be either synchronous or asynchronous. In synchronous transmission, groups of bits are combined into frames and frames are sent continuously with or without data to be transmitted. In asynchronous transmission, groups of bits are sent as independent units with start/stop flags and no data link synchronization, to allow for arbitrary size gaps between frames. However, start/stop bits maintain physical bit level synchronization once detected.

In parallel transmission, multiple bits (usually 8 bits or a byte/character) are sent simultaneously on different channels (wires, frequency channels) within the same cable as shown in Figure 1, or radio path, and synchronized to a clock. Parallel devices have a wider data bus than serial devices and can therefore, transfer data in words of one or more bytes at a time.

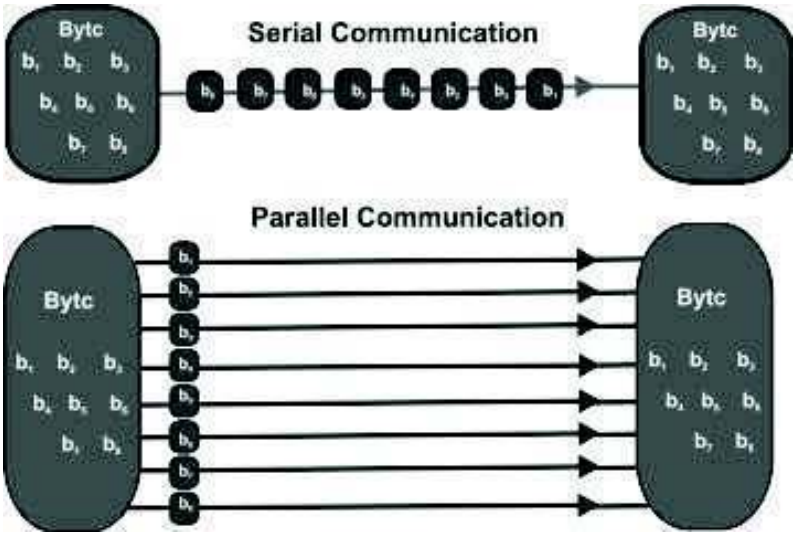


Figure 1: Serial and parallel communication

As a result, there is a speedup in parallel transmission bit rate over serial transmission bit rate. However, this speedup is a tradeoff versus cost since multiple wires cost more than a single wire, and as a parallel cable gets longer, the synchronization timing between multiple channels becomes more sensitive to distance. The timing for parallel transmission is provided by a constant clocking signal sent over a separate wire within the parallel cable; thus parallel transmission is considered synchronous.

4.3 GUIDED AND UNGUIDED MEDIUMS

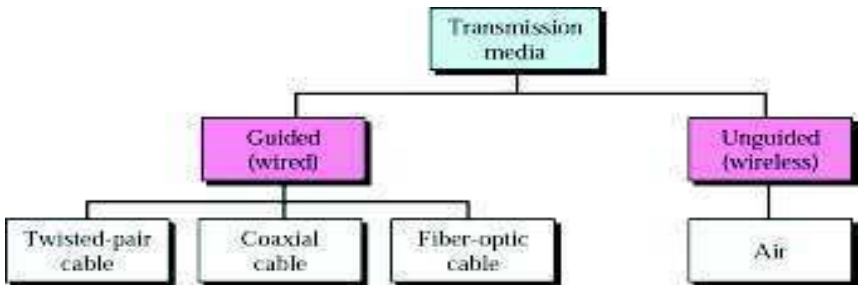


Figure 2: Classification of Transmission Mediums

Transmission Media: The transmission medium is the physical path between transmitter and receiver in a data transmission system. Transmission media can be classified as guided or unguided as depicted in Figure 2. With guided media, the waves are guided along a solid medium, such as twisted pair, coaxial cable, and optical fiber. The atmosphere and outer space are examples of unguided media that provide a means of transmitting electromagnetic signals but do not guide them; this form of transmission is usually referred to as wireless transmission.

The characteristics and quality of a data transmission are determined both by the characteristics of the medium and the characteristics of the signal. In the case of guided media, the medium itself is more important in determining the limitations of transmission.

For unguided media, the bandwidth of the signal produced by the transmitting antenna is more important than the medium in determining transmission characteristics. One key property of signals transmitted by antenna is directionality. In general, signals at lower frequencies are Omni-directional; that is, the signal propagates in all directions from the antenna. At higher frequencies, it is possible to focus the signal into a directional beam.

4.4 TWIATED PAIR

Twisted pair is most widely used media for local data distribution. Twisted-pair cable is a type of cabling that is used for telephone communications and most modern Ethernet networks. A pair of wires forms a circuit that can transmit data. The pairs are twisted to provide protection against crosstalk, and noise generated by adjacent pairs. When electrical current flows through a wire, it creates a small, circular magnetic field around the wire. When two wires in an electrical circuit are placed close together, their magnetic fields are the exact opposite of each other. Thus, the two magnetic fields cancel each other out. They also cancel out any outside magnetic fields. Twisting the wires can enhance this cancellation effect. Using cancellation together with twisting the wires, cable designers can effectively provide self shielding for wire pairs within the network media. The twisted pair cable is shown in Figure 3.

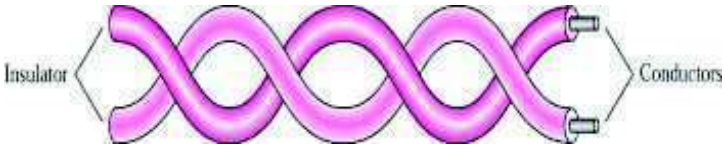


Figure 3: Twisted pair Cable

While twisted-pair cable is used by older telephone networks and is the least expensive type of local-area network (LAN) cable, most networks contain some twisted-pair cabling at some point along the network.

Since some telephone sets or desktop locations require multiple connections, twisted pair is sometimes installed in two or more pairs, all within a single cable. For some business locations, twisted pair is enclosed in a shield that functions as a ground. This is known as shielded twisted pair (STP). Ordinary wire to the home is unshielded twisted pair (UTP).

4.5 UTP CABLE

Unshielded twisted pair is the most common kind of copper telephone wiring. UTP cable is a medium that is composed of pairs of wires. UTP cable is used in a variety of networks. Each of the eight individual copper wires in UTP cable is covered by an insulating material. In addition, the wires in each pair are twisted around each other as shown in Figure 4 (a).

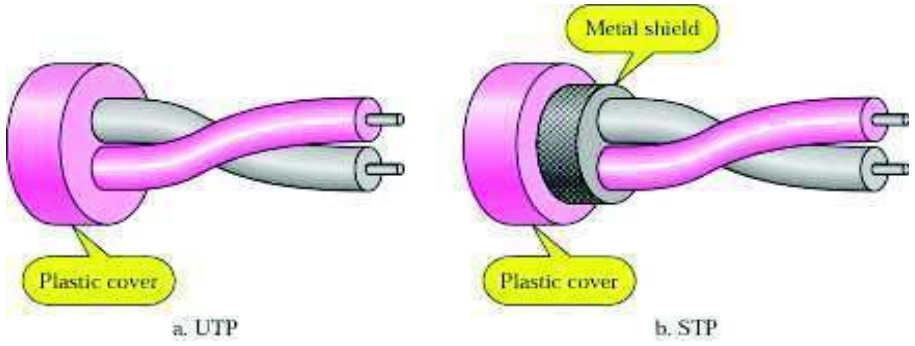


Figure 4: UTP and STP Cables

UTP cable relies solely on the cancellation effect produced by the twisted wire pairs to limit signal degradation caused by electromagnetic interference (EM!) and radio frequency interference (RFI). To further reduce crosstalk between the pairs in UTP cable, the number of twists in the wire pairs varies. UTP cable must follow precise specifications governing how many twists or braids are permitted per meter (3.28 feet) of cable.

4.6 STP CABLE

STP is similar to UTP in that the wire pairs are twisted around each other. STP also has shielding around the cable to further protect it from external interference. The shielding further reduces the chance of crosstalk but the shielding increases the overall diameter and weight of the cable. The maximum segment length of STP cable is 100 meters.

Shielded twisted pair is a special kind of copper telephone wiring used in some business installations. An outer covering or shield is added to the ordinary twisted pair telephone wires; the shield functions as a ground. The STP cable is shown in figure above in Figure 4(b).

Shielded twisted-pair (STP) cable combines the techniques of shielding, cancellation, and wire twisting. Each pair of wires is wrapped in a metallic foil. The four pairs of wires then are wrapped in an overall metallic braid or foil. It is usually a 150-ohm cable, as specified for use in Ethernet network installations. STP reduces electrical noise both within the cable (pair-to-pair coupling, or crosstalk) and from outside the cable (EMI and RFI).

☞ Check Your Progress 1

1. Define parallel transmission.

……

……

……

……

2. List guided transmission mediums?

……

……

……

……

3. What are the advantages of STP over UTP?

……

……

……

4.7 COAXIAL CABLE

Coaxial cable like twisted pair, consists of two conductors, but is constructed differently to permit it to operate over a wider range of frequencies. It consists of a hollow outer cylindrical conductor that surrounds a single inner wire conductor. The inner conductor is held in place by either regularly spaced insulating rings or a solid dielectric material. The outer conductor is covered with a jacket or shield. A single coaxial cable has a diameter of from 0.4 to about 1 inch. Because of its shielding, concentric construction, coaxial cable is much less susceptible to interference and cross-talk than is twisted pair. Coaxial cable can be used over longer distances and supports more stations on a shared line than twisted pair.

Coaxial cable is perhaps the most versatile transmission medium and has widespread use in a wide variety of applications; the most important of these are

i) Television distribution

ii) Long-distance telephone transmission

iii) Short-run computer system links

iv) Local Area Networks

Coaxial cable is spreading rapidly as a means of distributing TV signals to individual homes - cable TV. A cable TV system can carry dozens or even hundreds of TV channels ranging up to a few tens of miles.

Coaxial cable has traditionally been an important part of the long-distance telephone network. Today, it is getting replaced by optical fiber, terrestrial microwave, and satellite. Using frequency-division multiplexing, a coaxial cable can carry over 10,000 voice channels simultaneously. Coaxial cable is also commonly used for short-range connections between devices. Using digital signaling, coaxial cable can be used to provide high-speed I/O channels on computer systems. A co-axial cable is shown in Figure 5 below.

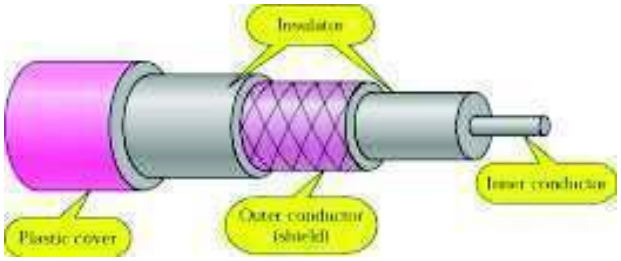


Figure 5: Coaxial cable

Another application area for coaxial cable is local area networks. Coaxial cable can support a large number of devices with a variety of data and traffic types, over distances that encompass a single building or a complex of buildings.

Coaxial cable is used to transmit both analog and digital signals. Coaxial cable has frequency characteristics that are superior to those of twisted pair, and can hence be used effectively at higher frequencies and data rates. The principal constraints on performance are attenuation, thermal noise, and inter modulation noise.

For long-distance transmission of analog signals, amplifiers are needed every few kilometers, with closer spacing required if higher frequencies are used. The usable spectrum for analog signaling extends to about 400 MHz. For digital signaling, repeaters are needed every kilometer or so, with closer spacing needed for higher data rates.

4.8 FIBRE OPTIC CABLES

Now day's optical fiber is widely used as a back bone for network due to its higher data transmission rate, lighter in weight, low interferences, less number of repeaters required, long distance coverage etc. An optical transmission system has three components; the light source, the transmission medium, and the detector.

Conventionally, a pulse of light indicates a bit 1 and absence of light indicates bit 0. Transmission medium is an ultra-thin fiber of glass. The transmitter generates the light pulses based on the input electrical signal. The detector regenerates the electrical signal based on the light signal it detects on the transmission medium. By attaching a light source to one end of an optical fiber and a detector to the other, we have an unidirectional data transmission system that accepts an electrical signal, converts and transmits it by light pulse, and then reconverts the output to an electrical signal at the receiving end. Figure 6 given blow shows optical fiber cable.

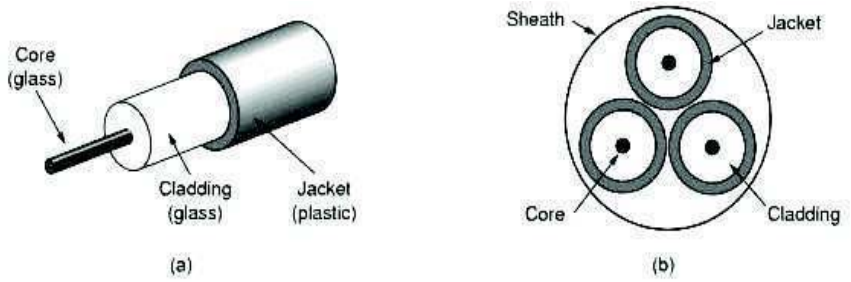


Figure 6: Optical Fiber Cable

An optical fiber is a thin (2 to 125 nm – nano meter – 10-9 meter), flexible medium capable of conducting an optical ray. Various glasses and plastics can be used to make optical fibers. The lowest losses have been obtained using fibers of ultrapure fused silica. Ultrapure fiber is difficult to manufacture; higher-loss multi-component glass fibers are more economical and still provide good performance. Plastic fiber is even less costly and can be used for short-haul links, for which moderately high losses are acceptable.

An optical fiber cable has a cylindrical shape and consists of three concentric sections: the core, the cladding, and the jacket. The core is the innermost section and consists of one or more very thin strands, or fibers, made of glass or plastic. Each fiber is surrounded by its own cladding, a glass or plastic coating that has optical properties different from those of the core. The outermost layer, surrounding one or a bundle of cladded fibers, is the jacket. The jacket is composed of plastic and other

material layered to protect against moisture, abrasion, crushing and other environmental dangers.

One of the most significant technological breakthroughs in data transmission has been the development of practical fiber optic communications systems. Optical fiber already enjoys considerable use in long-distance telecommunications. The continuing improvements in performance and decline in prices, together with the inherent advantages of optical fiber, have made it increasingly attractive for local area networking and metropolitan networks. Optical fiber is of two types.

i) Single mode optical fiber.

ii) Multimode Optical Fiber.

Single mode optical fiber: Single mode uses step-index fiber and a highly focused source of light that limits beams to a small range of angles, all close to the horizontal. The fiber itself is manufactured with a much smaller diameter than that of multimode fibers, and with substantially lowers density (index of refraction). The decrease in density results in a critical angle that is close enough to 90 degrees to make the propagation of beams delays are negligible. All of the beams arrive at the destination "together" and can be recombined without distortion to the signal as depicted in Figure 7 (c).

Multi-Mode: Multimode is so named because multiple beams from a light source move through the core in different paths. How these beams move within a cable depends on the structure of the core. Multi-mode is categorized into step-index multimode and graded index mode.

1. **Step-index Mult-mode:** In step-index multimode, the density of the core remains constant from the center to the edges. A beam of light moves through this constant density in a straight line until it reaches the interface of the core and the cladding. At the interface there is an abrupt change to a lower density that alters the angle of the beam's motion. The term step-index refers to the suddenness of this change. Figure 7 below shows various beams (or rays) traveling through a step-index fiber. Some beams in the middle travel in straight lines through the core and reach the destination without reflecting or refracting. Some beams strike the interface of the core and cladding at an angle smaller than the critical angle; these beams penetrate the cladding and are lost. Still others hit the edge of the core at angles greater than the critical angle and reflect back into the core and off the other side, bouncing back and forth down the channel until they reach the destination.

   Every beam reflects off the interface at an angle equal to its angle of incidence as shown in Figure 7(a). The greater the angle of incidence, the wider the angle of refraction. A beam with a smaller angle of incidence will require more bounces to travel the same distance than a beam with a larger angle of incidence. Consequently, the beam with the smaller incident angle must travel farther to reach the destination. This difference in path length means that different beams arrive at the destination at different times. As these different beams are recombined at the receiver, they result in a signal that is no longer an exact replica of the signal that was transmitted. Such a signal has been distorted by propagation delays. This distortion limits the available data rate and makes multimode step-index cable inadequate for certain precise applications.

2. **Graded-index Mode:** A second type of fiber, called graded-index, decreases this distortion of the signal through the cable. The word index here refers to the index of refraction. As we saw above, index of refraction is related to density. A graded-index fiber, therefore, is one with varying densities. Density is highest at

the center of the core and decreases gradually to its lowest at the edge. Figure 7(b) shows the impact of this variable density on the propagation of light beams.

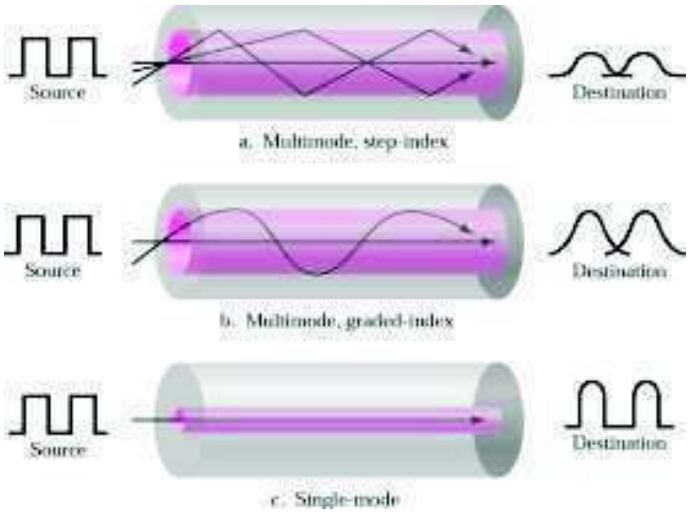


Figure 7: Types of Optical Fiber Cables

☞ Check Your Progress 2

1. List the applications of Coaxial cable.

……

……

……

……

2. What is Single mode optical fiber?

……

……

……

……

4.9 UNGUIDED MEDIUMS

Unguided Media: Unguided media transport electromagnetic waves without using a physical conductor. Signals are broadcast though air or water, and thus are available to anyone who has a device capable of receiving them. The EM spectrum covers frequencies from 3 Hz (ELF) to gamma rays (30 ZHz, Zetta Hertz - 10^{21} Hz) and beyond (cosmic rays). But only frequencies ranging from 3 KHz to 900 THz are used for wireless communication.

Propagation of Radio Waves: Radio technology considers the earth as surrounded by two layers of atmosphere: the troposphere and the ionosphere. The troposphere is the portion of the atmosphere extending outward approximately 30 miles from the earth's surface. The troposphere contains what we generally think of as air. Clouds, wind, temperature variations, and weather in general occur in the troposphere. The ionosphere is the layer of the atmosphere above the troposphere but below space. Unguided signals can travel from the source to destination in several ways. There is

ground propagation, sky propagation, and line-of-sight propagation. In ground propagation, radio waves travel through the lowest portion of the atmosphere, hugging the earth. These low-frequency signals emanate in all directions from the transmitting antenna and follow the curvature of the earth. The distance depends on the power of the signal. In Sky propagation, higher-frequency radio waves radiate upward into the ionosphere where they are reflected back to earth. This type of transmission allows for greater distances with lower power output. In Line-of-Sight Propagation, very high frequency signals are transmitted in straight lines directly from antenna to antenna. Antennas must be directional, facing each other and either tall enough or close enough together not to be affected by the curvature of the earth.

Radio Waves: Radio wave frequencies are between 3 KHz to 1 GHz, and uses omnidirectional antenna. Omniderectional antenna propagates signal in all direction. This means that the sending and receiving antennas do not have to be aligned. But it has disadvantage too, it is susceptible to interference wherein a radio wave transmitted by one antenna may be interfered by another antenna that may send signals using the same frequency or band.

Radio waves are used for multicast communications, such as radio (AM and FM radio), maritime radio, television, cordless phones and paging systems.

Microwaves: Frequencies between 1 and 300 GHz are called microwaves. Microwaves are unidirectional. When an antenna transmits microwave waves, they can be narrowly focused. This means that the sending and receiving antennas need to be aligned. Its advantage is that a pair of antennas can be aligned without interfering with another pair of aligned antennas.

The propagation of microwave is line-of-sight. The problem with this propagation is that towers that are far apart from each other need to be very tall. The curvature of the earth as well as other blocking obstacles does not allow two short towers to communicate. For long distance communication, repeaters are often needed. Another disadvantage is that very high frequency microwaves cannot penetrate walls.

In a unidirectional antenna, there are two types: the parabolic dish and the horn. A parabolic dish antenna is based on the geometry of the parabola. Every line parallel to the line of symmetry reflects off the curve at angles such that all the lines intersect in a common point called focus. The parabolic dish works as a funnel, catching a wide range of waves and directing them to a common point.

A horn antenna on the other hand looks like a gigantic scoop. Outgoing transmissions are broadcast up a stem and deflected outward in a series of narrow parallel beams by the curved head. Received transmissions are collected by the scooped shape of the horn, in a manner similar to the parabolic dish, and are deflected down into the stem.

There is another type of microwave transmission with the use of satellite relay. It requires geo-stationary orbit with the height of 35,784km to match the earth's rotation. It has uplink that receives transmission on one frequency and a downlink that transmits on a second frequency. It Operates on a number of frequency bands known as transponders.

It can operate in two ways:

a) Point to point- Ground station to satellite to ground station

b) Multipoint (Broadcast link)- Ground station to satellite to multiple receiving stations.

Microwaves are used in unicast communication such as cellular telephones, satellite networks, and wireless LANs.

Infrared Waves: Infrared signals with frequencies from 300 GHz to 400 THz (wavelengths from 1 mm to 700 nm), can be used for short-range communication. high frequencies cannot penetrate walls. This characteristic prevents interference between one system and another; a short-range communication cannot be affected by another system in the next room. The same characteristic makes infrared signals useless for long range communication. Infrared waves cannot be used outside a building because the sun's rays contained infrared waves can interfere with the communication. The infrared band, almost 400 THz, has an excellent potential for data transmission. Such a wide bandwidth can be used to transmit digital data with a very high data rate. The infrared Data Association (IrDA), an association for sponsoring the use of infrared waves, has established a standard for using these signals for communication between devices such as the keyboard, mice, PCs, and printers. Infrared signals defined by the IrDA transmit through line of sight; the IrDA port on the keyboard needs to point to the PC for transmission occurs.

4.10 CONNECTORS

The connectors are the interface for communication between computers/ computers to hub, switch, router etc. In LAN basically used connector are discussed as follows:

1. **RJ-45 Connector:** RJ stands for registered jack. RJ45 is a standard type of connector for network cables. RJ45 connectors are most commonly seen with Ethernet cables and networks. RJ45 connectors feature eight pins to which the wire strands of a cable interface electrically. Standard RJ-45 pin-outs define the arrangement of the individual wires needed when attaching connectors to a cable. RJ-45 connectors are of two types: male RJ-45 and female RJ-45. The Figure 8 shows RJ -45 connector.

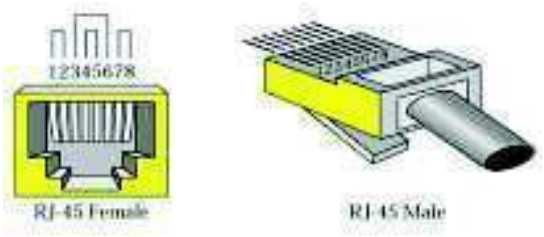


Figure 8: RJ-45 connectors

2. **BNC connector:** The BNC connector (Bayonet Neill–Concelman) is miniatures quick connect/disconnect RF connector used for coaxial cable. It features two bayonet lugs on the female connector; mating is achieved with only a quarter turn of the coupling nut. BNCs are ideally suited for cable termination for miniature-to-subminiature coaxial cable (e.g., RG-58, 59, to RG-179, RG-316). It is used with radio, television, and other radio-frequency electronic equipment, test instruments, video signals, and was once a popular computer network connector. BNC connectors are made to match the characteristic impedance of cable at either 50 ohms or 75 ohms. It is usually applied for frequencies below 3 GHz and voltages below 500 Volts.

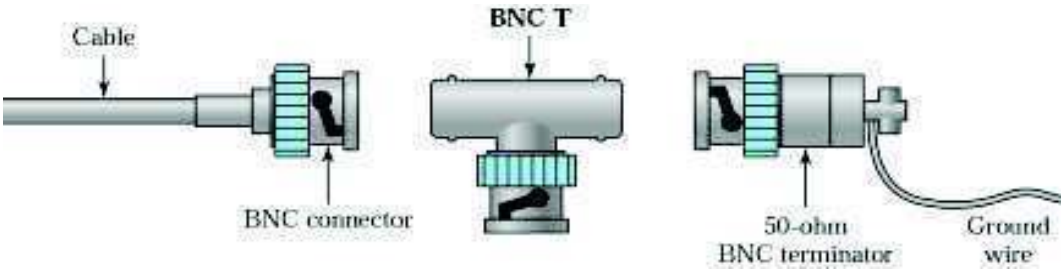


Figure 9: BNC connectors

Fiber optic cable connector: Fiber optic cable connectors are hardware installed on fiber cable ends to provide cable attachment to a transmitter, receiver or other cable. In order for information to be transmitted efficiently, the fiber cores must be properly aligned. They are usually devices that can be connected and disconnected repeatedly. There are many types of fiber optic cable connectors also shown in Figure 10:

1. **ST Connectors:** ST stands for Straight Tip. Slotted bayonet type connector with long ferrule, a common connector for multi-mode fibers. The ST connector has been the main stay of optical fiber connectors for many years. It can be found in almost every communications room worldwide, but used mainly in data communications systems. The simple to use bayonet locking mechanism reduces the risks of accidental disconnection of fiber connections.

2. **SC (Standard Connector) Connectors:** Push/pull connector that can also be used with duplex fiber connection. The SC connector comprises a polymer body with ceramic ferrule barrel assembly plus a crimp over sleeve and rubber boot. These connectors are suitable for, 900µm and 2-3mm cables. The connector is precision made to demanding specifications. The combination of a ceramic ferrule with precision polymer housing provides consistent long-term mechanical and optical performance.

3. **MT Connector:** The MT-RJ connector is a development of the now legendary MT ferrule. MT stands Multi-fiber Connector. The MT ferrule in its various designs has the ability to connect anything from 2 fibers in the MTRJ to 72 fibers in the latest versions of the MPO connector.

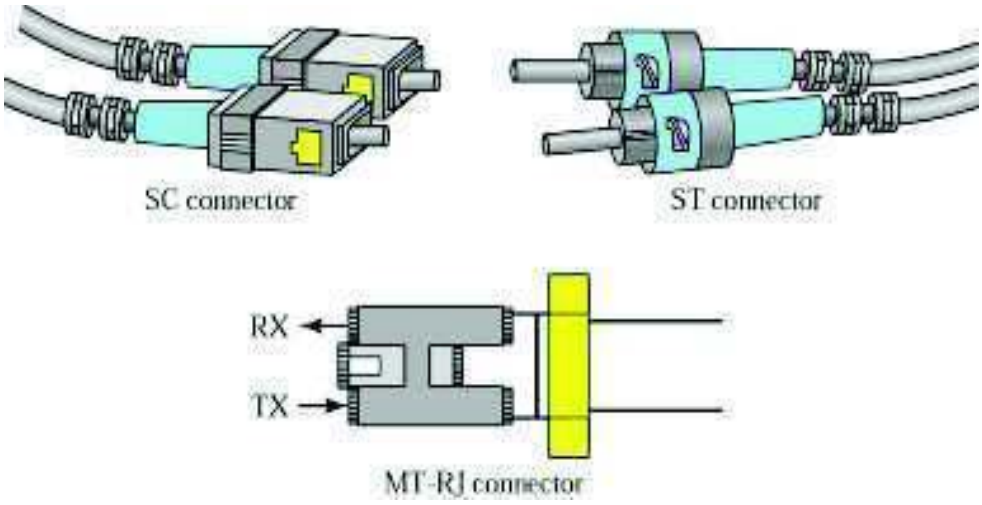


Figure 10: Fiber optic cable connector

☞ **Check Your Progress 3**

1. What are microwaves? Explain their properties.

   ……

   ……

   ……

2. What is BNC connector?

   ……

   ……

3. Explain the use of SC Connectors.

   ……

   ……

   ……

4.11 SUMMARY

After completing this unit, you must have knowledge of different transmission mediums, cables and connectors. In the beginning serial and parallel communication is explained. In serial transmission, bits are sent sequentially on the same channel (wire). In parallel transmission, multiple bits (usually 8 bits or a byte/character) are sent simultaneously on different channels (wires, frequency channels) within the same cable. In this unit, we have seen that transmission media can be classified as guided or unguided. Twisted-pair cable is a type of cabling that is used for telephone communications and most modern Ethernet networks. Coaxial cable like twisted pair, consists of two conductors, but is constructed differently to permit it to operate over a wider range of frequencies. Today's optical fiber is widely used as a back bone for network due to its higher data transmission rate, lighter in weight, low interferences, less number of repeaters required, long distance coverage etc.. Optical fiber is of two types i.e. Single mode optical fiber and Multimode Optical Fiber. Further medium of communication is unguided. Unguided media transport electromagnetic waves without using a physical conductor. Signals are broadcast though air or water, and thus are available to anyone who has a device capable of receiving them. The connectors are the interface for communication between computers/ computers to hub, switch, router etc. In LAN basically used connector.

4.12 REFERENCES/FURTHER READING

1. *Computer Networks*, A. S. *Tanenbaum* 4th Edition, Practice Hall of India, New Delhi. 2003.

2. *Introduction to Data Communication & Networking,* 3rd Edition, Behrouz Forouzan, Tata McGraw Hill.

3. *Computer Networking*, J.F. Kurose & K.W. Ross, A Top Down Approach Featuring the Internet, Pearson Edition, 2003.

4. *Communications Networks*, Leon Garcia, and Widjaja, Tata McGraw Hill, 2000.

5. www.wikipedia.org

6. *Data and Computer Communications,* Willian Stallings, 6th Edition, Pearson Education, New Delhi.

7. Larry L. Peterson, *Computer Networks*: A Systems Approach, 3rd Edition (The Morgan Kaufmann Series in Networking).

4.13 SOLUTIONS/ANSWERS

☞ Check Your Progress 1

1. In parallel transmission, multiple bits (usually 8 bits or a byte/character) are sent simultaneously on different channels (wires, frequency channels) within the same cable

2. Following are the guided transmission mediums
   i) twisted pair,
   ii) coaxial cable,
   iii) optical fiber

3. STP is similar to UTP in that the wire pairs are twisted around each other. STP also has shielding around the cable to further protect it from external interference. The maximum segment length of STP cable is 100 meters. Shielded twisted-pair (STP) cable combines the techniques of shielding, cancellation, and wire twisting. Each pair of wires is wrapped in a metallic foil.

☞ Check Your Progress 2

1. Following are the main applications of Coaxial cable.
   i) Television distribution
   ii) Long-distance telephone transmission
   iii) Short-run computer system links
   iv) Local Area Networks

2. Single mode uses step-index fiber and a highly focused source of light that limits beams to a small range of angles, all close to the horizontal. The fiber itself is manufactured with a much smaller diameter than that of multimode fibers, and with substantially lowers density (index of refraction). The decrease in density results in a critical angle that is close enough to 90 degrees to make the propagation of beams delays are negligible.

☞ Check Your Progress 3

1. Frequencies between 1 and 300 GHz are called microwaves. Microwaves are unidirectional. When an antenna transmits microwave waves, they can be narrowly focused. This means that the sending and receiving antennas need to be aligned. Its advantage is that a pair of antennas can be aligned without interfering with another pair of aligned antennas.

   The propagation of microwave is line-of-sight. The problem with this propagation is that towers that are far apart from each other need to be very tall. The curvature of the earth as well as other blocking obstacles does not allow two short towers to communicate. For long distance communication, repeaters are often needed. Another disadvantage is that very high frequency microwaves cannot penetrate walls.

2. The BNC connector (Bayonet Neill–Concelman) is miniatures quick connect/disconnect RF connector used for coaxial cable.
3. This is a fiber optics cable connector. Push/pull connector that can also be used with duplex fiber connection. The SC connector comprises a polymer body with ceramic ferrule barrel assembly plus a crimp over sleeve and rubber boot. These connectors are suitable for, 900μm and 2-3mm cables. The connector is precision made to demanding specifications.

UNIT 1 NETWORK CLASSIFICATIONS AND TOPOLOGIES

Structure **Page No.**

1.0	Introduction	5
1.1	Objectives	5
1.2	Network overview	5
	1.2.1 Classification of networks	
	1.2.2 Local area network (LAN)	
	1.2.3 Metropolitan area network (man)	
	1.2.4 Wide area network (wan)	
1.3	LAN Topologies	7
1.4	LAN /Mac Access Methods	12
1.5	Network Types Based on Size	15
1.6	Functional Classification of Networks	16
1.7	Wan Topologies	18
1.8	Wan Access Methods	18
1.9	Summary	20
1.10	References/Further Reading	20
1.11	Solutions/Answers	20

1.0 INTRODUCTION

As you know that a computer network is a group of computers that are connected with each other using some media for sharing of data and resources. It may connect other devices also like printers, scanners, etc. Information travels over the cables or other media, allowing network users to exchange documents & data with each other, print the data, and generally share any hardware or software that is connected to the network. In this unit we will learn about the different types of networks, their classifications based on topologies, size and functioning. We will also examine the access methods for LAN and WAN.

1.1 OBJECTIVES

After going through this unit, you should be able to:

- Define and classify network**;**
- distinguish between different types of networks,
- differentiate between different network (LAN and WAN) topologies
- understand LAN and WAN access methods

1.2 NETWORK OVERVIEW

In the simplest form, data transfer can take place between two devices which are directly connected by some form of communication medium. But it is not practical for two devices to be directly point to point connected. This is due to the following reasons:

i) The devices are situated at remote places.

ii) There is a set of devices, each of whom may require to connect to others at various times.

Solution to this problem is to connect each device to a communication network. Computer Networks means interconnected set of autonomous systems that permit distributed processing of information.

In order to meet the needs of various applications, networks are available with different interconnection layouts and plans, methods of access, protocols and data carrying capacities. Networks can be classified on the basis of geographical coverage.

1.2.1 Classification of Networks

- Local Area Network (LAN)
- Metropolitan Area Network (MAN)
- Wide Area Network (WAN).

1.2.2 Local Area Network (LAN)

A local area network is relatively smaller and privately owned network with the maximum span of 10 km. to provide local connectivity within a building or small geographical area. The LANs are distinguished from other kinds of networks by three characteristics:

i) Size (coverage area)

ii) Transmission technology (coverage area), and

iii) Topology.

1.2.3 Metropolitan Area Network (MAN)

Metropolitan Area Network is defined for less than 50 km. and provides regional connectivity typically within small geographical area. It is designed to extend over an entire city. It may be a single network such as cable television, network, or it may be a means of connecting a number of LANs into a large network, so that resources may be shared LAN to LAN as well as device to device. For example, a company can use a MAN to connect to the LANs in all of its offices throughout a city.

1.2.4 Wide Area Network (WAN)

Wide Area Network provides no limit of distance. In most WANs, the subnet consists of two distinct components. Transmission lines are also called circuits or channels or links and switching and routing devices (switches & routers). Transmission-lines are used for moving bits between machines, whereas routers are used to connect two or more transmission lines.

A WAN provides long distance transmission of data, voice, image and video information over large geographical areas that may comprise a country, a continent or even the whole world.

In contrast to LANs (which depend on their own hardware for transmission), WANs may utilise public, leased or private communication devices usually in combination and span own unlimited number of miles.

A WAN that is wholly owned by a single company is often referred to as an enterprise network.

A Local Area Network (LAN) is generally a privately owned network within a single office, building or campus, covering a distance of a few kilometers. The main reason for designing a LAN is to share resources such as disks, printers, programs and data. It also enables the exchange of information. Classically, LANs had data rates of 4-16 Megabits

per second (Mbps). Later, 100 Mbps LANs were introduced. Today, LANs with data rates of thousands of Mbps are possible. LANs typically can use the star, bus or a ring topology. However, bus topology is popular in the Ethernet LANs and Token Bus LANs and ring topology is popular in the Token Ring LANs of IBM. A modified version of Token Ring is Fiber Distributed Data Interface (FDDI). Of these, Ethernet and Token Ring are the most popular LANs.

☞ **Check Your Progress 1**

1. What are various types of networks?

   ……

   ……

   ……

2. What is the difference between Broadcasting and Multicasting?

   ……

   ……

   ……

1.3 LAN TOPOLOGIES

A topology is a generalized geometric configuration of some class of objects that join together. Topologies are the architectural "drawings" that show the overall physical configuration for a given communications system.

In networking, the term topology refers to the layout of connected devices on a network. It can be considered as the logical "shape" of the network wiring. This logical shape does not necessarily correspond to the actual physical layout of the devices on the network. For example, the computers on a home LAN may be arranged in a circle, but it would be highly unlikely to find an actual ring topology there 'Logical' means how it looks as a pure design concept, rather than how it actually looks physically.

Topology indicates the access methods and governs the rules that are used to design and implement the communication system. It is important to make a distinction between a topology and architecture. A topology is concerned with the physical arrangement of the network components. In contrast, architecture addresses the components themselves and how a system is structured (cable access methods, lower level protocols, topology, etc.). An example of architecture is 10baseT Ethernet that typically uses the star topology. Each topology has its advantages and disadvantages usually related to cost, complexity, reliability and traffic "bottlenecks". The different types of topologies are discussed below:

Bus Topology: --In a bus topology, all stations are attached to the same cable. In the Bus Network, messages are sent in both directions from a single point and are read by the node (computer or peripheral on the network) identified by the code with the message. Most Local Area Networks (LANs) are Bus Networks because the network will continue to function even if one computer is down. The purpose of the terminators (resistors) at either end of the network is to stop the signal being reflected back. If a bus network is not terminated, or if the terminator has the wrong level of resistance, each signal may travel across the bus several times instead of just once. This problem increases the number of signal collisions, degrading network performance. The figure 1 given below shows a bus Topology:

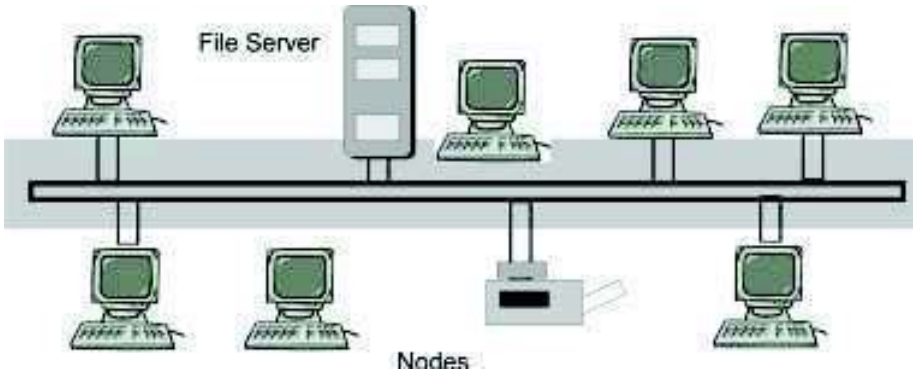


Figure 1: Bus Topology

In a bus topology, signals are broadcast to all stations. Each computer checks the address on the signal (data frame) as it passes along the bus. If the signal's address matches that of the computer, the computer processes the signal. If the address doesn't match, the computer takes no action and the signal travels down the bus.

Advantages of Bus Topology: The advantages of BUS topologies are as follows: --

i) Bus topologies are relatively easy to install and don't require much cabling compared to other topologies.

ii) Easy to connect a computer or peripheral to a linear bus.

iii) Requires less cable length than a star topology, as you only need to chain the stations together.

iv) There is no central point of failure on a bus because there is no switch. .

v) Simple and easy to implement and extend.

vi) Failure of one station does not affect others.

Disadvantages of a Linear Bus Topology: -- The disadvantages of BUS topologies are as follows: --

i) Entire network shuts down if there is a break in the main cable.

ii) Terminators are required at both ends of the backbone cable.

iii) Difficult to identify the problem if the entire network shuts down.

iv) Not meant to be used as a stand-alone solution in a large building.

v) Maintenance costs may be higher in the long run.

vi) More expensive cabling: Because the line is shared, the cable should have high bandwidth.

vii) Addition of nodes negatively affects the performance of the whole network, and if there is a lot of traffic throughput decreases rapidly.

viii) The more components share the signal, the more probable errors become. As the signal has to be multiplexed and demultiplexed and as every connected device is examining them. thus errors can more easily occur.

Star Topology: -- In a Star Network, all the nodes (PCs, printers and other shared peripherals) are connected to the central server. It has a central connection point - like a switch. A star topology is designed with each node (file server, workstations, and peripherals) connected directly to a central network hub or concentrator as shown in figure2 below.

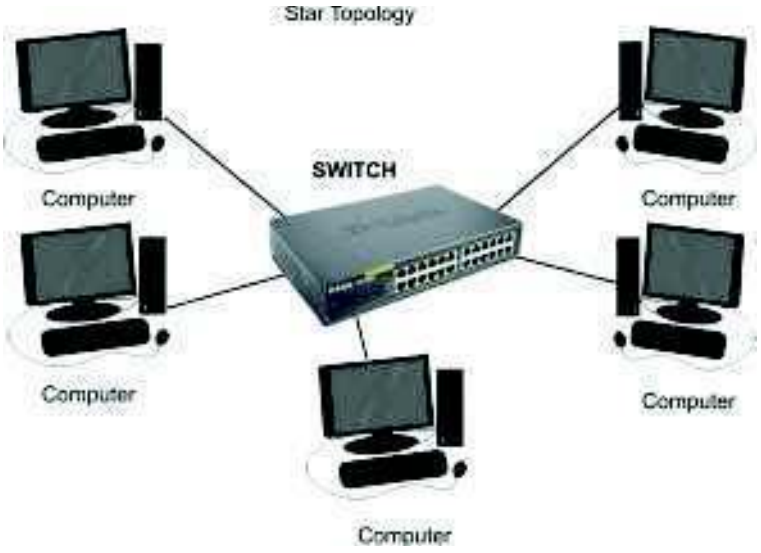


Figure 2: Star Topology

All traffic emanates from the switch of the star. Data on a star network passes through the switch or concentrator before continuing to its destination. The switch or concentrator manages and controls all functions of the network. It also acts as a repeater for the data flow. This configuration is common with twisted pair cable; however, it can also be used with coaxial cable or fiber optic cable. The switch offers a common connection for all stations on the network. Each station has its own direct cable connection to the switch.

Advantages of a Star Topology: -- The advantages of star topologies are as follows:

i) Easy to add new stations as each station has its own direct cable connection to the switch. If a cable is cut, it only affects the computer that was attached to it.

ii) It can accommodate different wiring. It can be installed using twisted pair, coaxial cable or fiber optic cable.

iii) Since all information in a star topology goes through a central point star, topologies are easy to troubleshoot. A star can simplify troubleshooting because stations can be disconnected from the switch one at a time until the problem is isolated.

iv) The main advantage is that one malfunctioning node does not affect the rest of the network.

Disadvantages of a Star Topology: --The advantages of star topologies are as follows:-

i) Depending on where the switches are located, star networks can require more cable length than a linear topology.

ii) If the switch / concentrator/switches fail, nodes attached are disabled.

iii) More expensive than linear bus topologies because of the cost of the switches.

Ring Topology: --All the nodes in a ring network are connected in a closed circle of cable as shown in figure 3. Messages that are transmitted travel around the ring until they reach the computer that they are addressed to. The signal being transmitted is refreshed by each node in the ring between the sender and receiver. In a ring network, every device has exactly two neighbors for communication purposes.

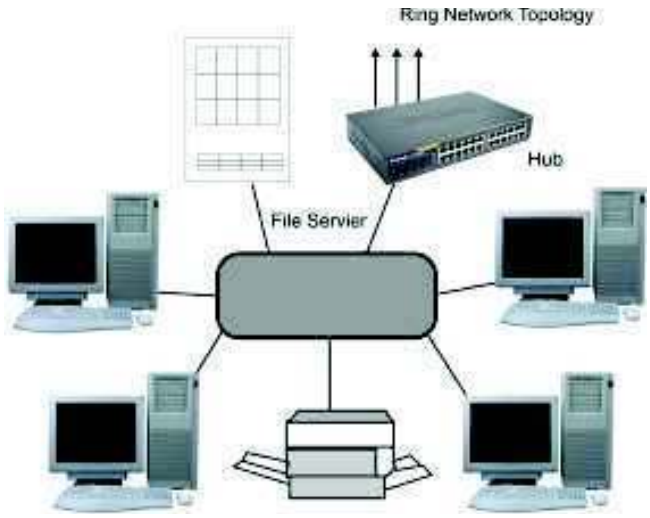


Figure 3: Ring Topology

All messages travel through a ring in the same direction. There are no terminated ends to the cable; the signal travels around the circle and terminated by the source.

Under the ring concept, a chance is given to each node sequentially via a "token" from one station to the next. When a station wants to transmit data, it "grabs" the token, attaches data and an address to it, and then sends it around the ring. The token travels along the ring until it reaches the destination. The receiving computer acknowledges receipt by stamping incoming message and passes it to the sender. The sender then releases the token to be used by another computer.

Each station in the ring has equal access but only one station can talk at a time. In contrast to the 'passive' topology of the bus, the ring employs an 'active' topology. Each station repeats or 'boosts' the signal before passing it on to the next station. Rings are normally implemented using twisted pair or fiber-optic cable.

Advantages of Ring Topology: -- The advantages of ring topologies are as follows: -

i) Growth of system has minimal impact on performance. The ring networks can be larger than bus or star because each node regenerates the signal.

ii) Degrade nicely under high utilization. Everybody gets to talk."

iii) Fault tolerance builds into the design (can bypass damaged nodes).

iv) Data packets travel at a greater speed.

Disadvantages of Ring Topology: -- The disadvantages of ring topologies are as follows: -

i) Expensive topology.

ii) Failure of one interface may impact others. A failure in any cable or device breaks the loop and will take down the entire segment.

iii) It is complex to implement and to extend the network; you must break the

iv) Ring (which brings the network down). If any device is added to or removed from the ring, the ring is broken and the segment fails.

Mesh Topology: -- In the topologies shown in figure 4, there is only one possible path from one node to another node. If any cable in the path is broken, the nodes cannot communicate. In a mesh topology, every device has a dedicated point-to point link to every other device. Such a network is called complete because between any two devices there is a special link; one could not add any non- redundant links.

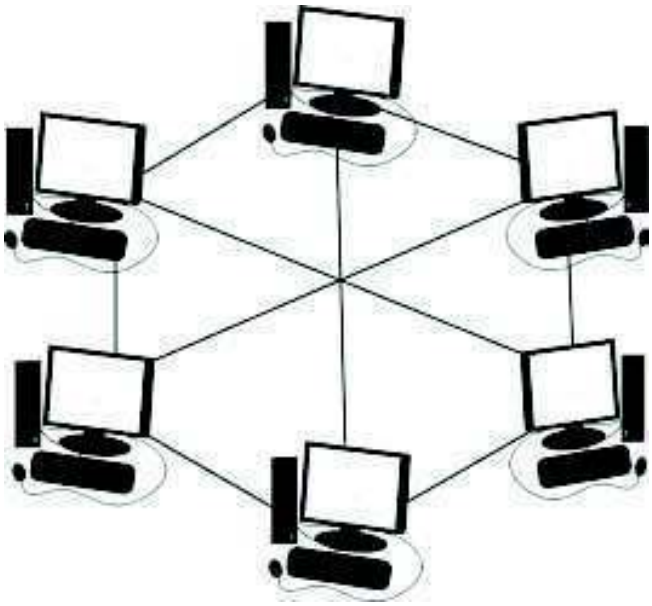


Figure 4: Mesh Topology

Mesh topology uses lots of cables to connect every node with every other node. It is very expensive to wire up, but if any cable fails, there are many other ways for two nodes to communicate. In mesh topology if we have to connect 'n' computers then we need n*(n-1)/2 cables/connections and each computer must have (n-1) Ethernet cards.

Advantages of Mesh Topology: -- The advantages of mesh topology are as follows:-

i) Redundant links between devices.

ii) Good security: If the line is not tapped only the intended recipient can see the data.

iii) Reliability: Increasing network traffic does not affect the speed of other connections.

iv) Easy fault identification and isolation, an unusable link does not incapacitate the entire system

Disadvantages of Mesh Topology: -- The disadvantages of mesh topology are as follows: -

i) Each node must have an interface for every other device.

ii) Large amounts of cable for many devices to be connected in a mesh environment. A mesh topology for n devices needs n (n - 1) connections. It is therefore hard to install and expensive because of the extensive cabling.

iii) Unless each station sends to every other station frequently, bandwidth is wasted. (Links that are not being used).

iv) Another disadvantage is that there is only limited of I/O-ports in a computer, but every connection needs one.

Tree Topology: -- The tree topology also known as the 'Hierarchical topology'. The tree topology is a combination of bus and star topologies. It consists of groups of star-configured workstations connected to' a linear bus backbone cable. Tree topologies allow for the expansion of an existing network and enable to configure a network to meet their needs. They are very common in larger networks. Figure 5 given below shows a typical tree topology.

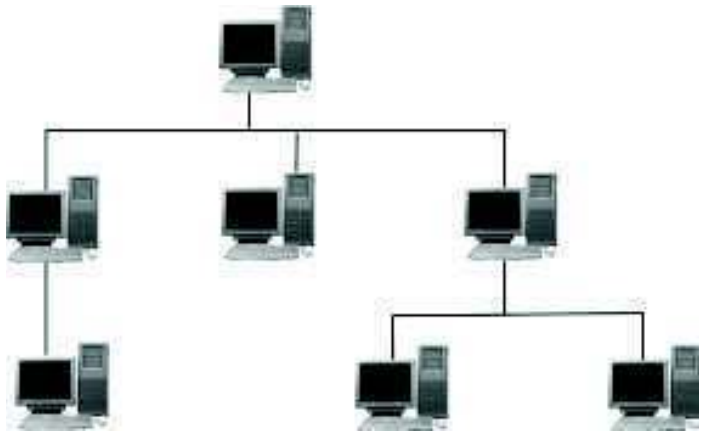


Figure 5: Tree Topology

For example, a file server is connected to a 24-port switch. A cable goes from the switch to a computer room where it connects to another switch. Many cables pass from this switch to the computers in the computer room. The node at the highest point in the hierarchy usually a file server-controls the network.

Advantages of a Tree Topology: -- The advantages of tree topology are as follows:-

i) Point-to-point wiring for individual segments.

ii) Supported by several hardware and software vendors.

Disadvantages of a Tree Topology: -- The disadvantages of tree topology are as follows:-

i) Overall length of each segment is limited by the type of cabling used.

ii) If the backbone line breaks, the entire segment goes down.

iii) More difficult to configure and wire than other topologies.

Considerations When Choosing a Topology

The considerations while choosing topologies are as follows: --

i) **Cost:** A linear bus network may be the least expensive way to install a network; you do not have to purchase concentrators

ii) **Length of cable needed:** The linear bus network uses shorter lengths of cable.

iii) **Future growth:** With a star topology, expanding a network is easily done by adding another switch.

iv) **Cable type:** The most common cable is unshielded twisted pair, which is most often used with bus, star topologies.

1.4 LAN /MAC ACCESS METHODS

Goals of MAC: Medium Access Control techniques are designed with the following goals in mind.

- **Initialisation:** The technique enables network stations, upon power-up, to enter the state required for operation.
- **Fairness:** The technique should treat each station fairly in terms of the time it is made to wait until it gains entry to the network, access time and the time it is allowed to spend for transmission.

- **Priority:** In managing access and communications time, the technique should be able to give priority to some stations over other stations to facilitate different type of services needed.
- **Limitations to one station:** The techniques should allow transmission by one station at a time.
- **Receipt:** The technique should ensure that message packets are actually received (no lost packets) and delivered only once (no duplicate packets), and are received in the proper order.
- **Error Limitation:** The method should be capable of encompassing an appropriate error detection scheme.
- **Recovery**: If two packets collide (are present on the network at the same time), or if notice of a collision appears, the method should be able to recover, i.e. be able to halt all the transmissions and select one station to retransmit.
- **Re-configurability**: The technique should enable a network to accommodate the addition or deletion of a station with no more than a noise transient from which the network station can recover.
- **Compatibility:** The technique should accommodate equipment from all vendors who build to its specification.
- **Robustness:** The technique should enable a network to confine operating in spite of a failure of one or several stations.

The MAC (Medium Access Control) techniques can be broadly divided into four categories; Contention-based, Round-Robin, Reservation-based and. Channelization-based. Under these four broad categories there are specific techniques, as shown in Figure 6 below:

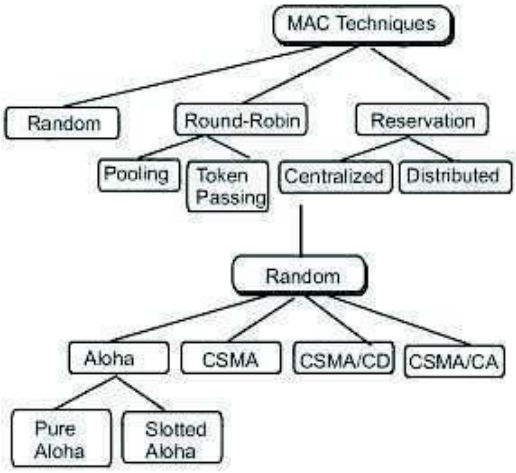


Figure 6: Classification of Medium Access Control techniques

There are different of methods used as access protocols in LANs, major techniques being token passing and CSMA/CD. Token passing can be used with ring or bus topologies. Token passing scheme is an access protocol that permits a terminal to transmit only on receipt of a special circulating bit sequence. CSMA/CD (carrier sense multiple access, with collision detected) is used with bus and some star topologies.

Random Access (Contention-based Approaches) : Round-Robin techniques work efficiently when majority of the stations have data to send most of the time. But, in situations where only a few nodes have data to send for brief periods of time, Round-Robin techniques are unsuitable. Contention techniques are suitable for bursty nature of

traffic. In contention techniques, there is no centralised control and when a node has data to send, it contends for gaining control of the medium. The principle advantage of contention techniques is their simplicity. They can be easily implemented in each node. The techniques work efficiently under light to moderate load, but performance rapidly falls under heavy load.

In the 1970s, Norman Abramson and his colleagues at the University of Hawaii devised a new and elegant method to solve the channel allocation problem. Their work has been extended by many researchers since then (Abramson, 1985). Although Abramson's work, called the ALOHA system, used ground-based radio broadcasting, the basic idea is applicable to any system in which uncoordinated users are competing for the use of a single shared channel.

ALOHA have two versions pure and slotted. They differ with respect to whether time is divided into discrete slots into which all frames must fit. Pure ALOHA does not require global time synchronization; slotted ALOHA does. These pure and slotted ALOHA schemes will be discussed further in this block.

CSMA/CD: CSMA/CD stands for Carrier Sense Multiple Access with Collision Detection. It refers to the means of media access, or deciding "who gets to talk" in an Ethernet network. In detailed mechanisms of CSMA/CD will be discussed further in this block.

Round Robin Techniques: In Round Robin techniques, each and every node is given the chance to send or transmit by rotation. When a node gets its turn to send, it may either decline to send, or it may send if it has got data to send. After getting the opportunity to send, it must relinquish its turn after some maximum period of time. The right to send then passes to the next node based on a predetermined logical sequence. The right to send may be controlled in a centralised or distributed manner. Polling is an example of centralised control and token passing is an example of distributed control.

i) **Polling:** The mechanism of polling is similar to the roll-call performed in a classroom. Just like the teacher, a controller sends a message to each node in turn. The message contains the address of the node being selected for granting access. Although all nodes receive the message, only the addressed node responds and then it sends data, if any. If there is no data, usually a “poll reject” message is sent back. In this way, each node is interrogated in a round-robin fashion, one after the other, for granting access to the medium. The first node is again polled when the controller finishes with the remaining codes. The polling scheme has the flexibility of either giving equal access to all the nodes, or some nodes may be given higher priority than others. In other words, priority of access can be easily implemented.

ii) **Token Passing:** In token passing scheme, all stations are logically connected in the form of a ring and control of the access to the medium is performed using a token. A token is a special bit pattern or a small packet, usually several bits in length, which circulate from node to node. Token passing can be used with both broadcast (token bus) and sequentially connected (token ring) type of networks with some variation.

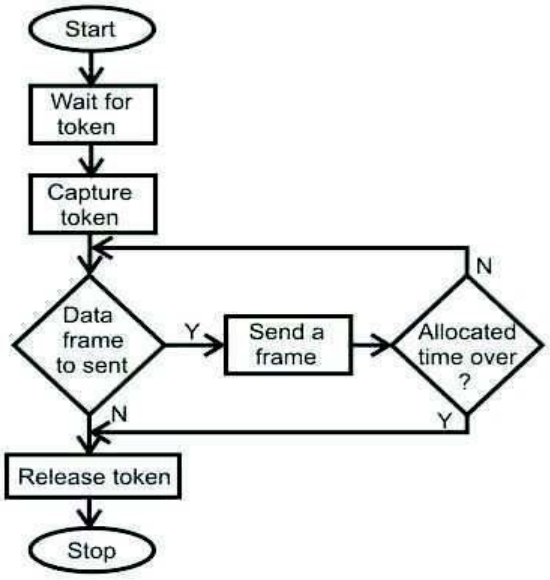


Figure 7: Mechanism of Token Passing

In case of token ring as shown in flowchart of figure 7, token is passed from a node to the physically adjacent node. On the other hand, in the token bus, token is passed with the help of the address of the nodes, which form a logical ring. In either case a node currently holding the token has the 'right to transmit'. When it has got data to send, it transmits the data and then forwards the token to the next logical or physical node in the ring. If a node currently holding the token has no data to send, it simply forwards the token to the next node. The token passing scheme is efficient compared to the polling technique, but it relies on the correct and reliable operation of all the nodes. There exists a number of potential problems, such as lost token, duplicate token, and insertion of a node, removal of a node, which must be tackled for correct and reliable operation of this scheme.

☞ Check Your Progress 2

1. Write an advantage and one disadvantage of star topology?

   ……

   ……

   ……

   ……

2. What is the difference Considerations while Choosing a Topology for a network?

   ……

   ……

   ……

   ……

1.5 NETWORK TYPES BASED ON SIZE

As you know that In order to meet the needs of various applications, networks are available with different interconnection layouts and plans, methods of access, protocols and data carrying capacities. Networks can be classified on the basis on size classified

are following. You have already studied the brief about LAN, MAN and WAN in the beginning of this unit. Now, in this section lets us again discuss them further.

Personal area network (PAN)

1. Local area network (LAN)
2. Metropolitan area network (MAN)
3. Wide area network (WAN)

    1. **PAN:** A personal area network (PAN) is a computer network organized around an individual person. Personal area networks typically involve network of a mobile computer, a cell phone and/or a handheld computing device such as a PDA. You can use these networks to transfer files including email and calendar appointments, digital photos and music. Personal area networks can be constructed with cables or wirelessly. USB and FireWire technologies often link together a wired PAN while wireless PANs typically use Bluetooth or sometimes infrared connections. Bluetooth PANs are also called piconets. Personal area networks generally cover a range of less than 10 meters (about 30 feet).

    2. **LAN:** A local area network (LAN) supplies networking capability to a group of computers in close proximity to each other such as in an office building, a school, or a home. A LAN is useful for sharing resources like files, printers, games or other applications. A LAN in turn often connects to other LANs, and to the Internet or other WAN. Most local area networks are built with relatively inexpensive hardware such as Ethernet cables, network adapters, and hubs. Wireless LAN and other more advanced LAN hardware options also exist. Specialized operating system software may be used to configure a local area network. For example, most flavors of Microsoft Windows provide a software package called Internet Connection Sharing (ICS) that supports controlled access to LAN resources.

    3. **MAN**: A Metropolitan Area Network (MAN) is a network that is designed to cover an entire city. As we have seen, organizations create smaller networks called as Local Area Networks (LANs). LANs are privately owned networks within the premises of an organization. However, suppose that an organization wants to connect the computers in its three city offices to each other. In such a case, the organization cannot obviously lay a private network all around the city. **WAN**: A Wide Area Network (WAN) is huge compared to a LAN or a MAN. A WAN spans across city, state, country or even continent boundaries. For instance, a WAN could be made up of a LAN in India, another LAN in the US and a third LAN in Japan, all connected to each other to form a big network of networks. The technical specifications of WAN differ from that of a LAN, although in principle, a WAN looks like a very big LAN.

1.6 FUNCTIONAL CLASSIFICATION OF NETWORKS

On the basis of functional relationship network is classified as follows:

1. Peer-to-peer
2. Client-server

    1. **Peer-to-Peer: --** Peer-to-peer network operating systems allow users to share resources and files located on their computers and to access shared resources found on other computers. However, they do not have a file server or a centralized management source (See figure 8 given below). In

a peer-to-peer network, all computers are considered equal; they all have the same abilities to use the resources available on the network. Peer-to-peer networks are designed primarily for small to medium local area networks. AppleShare and Windows for Workgroups are examples of programs that can function as peer-to-peer network operating systems.

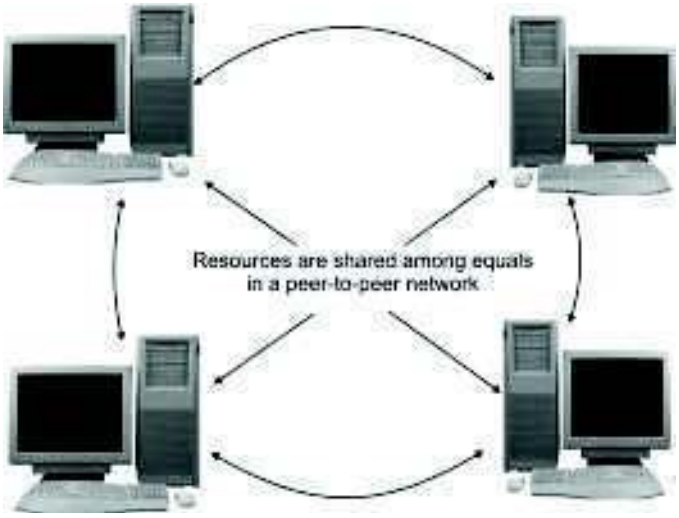


Figure 8: Peer to Peer network

The advantages of peer-to-peer over client-server NOSs include:

i) No need for a network administrator

ii) Network is fast/inexpensive to setup & maintain

iii) Each PC can make backup copies of its data to other PCs for security.

By far the easiest type of network to build, peer-to-peer is perfect for both home and office use.

2. **Client/Server: --** Client/server network operating systems allow the network to centralize functions and applications in one or more dedicated file servers. The file servers become the heart of the system, providing access to resources and providing security. Individual workstations (clients) have access to the resources available on the file servers. The network operating system provides the mechanism to integrate all the components of the network and allow multiple users to simultaneously share the same resources irrespective of physical location. Novell Netware and Windows NT Server are examples of client/server network operating system.

   In a client-server environment like Windows NT or Novell NetWare, files are stored on a centralized, high speed file server PC that is made available to client PCs. Network access speeds are usually faster than those found on peer-to-peer networks, which is reasonable given the vast numbers of clients that this architecture can support. Nearly all network services like printing and electronic mail are routed through the file server, which allows networking tasks to be tracked. Inefficient network segments can be reworked to make them faster, and users' activities can be closely monitored. Public data and applications are stored on the file server, where they are run from client PCs' locations, which make upgrading software a simple task network administrators can simply upgrade the applications stored on the file server, rather than having to physically upgrade each client PC.

1.7 WAN TOPOLOGIES

A wide area network (WAN) is a network connecting geographically distinct locations, which may or may not belong to the same organization. WAN topologies use both LAN add enterprise-wide topologies as building blocks, but add more complexity because of the distance they must cover, the larger number of users they serve, and the heavy traffic they often handle. For example, although a simple ring topology may suffice for a small office with 10 users, it does not scale well and therefore cannot serve 1000 users. The particular WAN topology you choose will depend on the number of sites you must connect, the distance between the sites, and any existing infrastructure.

WAN Ring Topology: In a ring WAN topology, each site is connected to two other sites so that the entire WAN forms a ring pattern. This architecture is similar to the ring LAN topology, except that a ring WAN topology connects locations rather than local nodes. The advantages of a ring WAN over a peer-to-peer WAN are twofold: a single cable problem will not affect the entire network, and routers at any site can redirect data to another route if one route becomes too busy. On the other hand, expanding a peer-to-peer WAN because it requires at least one additional link. For those reasons, WANs that use the ring topology are only practical for connecting fewer than four or five locations.

WAN Star Topology: The star WAN topology mimics the arrangement of a star LAN. A single site acts as the central connection point for several other points. This arrangement provides separate routes for data between any two sites. As a result, star WANs are more reliable than the peer-to-peer or ring WANs. As a general rule, reliability increases with the number of potential routes data can follow. Another advantage of a star WAN is that when all of its dedicated circuits are functioning, a star WAN provides shorter data paths between any two sites.

WAN Mesh Topology: Like an enterprise-wide mesh, a mesh WAN topology incorporates many directly interconnected nodes--in this case, geographical locations. Because every site is interconnected, data can travel directly from its origin to its destination. If one connection suffers a problem, routers can redirect data easily and quickly. Mesh WANs are the most fault-tolerant type of WAN configuration because they provide multiple routes for data to follow between any two points.

One drawback to a mesh WAN is the cost; connecting every node on a network to every other entails leasing a large number of dedicated circuits. With larger WANs, the expense can become enormous. To reduce costs, you might choose to implement a partial mesh, in which critical WAN nodes are directly interconnected and secondary nodes are connected through star or ring topologies. Partial-mesh WANs are more practical and therefore more common in today's business world, than full-mesh WANs.

1.8 WAN ACCESS METHODS

WAN access methods are as follows:

1. **Lease Line:** A permanent telephone connection between two points set up by a telecommunications common carrier. Typically, leased lines are used by businesses to connect geographically distant offices. Unlike normal dial-up connections, a leased line is always active. The fee for the connection is a fixed monthly rate. The primary factors affecting the monthly fee are distance between end points and the speed of the circuit. Because the connection doesn't carry anybody else's communications, the carrier can assure a given level of quality.

For example, a T-1 channel is a type of leased line that provides a maximum transmission speed of 1.544 Mbps. You can divide the connection into different lines for data and voice communication or use the channel for one high speed data circuit. Dividing the connection is called multiplexing.

Increasingly, leased lines are being used by companies, and even individuals, for Internet access because they afford faster data transfer rates and are cost-effective if the Internet is used heavily.

2. **Packet Switching: --**Packet switching is used to overcome from limitations of circuit switching, packet switching has emerged as the standard switching technology for computer-to-computer communications, and therefore, used by most of the communication protocols such as X.25, TCP/IP, Frame Relay, ATM, etc. Unlike in a circuit switching, in packet switching, data to be sent is divided into and then sent as discrete blocks, called packets, which are of potentially variable length. The underlying network mandates the maximum size of data called packet size or packet length-that can be transmitted at a given time. Each packet contains data to be transferred, and also the control information such as the sender's address and the destination's address. Packets also help in recovering from erroneous transmission quicker and more easily. This is because, in this case, only the packers in error need to be retransmitted.

3. **ISDN:** Integrated Services Digital Network (ISDN) was developed by ITU- Tin 1976. It is a set of protocols that combines digital telephony and data transport services. The whole idea is to digitize the telephone network to permit the transmission of audio, video, and text over existing telephone lines.

   ISDN is an effort to standardize subscriber services, provide user/network inter-faces, and facilitate the internetworking capabilities of existing voice and data networks. The goal of ISDN is to form a wide area network that provides universal end-to end connectivity over digital media. This can be done by integrating all of the separate transmission services into one without adding new links or subscriber lines.

   DSL: Digital subscriber line (DSL) is a family of technologies that provide Internet access by transmitting digital data over the wires of a local telephone network. It is a high-speed Internet service like cable Internet. DSL provides high-speed networking over ordinary phone lines using broadband modem technology. DSL technology allows Internet and telephone service to work over the same phone line without requiring customers to disconnect either their voice or Internet connections. DSL technology theoretically supports data rates of 8.448 Mbps, although typical rates are 1.544 Mbps or lower. DSL Internet services are used primarily in homes and small businesses. DSL Internet service only works over a limited physical distance and remains unavailable in many areas where the local telephone infrastructure does not support DSL technology.

☞ Check Your Progress 3

1. Write the advantage of peer-to-peer over client-server.

   ……

   ……

   ……

   ……

2. List any three WAN access methods.

……

……

……

1.9 SUMMARY

A communication system that supports many users is called a network. In a network many computers are connected to each other by various topologies like star, ring, complete, interconnected or irregular. Depending on the area of coverage a network can be classified as LAN, MAN or WAN. A network is required for better utilisation of expensive resources, sharing information, collaboration among different groups, multimedia communication and video conferencing.

The two different types of networking models OSI and TCP/IP are existing. The difference between these models was discussed in detail.

1.10 REFERENCES/FURTHER READING

1. *Computer Networks*, A. S. *Tanenbaum* 4th Edition, Practice Hall of India, New Delhi. 2003.
2. *Introduction to Data Communication & Networking,* 3rd Edition, Behrouz Forouzan, Tata McGraw Hill.
3. *Computer Networking*, J.F. Kurose & K.W. Ross, A Top Down Approach Featuring the Internet, Pearson Edition, 2003.
4. *Communications Networks*, Leon Garcia, and Widjaja, Tata McGraw Hill, 2000.
5. *Data and Computer Communications,* Willian Stallings, 6th Edition, Pearson Education, New Delhi.
6. www.wikipedia.org
7. Larry L. Peterson, *Computer Networks*: A Systems Approach, 3rd Edition (The Morgan Kaufmann Series in Networking).

1.11 SOLUTIONS/ANSWERS

☞ Check Your Progress 1

1. There are basically two types of networks:
   i) Point to point network or switched networks
   ii) Broadcast Networks.
2. Broadcasting refers to addressing a packet to all destinations in a network whereas multicasting refers to addressing a packet to a subset of the entire network.

☞ Check Your Progress 2

1. **Advantages of a Star Topology: --** The advantages of star topologies are as follows:
   i) Easy to add new stations as each station has its own direct cable

connection to the switch. If a cable is cut, it only affects the computer that was attached to it.

ii) It can accommodate different wiring. It can be installed using twisted pair, coaxial cable or fiber optic cable.

Disadvantages of a Star Topology: --The advantages of star topologies are as follows:-

i) Depending on where the switches are located, star networks can require more cable length than a linear topology.

ii) If the switch / concentrator/switches fail, nodes attached are disabled.

2. The considerations while choosing topologies are as follows: --

i) **Cost:** A linear bus network may be the least expensive way to install a network; you do not have to purchase concentrators

ii) **Length of cable needed:** The linear bus network uses shorter lengths of cable.

iii) **Future growth:** With a star topology, expanding a network is easily done by adding another switch.

iv) **Cable type:** The most common cable is unshielded twisted pair, which is most often used with bus, star topologies.

☞ Check Your Progress 3

1. The advantages of peer-to-peer over client-server based networks are:

i) No need for a network administrator

ii) Network is fast/inexpensive to setup & maintain

iii) Each PC can make backup copies of its data to other PCs for security.

By far the easiest type of network to build, peer-to-peer is perfect for both home and office use.

2. WAN access methods are as follows:

i) **Packet Switching: --**Packet switching is used to overcome from limitations of circuit switching, packet switching has emerged as the standard switching technology for computer-to-computer communications, and therefore, used by most of the communication protocols such as X.25, TCP/IP, Frame Relay, ATM, etc.

ii) **Lease Line:** A permanent telephone connection between two points set up by a telecommunications common carrier.

iii) **ISDN:** Integrated Services Digital Network (ISDN) was developed by ITU- Tin 1976. It is a set of protocols that combines digital telephony and data transport services.

iv) **DSL**: Digital subscriber line (DSL) is a family of technologies that provide Internet access by transmitting digital data over the wires of a local telephone network.

UNIT 2 OSI AND TCP/IP MODELS

Structure **Page Nos**

2.0 Introduction 22
2.1 Objectives 22
2.2 OSI Reference Model 23
- 2.2.1 Layers in the OSI model
- 2.2.2 Layer 1: the physical layer
- 2.2.3 Layer 2: the data-link layer
- 2.2.4 Layer 3: the network layer
- 2.2.5 Layer 4: the transport layer
- 2.2.6 Layer 5: the session layer
- 2.2.7 Layer 6: the presentation layer
- 2.2.8 Layer 7: the application layer

2.3 TCP/IP Model 28
- 2.3.1 Layers in the TCP/IP model
- 2.3.2 TCP/IP application layer
- 2.3.3 TCP/IP transport layer
- 2.3.4 TCP/IP internet layer
- 2.3.5 TCP/IP network access layer

2.4 Comparison of OSI and TCP/IP Models 31
2.5 TCP/IP Protocols 32
- 2.5.1 Application layer protocols
- 2.5.2 Transport layer protocols
- 2.5.3 Internet layer protocols

2.6 Summary 38
2.7 References/Further Readings 38
2.8 Solutions/Answers 39

2.0 INTRODUCTION

In order for a computer to send information to another computer, and for that computer to receive and understand the information, there has to exist a set of rules or standards for this communication process. These standards ensure that varying devices and products can communicate with each other over any network. This set of standards is called a network reference model. There are a variety of networked models currently being implemented. However, in this unit, the focus will be on the OSI and TCP/IP models.

2.1 OBJECTIVES

After going through this unit, you should be able to know:

- The seven layers of OSI reference model
- Understand each layer of OSI model
- Functions of each layer of OSI model
- Understanding of TCP/IP model and its four Layers
- Detail Description of protocol used in each layer
- Similarities of OSI and TCP/IP

2.2 OSI REFERNCE MODEL

In 1983, the International Standards Organization (ISO) developed a model called Open Systems Interconnection (OSI) which is a standard reference model for

communication between two end users in a network. The model is used in developing products and understanding networks. It is a prescription of characterizing and standardizing the functions of a communications system in terms of abstraction layers. Similar communication functions are grouped into logical layers. A layer serves the layer above it and is served by the layer below it.

2.2.1 Layers in the OSI Model

OSI divides Telecommunications into Seven Layers as shown below in the Figure 1 given below. Each layer is responsible for a particular aspect of data communication. For example, one layer may be responsible for establishing connections between devices, while another layer may be responsible for error checking during transfer.

The layers of the OSI model are divided into two groups: the upper layers and lower layers. The upper layers (Host layers) focus on user applications and how files are represented on the computers prior to transport. The lower layers (Media Layers) concentrate on how the communication across a network actually occurs. Each layer has a set of functions that are to be performed by a specific protocol(s). The OSI reference model has a protocol suit for all of its layers.

In computing, a protocol is a convention or standard that controls or enables the connection, communication, and data transfer between two computing endpoints. In its simplest form, a protocol can be defined as the rules governing the syntax, semantics, and synchronization of communication.

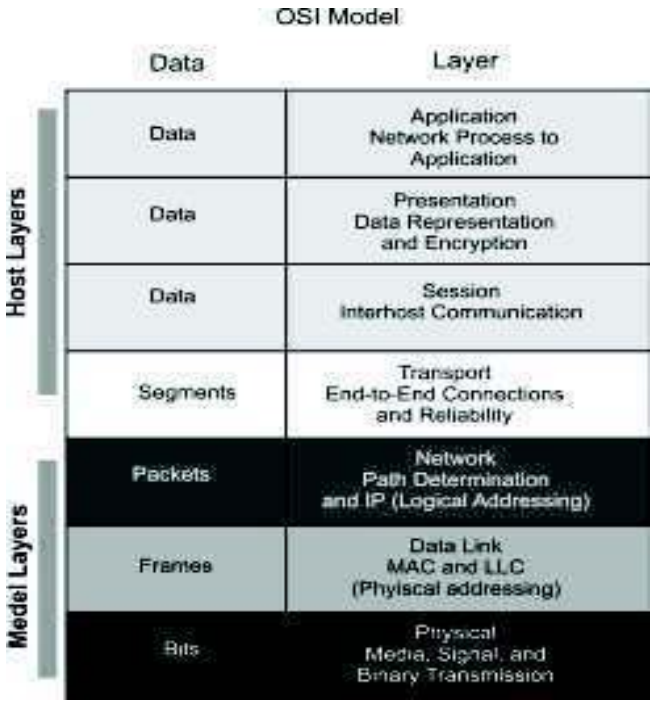


Figure 1: The OSI Model

2.2.2 Layer 1: The Physical Layer

The physical layer, the lowest layer of the OSI model, is concerned with the transmission and reception of the unstructured raw bit stream over a physical medium. It describes the electrical/optical, mechanical, and functional interfaces to the physical medium, and carries the signals for all of the higher layers. It provides:

- **Data encoding:** modifies the simple digital signal pattern (1s and 0s) used by the PC to better accommodate the characteristics of the physical medium, and to aid in bit and frame synchronization.
- **Transmission technique:** determines whether the encoded bits will be transmitted by baseband (digital) or broadband (analog) signaling.
- **Physical medium transmission:** transmits bits as electrical or optical signals appropriate for the physical medium, and determines: What physical medium options can be used? And How many volts/db should be used to represent a given signal state, using a given physical medium?

2.2.3 Layer 2: The data-link layer

The data link layer provides error-free transfer of data frames from one node to another over the physical layer, allowing layers above it to assume virtually error-free transmission over the link. To do this, the data link layer provides:

- **Frame Traffic Control:** tells the transmitting node to "stop" when no frame buffers are available.
- **Frame Sequencing:** transmits/receives frames sequentially.
- **Frame Acknowledgment:** provides/expects frame acknowledgments. Detects and recovers from errors that occur in the physical layer by retransmitting non-acknowledged frames and handling duplicate frame receipt.
- **Frame Delimiting:** creates and recognizes frame boundaries.
- **Link Establishment and Termination:** establishes and terminates the logical link between two nodes.
- **Frame Error Checking:** checks received frames for integrity.
- **Media access management:** determines when the node "has the right" to use the physical medium.

Data Link Sub layers

The Data Link Layer is described in more detail with Media Access Control (MAC) and Logical Link Control (LLC) sub layers; where LLC is consider as upper data link layer and MAC as lower data link layer as shown below in the Figure 2.

- Logical Link Control (LLC): The LLC is concerned with managing traffic (flow and error control) over the physical medium and may also assign sequence numbers to frames and track acknowledgements. LLC is defined in the IEEE 802.2 specification and supports both connectionless and connection-oriented services used by higher-layer protocols.
- Media Access Control (MAC): The MAC sub layer controls how a computer on the network gains access to the data and permission to transmit it.

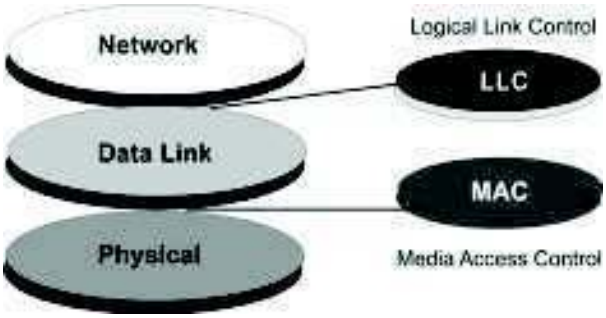


Figure 2: Data Link Sub-Layers

2.2.4 Layer 3: The Network Layer

The network layer provides the functional and procedural means of transferring variable length data sequences from a source host on one network to a destination host on a different network. The network layer performs network routing functions, and might also perform fragmentation and reassembly, and report delivery errors. Routers operate at this layer, sending data throughout the extended network and making the Internet possible.

Functions of the network layer include:

- Connection setup
- Addressing
- Routing
- Security
- Quality of Service
- Fragmentation

The Network Layer identifies computers on a network. Two types of packets are used at the Network layer; Data packets and Route update packets. Data packets are used to transport user data through the Internet work. Protocols used to support data traffic are called routed protocols. Route update packets are used to update neighboring routers about the network connected to all routers within the internet work. Protocols that send route updates are called routing protocols. This layer is concerned with two functions Routing and Fragmentation / Reassembly:

Routing: It is the process of selecting the best paths in a network along which to send data on physical traffic as shown in Figure 3.

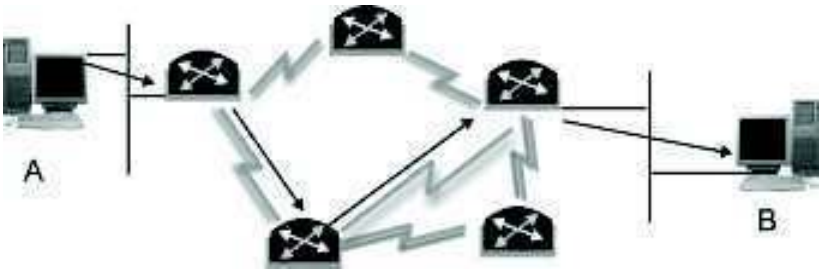


Figure 3: Routing at Network Layer

Fragmentation / Reassembly: if the network layer determines that a next router's maximum transmission unit (MTU) size is less than the current frame size, a router can fragment a frame for transmission and re-assembly at the destination station.

2.2.5 Layer 4: The Transport Layer

The transport layer provides transparent transfer of data between end users, providing reliable data transfer services to the upper layers. The transport layer controls the reliability of a given link through flow control, segmentation/de-segmentation, and error control. This layer manages the end-to-end control (for example, determining whether all packets have arrived). It ensures complete data transfer. The Basic Transport Layer Services are:

- **Resource Utilization (multiplexing):** Multiple applications run on the same machine but use different ports.
- **Connection Management (establishing & terminating):** The second major task of Transport Layer is establishing connection between sender & the receiver before data transmission starts & terminating the connection once the data transmission is finished
- **Flow Control (Buffering / Windowing):** Once the connection has occurred and transfer is in progress, congestion of the data flow can occur at a destination for a variety of reasons. Possible options include:
    - The destination can become overwhelmed if multiple devices are trying to send it data at the same time.
    - The destination can become overwhelmed if the source is sending faster than it can physically receive.

The Transport Layer is responsible for providing flow control to alleviate the issue of congestion in the data transfer. Two main methods for flow control include:

Buffering: Buffering is a form of data flow control regulated by the Transport Layer as depicted in Figure 4. It is responsible for ensuring that sufficient buffers (Temporary Memory) are available at the destination for the processing of data and that the data is transmitted at a rate that does not exceed what the buffer can handle.

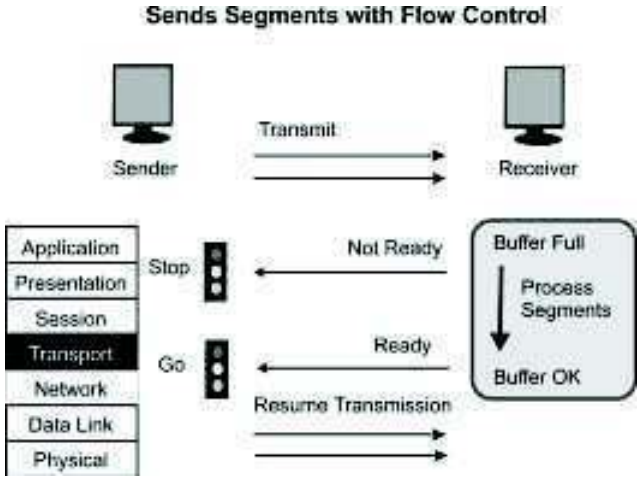


Figure 4: Buffering at Work

Windowing: Windowing is a flow control scheme in which the source computer will monitor and make adjustments to the amount of information sent based on successful, reliable receipt of data segments by the destination computer as shown in Figure 5. The size of the data transmission, called the "window size", is negotiated at the time of connection establishment, which is determined by the amount of memory or buffer that is available.

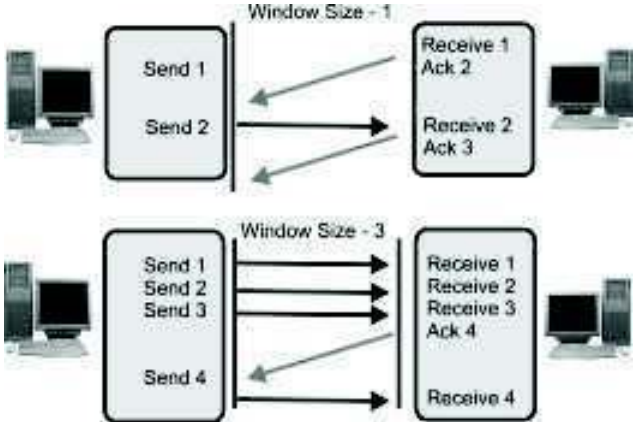


Figure 5: Flow Control & Reliability through Windowing

Reliable Transport (positive acknowledgment): Transport layer provides reliable transport of data by sending positive acknowledgements back to the sender once the data has reached the receiving side, if the data is lost or is corrupted, a negative acknowledgement is sent.

2.2.6 Layer 5: The Session Layer

The session Layer establishes, manages, and terminates sessions (different from connections) between applications as they interact on different hosts on a network. Its main job is to coordinate the service requests and responses between different hosts for applications.
The session established between hosts can be Simplex, half duplex and full duplex:

- **Simplex:** Simplex transmission is like a one-way street where traffic moves in only one direction. Simplex mode is a one-way-only transmission, which means that data can flow only in one direction from the sending device to the receiving device.
- **Half Duplex:** Half Duplex is like the center lane on some three-lane roads. It is a single lane in which traffic can move in one direction or the other, but not in both directions at the same time. Half-duplex mode limits data transmission because each device must take turns using the line. Therefore, data can flow from A to B and from B to A, but not at the same time.
- **Full Duplex:** is like a major highway with two lanes of traffic, each lane accommodating traffic going in opposite directions. Full-duplex mode accommodates two-way simultaneous transmission, which means that both sides can send and receive at the same time. In full-duplex mode, data can flow from A to B and B to A at the same time.

Note: Full-duplex transmission is, in fact, two simplex connections: One connection has traffic flowing in only one direction; the other connection has traffic flowing in the opposite direction of the first connection.

2.2.7 Layer 6: The Presentation Layer

The presentation layer transforms data into the form that the application accepts. This layer formats and encrypts data to be sent across a network. This layer, usually part of an operating system, that converts incoming and outgoing data from one presentation format to another (for example, from a text stream into a popup window with the newly arrived text). This layer is sometimes called the syntax layer. The Presentation Layer is responsible for the following services:

- **Data representation:** The presentation layer of the OSI model at the receiving computer is also responsible for the conversion of “the external format” with

which data is received from the sending computer to one accepted by the other layers in the host computer. Data formats include postscript, ASCII, or BINARY such as EBCDIC (fully Extended Binary Coded Decimal Interchange Code).

- **Data security:** Some types of encryption (and decryption) are performed at the presentation layer. This ensures the security of the data as it travels down the protocol stack.
- **Data compression:** Compression (and decompression) may be done at the presentation layer to improve the throughput of data.

2.2.8 Layer 7: The Application Layer

The application layer is closest to the end user, which means that both the OSI application layer and the user interact directly with the software application. This layer interacts with software applications that implement a communicating component.

The Application Layer is the highest layer in the protocol stack and the layer responsible for introducing data into the OSI stack. The functions of Application Layer are:

- Resource sharing and device redirection
- Remote file access
- Remote printer access
- Network management
- Directory services
- Electronic messaging (such as mail) etc

2.3 TCP/IP MODEL

The TCP/IP Model is a specification for computer network protocols created in the 1970s by DARPA, an agency of the United States Department of Defense. It laid the foundation for ARPANET, which was the world's first wide area network and a predecessor of the Internet.

2.3.1 Layers in the TCP/IP Model

TCP/IP is generally described as having four 'layers' or five if we include the bottom physical layer. The layers near the top are logically closer to the user application, while those near the bottom are logically closer to the physical transmission of the data.

2.3.2 TCP/IP Application Layer

TCP/IP application layer protocols provide services to the application software running on a computer. The application Layer identifies the application running on the computer through Port Numbers.

The various protocols that are used at the Application Layer are:

- **Telnet**: Terminal Emulation, Telnet is a program that runs on your computer and connects your PC to a server on the network. You can then enter commands through the Telnet program and they will be executed as if you were entering them directly on the server console. Port Number :23

- **FTP**: File Transfer Protocol, the protocol used for exchanging files over the Internet. FTP is most commonly used to download a file from a server using the Internet or to upload a file to a server. Port Number : 20(data port) ,21(control port)
- **HTTP**: Hyper Text Transfer Protocol is the underlying protocol used by the World Wide Web. HTTP defines how messages are formatted and transmitted, and what actions Web servers and browsers should take in response to various commands. For example, when we enter a URL in the browser, this actually sends an HTTP command to the Web server directing it to fetch and transmit the requested Web page. Port Number :80
- **NFS**: Network File System, a client/server application that allows all network users to access shared files stored on computers of different types. Users can manipulate shared files as if they were stored locally on the user's own hard disk. Port Number :2049
- **SMTP**: Simple Mail Transfer Protocol, a protocol for sending e-mail messages between servers. In addition, SMTP is generally used to send messages from a mail client to a mail server. Port Number :25
- **POP3**: Post Office Protocol, a protocol used to retrieve e-mail from a mail server. Most e-mail applications (sometimes called an e-mail client) use the POP, although some can use the newer IMAP (Internet Message Access Protocol)as a replacement for POP3 Port Number :110
- **TFTP**: Trivial File Transfer Protocol, a simple form of the File Transfer Protocol (FTP). TFTP provides no security features. It is often used by servers to boot diskless workstations, X-terminals, and routers. Port Number :69
- **DNS**: Domain Name System (or Service or Server), an Internet service that translates domain names into IP addresses. Because domain names are alphabetic, they're easier to remember. The Internet however, is really based on IP addresses. Every time you use a domain name, a DNS service must translate the name into the corresponding IP address. For example, the domain name www.example.com might translate to 198.105.232.4. Port Number :53
- **DHCP**: Dynamic Host Configuration Protocol, a protocol for assigning dynamic IP addresses to devices on a network. With dynamic addressing, a device can have a different IP address every time it connects to the network. Dynamic addressing simplifies network administration because the software keeps track of IP addresses rather than requiring an administrator to manage the task. Port Number : 67(Server),68(Client)
- **BOOTP**: Bootstrap Protocol (BOOTP) is utilized by diskless workstations to gather configuration information from a network server. This enables the workstation to boot without requiring a hard or floppy disk drive. Port Number : 67(Server),68(Client)
- **SNMP**: Simple Network Management Protocol, a set of protocols for managing complex networks. SNMP works by sending messages, called protocol data units (PDUs), to different parts of a network. Port Number :161

2.3.3 TCP/IP Transport Layer

The protocol layer just below the Application layer is the *host-to-host layer (Transport layer)*. It is responsible for end-to-end data integrity. Transport Layer identifies the segments through *Socket address* (Combination of Port Number & I.P. address).
The two most important protocols employed at this layer are the

- *Transmission Control Protocol (*TCP*):* TCP provides *reliable, full-duplex connections* and *reliable service* by ensuring that data is retransmitted when transmission results in an error (end-to-end error detection and correction). Also, TCP enables hosts to maintain multiple, simultaneous connections.
- *User Datagram Protocol (UDP):* When error correction is not required, UDP provides *unreliable datagram service* (connectionless) that enhances network throughput at the host-to-host transport layer. It's used primarily for *broadcasting* messages over a network.

2.3.4 TCP/IP Internet Layer

The best known TCP/IP protocol at the internetwork layer is the *Internet Protocol (*IP*),* which provides the basic packet delivery service for all TCP/IP networks node addresses, the IP implements a system of logical host addresses called IP addresses.

The IP addresses are used by the internetwork and higher layers to identify devices and to perform internetwork routing. IP is used by all protocols in the layers above and below it to deliver data, which means all TCP/IP data flows through IP when it is sent and received, regardless of its final destination.

The basic protocols used at the Internet Layer are:

- I.P. (Internet Protocol): It is a protocol used at the internet layer of TCP/IP model by which data is encapsulated and is sent from one computer to another on the Internet.
- ARP (Address Resolution Protocol): It is used to map the known I.P. addresses into Physical address.
- RARP(Reverse Address Resolution Protocol): It is used to map Physical address into I.P. address
- I.C.M.P.(Internet Control Message Protocol): It is used to send error & control Messages in the network
- I.G.M.P. (Internet Group Management Protocol): It is a protocol which is used to form multicast groups in a network to receive multicast messages.

2.3.5 TCP/IP Network Access Layer

The *network access layer* is the lowest layer in the TCP/IP model. This layer contains the protocols that the computer uses to deliver data to the other computers and devices that are attached to the network. The protocols at this layer perform three distinct functions:

- They define how to use the network to transmit a *frame*, which is the data unit passed across the physical connection.
- They exchange data between the computer and the physical network.
- They deliver data between two devices on the same network using the physical address.

The network access layer includes a large number of protocols. For instance, the network access layer includes all the variations of Ethernet protocols and other LAN standards. This layer also includes the popular WAN standards, such as the Point-to-Point Protocol (PPP) and Frame Relay.

2.4 COMPARISON OF OSI AND TCP/IP MODELS

As it can be seen from the previous pages, there are a number of comparisons, which can be drawn between the two models as shown below in the Figure 6. This section will therefore be focusing on highlighting the similarities and differences between the OSI and TCP/IP models.

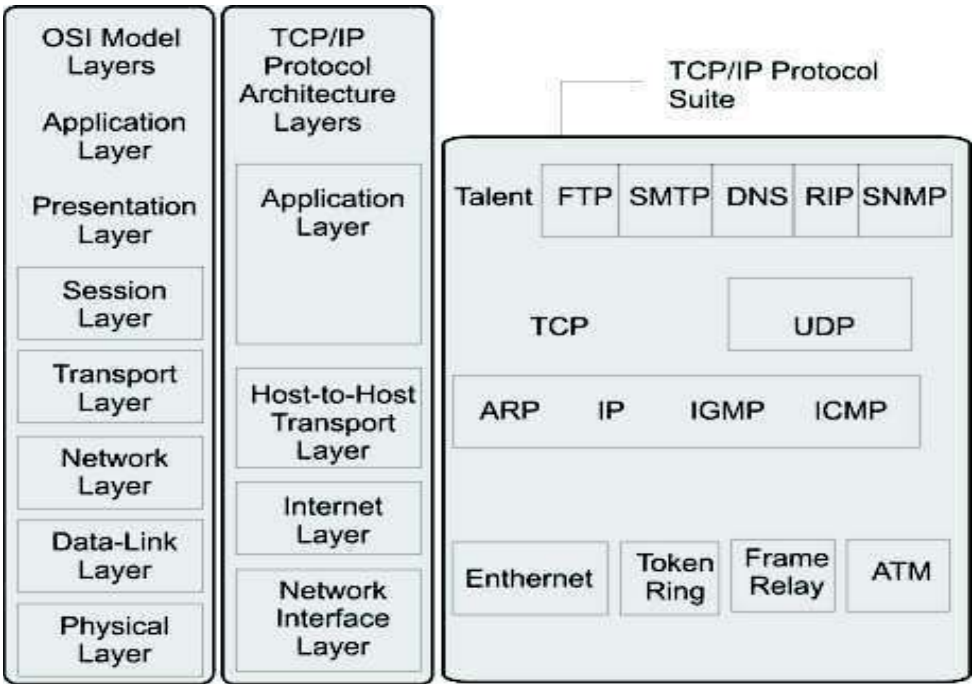


Figure 6: OSI Vs TCP/IP

Similarities

The main similarities between the OSI and TCP/IP models include the following:

- They share similar architecture. - Both of the models share a similar architecture. This can be illustrated by the fact that both of them are constructed with layers.

- They share a common application layer.- Both of the models share a common "application layer". However in practice this layer includes different services depending upon each model.

- Both models have comparable transport and network layers.- This can be illustrated by the fact that whatever functions are performed between the presentation and network layer of the OSI model similar functions are performed at the Transport layer of the TCP/IP model.

- Both models assume that packets are switched.- Basically this means that individual packets may take differing paths in order to reach the same destination.

Differences

The main differences between the two models are as follows:

- TCP/IP Protocols are considered to be standards around which the internet has developed. The OSI model however is a "generic, protocol- independent standard."
- TCP/IP combines the presentation and Chapter layer issues into its application layer.
- TCP/IP combines the OSI data link and physical layers into the network access layer.
- TCP/IP appears to be a simpler model and this is mainly due to the fact that it has fewer layers.
- TCP/IP is considered to be a more credible model- This is mainly due to the fact because TCP/IP protocols are the standards around which the internet was developed therefore it mainly gains creditability due to this reason. Where as in contrast networks are not usually built around the OSI model as it is merely used as a guidance tool.

☞ Check Your Progress 1

1. How transport layer of OSI model provide flow control to improve the issue of congestion in the data transfer?

   ……

   ……

   ……

   ……

2. Write the main similarities between the TCP/IP and OSI reference models.

   ……

   ……

   ……

2.5 TCP/IP PROTOCOLS

Transmission Control Protocol (TCP)/Internet Protocol (IP) is a set of protocols developed to allow computers of all sizes from different vendors, running different operating systems, to communicate or to share resources across a network. A packet switching network research project was started by the USA Government in the late 1960s in 1990s, became the most widely used form of computer networking. This project centered on ARPANET. ARPANET is best-known TCP/IP network. TCP/IP is the principal UNIX networking protocol and was designed to provide a reliable end-to-end byte stream over an unreliable internetwork. TCP is a connection-oriented protocol while IP is a connectionless protocol. TCP supplies logic to give a reliable connection-oriented protocol above IP. It provides a virtual-circuit that two processes can use to communicate. IP provides a connectionless and unreliable delivery system and transfer each datagram independently in the network. UDP is a connectionless and unreliable protocol running over IP. It adds a checksum to IP for the contents of the datagram and pass members. In this section, we are going to discuss all the protocols of TCP/IP in brief.

2.5.1 Application Layer Protocols

The Application layer provides applications the ability to access the services of the other layers and defines the protocols that applications use to exchange data. There are many Application layer protocols and new protocols are always being developed. The major functions of Application Layer are:

- Transfer of file that make up of Web pages
- Interactive file transfer(FTP)
- Transfer of mail messages and attachments
- Logging on remotely to networks hosts
- Resolving host name of an IP address
- Exchanging routing information on an IP internetwork.
- Collecting and exchanging network management information.

The Most common Application Layer Protocols are:

- Telnet (Network Terminal Protocol)
- FTP (File Transfer Protocol)
- SMTP(Simple Mail Transfer Protocol)
- DNS(Domain Name Server)
- RIP(Routing Information Protocol)
- SNMP(Simple Network Management Protocol)

Network Terminal Protocol

The purpose of the Telnet protocol is to provide a fairly general, bi-directional, eight-bit byte-oriented communications facility. Its primary goal is to allow a standard method of interfacing terminal devices and terminal-oriented processes to each other.

Telnet not only allows the user to log in to a remote host, it allows that user to execute commands on that host. Thus, an individual in Los Angeles can Telnet to a machine in New York and begin running programs on the New York machine just as though the user were actually in New York.

File Transfer Protocol

FTP (File Transfer Protocol) is the simplest and most secure way to exchange files over the Internet. Whether you know it or not, you most likely use FTP all the time. The most common use for FTP is to *download* files from the Internet. When *downloading* a file from the Internet you're actually *transferring* the file to your computer from another computer over the Internet. This is why the \T (transfer) is in FTP. You may not know where the computer is that the file is coming from but you most likely know its URL or Internet address.

An FTP address looks a lot like an HTTP, or Website, address except it uses the prefix *ftp://* instead of *http://*.

Example Website address:	http://www.ignou.ac.in
Example FTP site address:	ftp://www.ignou.ac.in

Simple Mail Transfer Protocol

SMTP is a relatively simple, text-based protocol, in which one or more recipients of a message are specified (and in most cases verified to exist) along with the message text and possibly other encoded objects. The message is then transferred to a remote server using a procedure of queries and responses between the client and server. Either an end-user's email client, a.k.a. MUA (Mail User Agent), or a relaying server's MTA (Mail Transport Agents) can act as an SMTP client.

An email client knows the outgoing mail SMTP server from its configuration. A relaying server typically determines which SMTP server to connect to by looking up the MX (Mail eXchange) DNS record for each recipient's domain name (the part of the email address to the right of the at (@) sign). Conformant MTAs (not all) fall back to a simple A record in the case of no MX. (Relaying servers can also be configured to use a smart host.)

The SMTP client initiates a TCP connection to server's port 25 (unless overridden by configuration). It is quite easy to test an SMTP server using the telnet program. SMTP is a "push" protocol that does not allow one to "pull" messages from a remote server on demand. To do this a mail client must use POP3 or IMAP. Another SMTP server can trigger a delivery in SMTP using ETRN.

HyperText Transfer Protocol

Hypertext Transfer Protocol (HTTP) is a communications protocol for the transfer of information on the intranet and the World Wide Web. Its original purpose was to provide a way to publish and retrieve hypertext pages over the Internet.

HTTP is a request/response standard between a client and a server. A client is the end-user; the server is the web site. The client making an HTTP request - using a web browser, spider, or other end-user tool - is referred to as the user agent. The responding server - which stores or creates resources such as HTML files and images - is called the origin server. In between the user agent and origin server may be several intermediaries, such as proxies, gateways, and tunnels. HTTP is not constrained to using TCP/IP and its supporting layers, although this is its most popular application on the Internet. Indeed HTTP can be "implemented on top of any other protocol on the Internet, or on other networks. HTTP only presumes a reliable transport; any protocol that provides such guarantees can be used."

Typically, an HTTP client initiates a request. It establishes a Transmission Control Protocol (TCP) connection to a particular port on a host (port 80 by default; see List of TCP and UDP port numbers). An HTTP server listening on that port waits for the client to send a request message. Upon receiving the request, the server sends back a status line, such as "HTTP/1.1 200 OK", and a message of its own, the body of which is perhaps the requested file, an error message, or some other information.

The reason that HTTP uses TCP and not UDP is because much data must be sent for a webpage, and TCP provides transmission control, presents the data in order, and provides error correction.

Domain Name Server

The most basic task of DNS is to translate hostnames to IP addresses. In very simple terms, it can be compared to a phone book. DNS also has other important uses.

Above all, DNS makes it possible to assign Internet names to organizations (or concerns they represent) independent of the physical routing hierarchy represented by the numerical IP address. Because of this, hyperlinks and Internet contact information

can remain the same, whatever the current IP routing arrangements may be, and can take a human-readable form (such as "example.com"), which is easier to remember than the IP address 208.77.188.166. People take advantage of this when they recite meaningful URLs and e-mail addresses without caring how the machine will actually locate them.

The Domain Name System distributes the responsibility for assigning domain names and mapping them to IP networks by allowing an authoritative name server for each domain to keep track of its own changes, avoiding the need for a central register to be continually consulted and updated.

Simple Network Management Protocol

Simple Network Management Protocol (SNMP) is a popular protocol for network management. It is used for collecting information from, and configuring, network devices, such as servers, printers, hubs, switches, and routers on an Internet Protocol (IP) network. Using SNMP, you can monitor network performance, audit network usage, detect network faults or inappropriate access, and in some cases configure remote devices. SNMP is designed to be deployed on the largest possible number of network devices, to have minimal impact on the managed nodes, to have minimal transport requirements, and to continue working when most other network applications fail.

Network File System

NFS stands for Network File System, a file system developed by Sun Microsystems, Inc. It is a client/server system that allows users to access files across a network and treats them as if they resided in a local file directory. For example, if you were using a computer linked to a second computer via NFS, you could access files on the second computer as if they resided in a directory on the first computer. This is accomplished through the processes of exporting (the process by which an NFS server provides remote clients with access to its files) and mounting (the process by which file systems are made available to the operating system and the user).

The NFS protocol is designed to be independent of the computer, operating system, network architecture, and transport protocol. This means that systems using the NFS service may be manufactured by different vendors, use different operating systems, and be connected to networks with different architectures. These differences are transparent to the NFS application, and thus, the user.

2.5.2 Transport Layer Protocols

In the TCP/IP model, the transport layer is responsible for delivering data to the appropriate application process on the host computers. This involves multiplexing of data from different application processes, i.e. forming a *segment* by adding source and destination port numbers in the header of each transport layer data packet. Together with the source and destination IP address (from the internet layer), the port numbers constitutes a network socket, i.e. an identification address of the process-to-process communication.

The major functions of Transport Layer are:

- It sets up and maintains a Chapter connection between two devices.
- It can provide for the reliable or unreliable delivery of data across the connection.
- It can implement flow control through ready/not ready signals or Windowing to ensure that the sender do not overwhelm the receiver with too many segments.

- It multiplexes the connections, allowing multiple applications to simultaneously send and receive data through port or socket numbers

The Most common Transport Layer Protocols are:

- T.C.P (Transmission Control Protocol)
- U.D.P (User Datagram Protocol)

Transmission Control Protocol

TCP is a Reliable (guarantees that the data sent across the connection will be delivered exactly as sent, without missing or duplicate data), Connection oriented (An application requests a connection, and then uses it for data transfer) protocol on the transport layer that provides in-order delivery of data and also use buffering and windowing to implement flow control.

User Datagram Protocol

The UDP is an unreliable connectionless protocol of the transport layer. UDP is *unreliable*, means that UDP does not provide mechanisms for error detection and error correction between the source and the destination. Because of this, UDP utilized bandwidth more efficiently than TCP. *Connectionless*, means that a network node can communicate with another network node using UDP without first negotiating any kind of handshaking or creating a connection. Because of this, UDP is very efficient for protocols that send very small amounts of data at irregular intervals.

2.5.3 Internet Layer Protocols

The TCP/IP internet-layer functionality includes transmitting data to and from the TCP/IP network interface layer, routing data to the correct network and station on the destination network, and handling packet errors and fragmentation.

Internet Protocol

The Internet Protocol is the building block of the Internet. IP is a **connectionless protocol**, means it does not exchange control information (handshake) to provide end-to-end control of communications flow. It relies on other layers to provide this function if it is required. IP also relies on other layers to provide error detection and correction. Because of this IP is sometimes referred to as an **unreliable protocol** because it contains no error detection and recovery code. IP can be relied upon to accurately deliver your data to the connected network, but it doesn't check whether that data was correctly received.

Its functions include:

- Defining the datagram, which is the basic unit of transmission in the Internet
- Defining the Internet addressing scheme
- Moving data between the Network Access Layer and the Host-to-Host Transport Layer
- Routing datagrams to remote hosts
- Performing fragmentation and re-assembly of datagrams

Address Resolution Protocol (ARP)

The address resolution protocol is a protocol used by the Internet Protocol (IP), specifically IPv4 (IP version 4), to map IP network addresses to the hardware addresses used by a data link protocol as depicted in figure 7. It is used when IPv4 is used over Ethernet. ARP works on Ethernet networks as follows. Ethernet network

adapters are produced with a physical address embedded in the hardware called the Media Access Control (MAC) address.

Manufacturers take care to ensure these 6-byte (48-bit) addresses are unique, and Ethernet relies on these unique identifiers for message delivery. When any device wishes to send data to another target device over Ethernet, it must first determine the MAC address of that target given its IP address These IP-to-MAC address mappings are derived from an **ARP cache** maintained on each device.

If the given IP address does not appear in a device's cache, that device cannot direct messages to that target until it obtains a new mapping. To do this, the initiating device first sends an *ARP request broadcast message* on the local subnet. The host with the given IP address sends an *ARP reply* in response to the broadcast, allowing the initiating device to update its cache and proceed to deliver messages directly to the target.

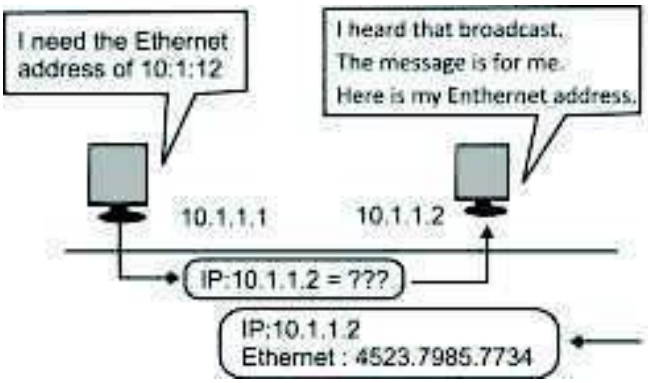


Figure 7: Working of ARP

Reverse Address Resolution Protocol (RARP)

Reverse Address Resolution Protocol, a TCP/IP protocol that permits a physical address, such as an Ethernet address, to be translated into an IP address. Hosts such as diskless workstations often only know their hardware interface addresses, or MAC address, when booted but not their IP addresses. They must discover their IP addresses from an external source, usually a RARP server.

To obtain the I.P. address, diskless workstations broadcast their MAC address in the whole network, when the RARP server receives the request it responds the workstation with a unique I.P. address.

Internet Control Message Protocol (ICMP)

Internet Control Message Protocol (ICMP) is a protocol tightly integrated with IP. ICMP messages, delivered in IP packets, are used for sending *error & control messages* i.e. information about the status of the network itself. Since ICMP uses IP, ICMP packet delivery is unreliable, so hosts can't count on receiving ICMP packets for any network problem. Some of ICMP's functions are to:

- **Announce network errors,** such as a host or entire portion of the network being unreachable, due to some type of failure. A TCP or UDP packet directed at a port number with no receiver attached is also reported via ICMP.
- **Announce network congestion.** When a router begins buffering too many packets, due to an inability to transmit them as fast as they are being received, it will generate ICMP *Source Quench* messages. Directed at the sender, these messages should cause the rate of packet transmission to be slowed. Of course,

generating too many Source Quench messages would cause even more network congestion, so they are used sparingly.

- **Assist Troubleshooting.** ICMP supports an *Echo* function, which just sends a packet on a round--trip between two hosts. Ping, a common network management tool, is based on this feature. Ping will transmit a series of packets, measuring average round--trip times and computing loss percentages.
- **Announce Timeouts.** If an IP packet's TTL (Time To Live) field drops to zero, the router discarding the packet will often generate an ICMP packet announcing this fact. TraceRoute is a tool which maps network routes by sending packets with small TTL values and watching the ICMP timeout announcements.

☞ Check Your Progress 2

1. How does the HTTP protocol transfer the information on the World Wide Web?

   ……

   ……

   ……

   ……

2. Explain the working of Address Resolution Protocol (ARP).

   ……

   ……

   ……

   ……

2.6 SUMMARY

This unit began with an introduction to OSI reference model. It gave detailed information about various layers and functions of each layer of OSI reference model. The unit covers on understanding of how does the communication happen in a network. It also covered TCP/IP model. Comparison was made between OSI and TCP/IP models along with similarities and differences. Some of useful protocols of each layer of TCP/IP were described.

2.7 REFERENCES/FURTHER READINGS

1. *Computer Networks*, A. S. *Tanenbaum* 4th Edition, Practice Hall of India, New Delhi. 2003.
2. *Introduction to Data Communication & Networking,* 3rd Edition, Behrouz Forouzan, Tata McGraw Hill.
3. *Computer Networking*, J.F. Kurose & K.W. Ross, A Top Down Approach Featuring the Internet, Pearson Edition, 2003.
4. *Communications Networks*, Leon Garcia, and Widjaja, Tata McGraw Hill, 2000.
5. *www.wikipedia.org*
6. *Data and Computer Communications,* Willian Stallings, 6th Edition, Pearson Education, New Delhi.

7. Larry L. Peterson, *Computer Networks*: A Systems Approach, 3rd Edition (The Morgan Kaufmann Series in Networking).

2.8 SOLUTIONS/ANSWERS

☞ Check Your Progress 1

1. The Transport Layer is responsible for providing flow control to alleviate the issue of congestion in the data transfer. Two main methods for flow control include:

Buffering: Buffering is a form of data flow control regulated by the Transport Layer as depicted in Figure 8. It is responsible for ensuring that sufficient buffers (Temporary Memory) are available at the destination for the processing of data and that the data is transmitted at a rate that does not exceed what the buffer can handle.

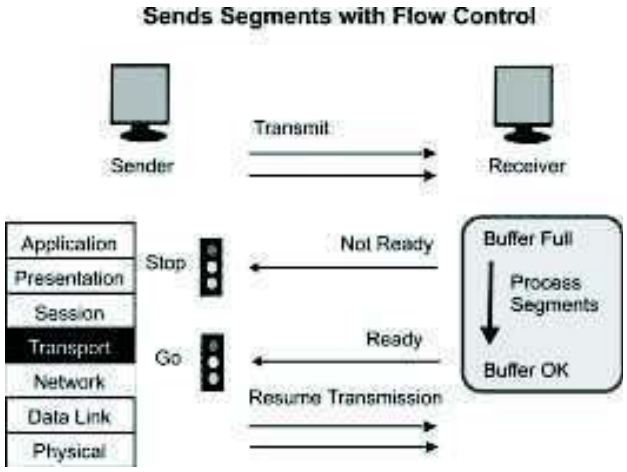


Figure 8: Buffering at Work

Windowing: Windowing is a flow control scheme in which the source computer will monitor and make adjustments to the amount of information sent based on successful, reliable receipt of data segments by the destination computer as shown in Figure 9. The size of the data transmission, called the "window size", is negotiated at the time of connection establishment, which is determined by the amount of memory or buffer that is available.

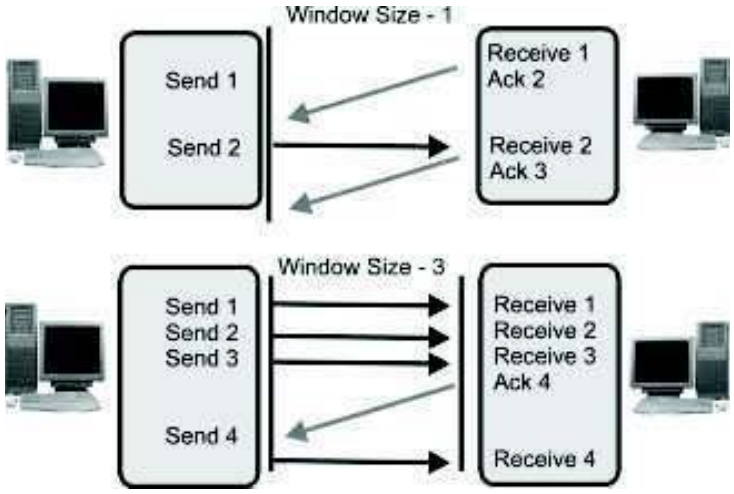


Figure 9: Flow Control & Reliability through Windowing

2. The main similarities between the OSI and TCP/IP models include the following:

   - They share similar architecture. - Both of the models share a similar architecture. This can be illustrated by the fact that both of them are constructed with layers.
   - They share a common application layer.- Both of the models share a common "application layer". However in practice this layer includes different services depending upon each model.
   - Both models have comparable transport and network layers.- This can be illustrated by the fact that whatever functions are performed between the presentation and network layer of the OSI model similar functions are performed at the Transport layer of the TCP/IP model.
   - Both models assume that packets are switched.- Basically this means that individual packets may take differing paths in order to reach the same destination.

☞ Check Your Progress 2

1. Hypertext Transfer Protocol (HTTP) is a communications protocol for the transfer of information on the intranet and the World Wide Web. Its original purpose was to provide a way to publish and retrieve hypertext pages over the Internet.

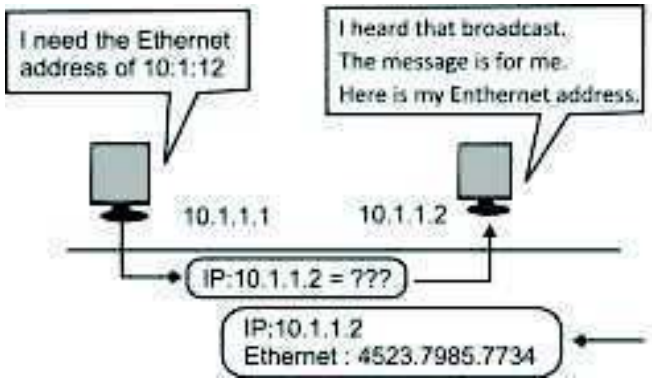
   HTTP is a request/response standard between a client and a server. A client is the end-user, the server is the web site. The client making an HTTP request - using a web browser, spider, or other end-user tool - is referred to as the user agent. The responding server - which stores or creates resources such as HTML files and images - is called the origin server. In between the user agent and origin server may be several intermediaries, such as proxies, gateways, and tunnels. HTTP is not constrained to using TCP/IP and its supporting layers, although this is its most popular application on the Internet. Indeed HTTP can be "implemented on top of any other protocol on the Internet, or on other networks. HTTP only presumes a reliable transport; any protocol that provides such guarantees can be used."

   Typically, an HTTP client initiates a request. It establishes a Transmission Control Protocol (TCP) connection to a particular port on a host (port 80 by default; see List of TCP and UDP port numbers). An HTTP server listening on that port waits for the client to send a request message. Upon receiving the request, the server sends back a status line, such as "HTTP/1.1 200 OK", and a message of its own, the body of which is perhaps the requested file, an error message, or some other information.

2. The address resolution protocol is a protocol used by the Internet Protocol (IP), specifically IPv4 (IP version 4), to map IP network addresses to the hardware addresses used by a data link protocol. It is used when IPv4 is used over Ethernet. ARP works on Ethernet networks as follows. Ethernet network adapters are produced with a physical address embedded in the hardware called the Media Access Control (MAC) address.

   Manufacturers take care to ensure these 6-byte (48-bit) addresses are unique, and Ethernet relies on these unique identifiers for message delivery. When any device wishes to send data to another target device over Ethernet, it must first determine the MAC address of that target given its IP address These IP-to-MAC address mappings are derived from an **ARP cache** maintained on each device.

If the given IP address does not appear in a device's cache, that device cannot direct messages to that target until it obtains a new mapping. To do this, the initiating device first sends an *ARP request broadcast message* on the local subnet. The host with the given IP address sends an *ARP reply* in response to the broadcast, allowing the initiating device to update its cache and proceed to deliver messages directly to the target.



UNIT 3 PHYSICAL AND DATA LINK LAYER

Structure **Page Nos.**

3.0 Introduction 42
3.1 Objectives 42
3.2 Physical and Data Link Layer Services 42
3.3 Error Detection and Correction 44
3.4 Flow and Error Control 48
3.5 Medium Access Control (MAC) Sublayer 51
3.5.1 Contention based media access protocols
3.5.2 Random access protocols
3.5.3 Polling based MAC protocols
3.5.4 IEEE standard 802.3 and Ethernet
3.5.5 IEEE standard 802.4 token bus
3.5.6 IEEE standard 802.5 token ring
3.5.7 Address resolution protocol (ARP)
3.5.8 Reverse address resolution protocol (RARP)
3.6 Summary 55
3.7 References/Further Reading 56
3.7 Solutions/Answers 56

3.0 INTRODUCTION

As you have studied earlier that the physical layer provides an electrical, mechanical, and functional interface to the transmission medium also the data link layer together with physical layer provide a data link connection for reliable transfer of data bits over an imperfect physical connection, between two adjacent nodes. In this unit, we will study about design of Data Link Layer and its Medium Access Control Sublayer. This includes various protocols for achieving reliable, efficient communication. It also covers the study of nature of errors, causes and how they can be detected and corrected. The MAC sublayer contains protocols which determine what goes next on a multi-access channel. In the end of this unit you will learn about working of ARP and RARP protocols.

3.1 OBJECTIVES

After going through this unit, you should be able to:

- Know the services of physical and data link layer
- Understand the concept of framing
- Understand various error handling methods;
- Know the Retransmission Strategies at data link layer
- Understand various flow control methods,
- Understand the working of MAC sub-layer protocols
- Differentiate between CSMA/CD, Polling and Token Passing.
- Understand the working of ARP and RARP

3.2 PHYSICAL AND DATA LINK LAYER SERVICES

To exchange digital information between devices A and B, we require an interconnecting transmission medium to carry the electrical signals; a standard interface and the physical layer to convert bits into electrical signals and vice-versa.

This is an elementary layer below the logical data structures of the higher level functions in a network. The physical layer deals with transmitting raw bits rather than logical data packets over a physical network. The bit stream may be grouped into code words or symbols and converted to an electrical signal that is transmitted over a hardware transmission medium.
The physical layer provides an electrical, mechanical, and functional interface to the transmission medium. This layer has certain limitations, for example assume:

- If the electrical signal gets impaired due to the encountered interference with other signals or electromagnetic waves from external sources, errors may be introduced in the data bits.
- Errors can also be introduced if the receiving device is not ready for the incoming signal, hence resulting in the loss of some information.

The data link layer constitutes the second layer of the hierarchical OSI Model. The Data Link layer together with physical layer provide a data link connection for reliable transfer of data bits over an imperfect physical connection, between two adjacent nodes. It accomplishes this task by having the sender break the input data into data frames, transmit the frames sequentially and process the acknowledgement frames sent back by the receiver. Remember, like other layers of OSI model this layer also create its own protocol data unit. Data link layer add some control bits to the protocol data unit received from network layer and convert into different protocol data unit called frames. The data link layer creates and recognises frame boundaries too.

Another issue that arises in data link layer is how to keep a fast transmitter from overflowing a slow receiver in data. The data link layer (Figure 1) incorporates certain processes, which carry out error control, flow control and the associated link management functions. The data block along with the control bits is called a frame.

Data link layer (Figure 1) is divided into two sublayers:

Logical Link Control (LLC) concerned with providing a reliable communication part between two devices. It is also involved with flow control and sequencing. The LLC is non-architecture-specific and is the same for all IEEE defined LANs.

Medium Access Control (MAC) focuses on methods of sharing a single transmission medium.

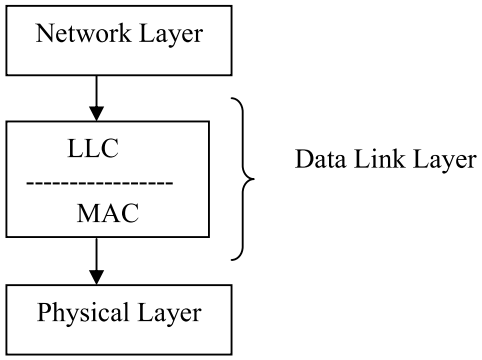


Figure 1: Division of Data Link Layer

The data link layer provides the functional means to transfer data between network entities and might provide the means to detect and possibly correct errors that may occur in the physical layer. Following are some of the main services provided by data

Link layer:

1. **Framing:** Encapsulation of network layer data packets into frames, and Frame synchronization
2. **Flow Control:** Flow control deals with how to keep the fast sender from overflowing a slow receiver by buffering and acknowledgement procedures. This flow control at data link layer is provided in addition to the one provided on the transport layer.
3. **Error detection and correction codes:** Various methods used for error-detection and corrections are – Parity bit, cyclic redundancy check, checksum, Hamming code, etc.
4. Multiple access protocols for channel-access control
5. Physical addressing (MAC addressing)
6. Quality of Service (QoS) control

3.3 ERROR DETECTION AND CORRECTION

Data that is either transmitted over communication channel or stored in memory is not completely error free. Transmission Errors may be caused by many reasons like Signal distortion or attenuation, synchronization problems, distorted channel, etc. Error detection and corrections are two different but related thing, error detection is the ability to detect errors but the error correction has an additional feature that enables identification and correction of the errors. Error detection always precedes error correction. Both can be achieved by having extra/ redundant/check bits in addition to data deduce that there is an error.

Error Detection

In the following section parity bit and CRC methods for error detection are discussed.

Parity bits Method

Parity bit method is very simple error detection method in the digital communication. A binary digit called “parity” is used to indicate whether the number of bits with “1” in a given set of bits is even or odd. The parity bit is then attached to original bits. In this method sender adds the parity bit to existing data bits before transmission. At the receiver side, it checks for the expected parity, if wrong parity found, the received data is discarded and retransmission is requested. It is a very simple scheme that can be used to detect single or any other odd/even number of errors in the output.

The parity bit is only suitable for detecting errors; it cannot correct any errors, as there is no way to determine which particular bit is corrupted. The data must be discarded entirely, and re-transmitted from scratch. Following are some of the examples for parity bit methods:

Assume, sender wants to send some bit streams like 001 0101 and 101 0011. If we are using even parity bit method, we will add “0” with the bit steam having even number of 1’s otherwise add “1”. So our bit steams will be changed after adding parity bit as **1**001 0101and **0**101 0011. At the receiver again the number of 1’s are counted in the original message, if the parity bit is mismatched we can say an error has occurred in the message. Just like the even parity we may have odd parity bit method. Parity bit method has many limitations, like it cannot identify the error if more than one bit has been changed or parity bit itself has been changed during the transmission. Further it cannot determine which bit position has a problem.

Cyclic redundancy checks (CRCs)

A cyclic redundancy check (CRC) is an error-detecting code commonly used in digital networks and storage devices to detect transmission error.

When n-bits of message M(x) is transmitted from sender to receiver, first the n- bits of message is converted in such a way that when a selected k-bits divisor code G(k) (so-called generator polynomial) is divided with the x+k-bits message M(x+k) the remainder is zero.

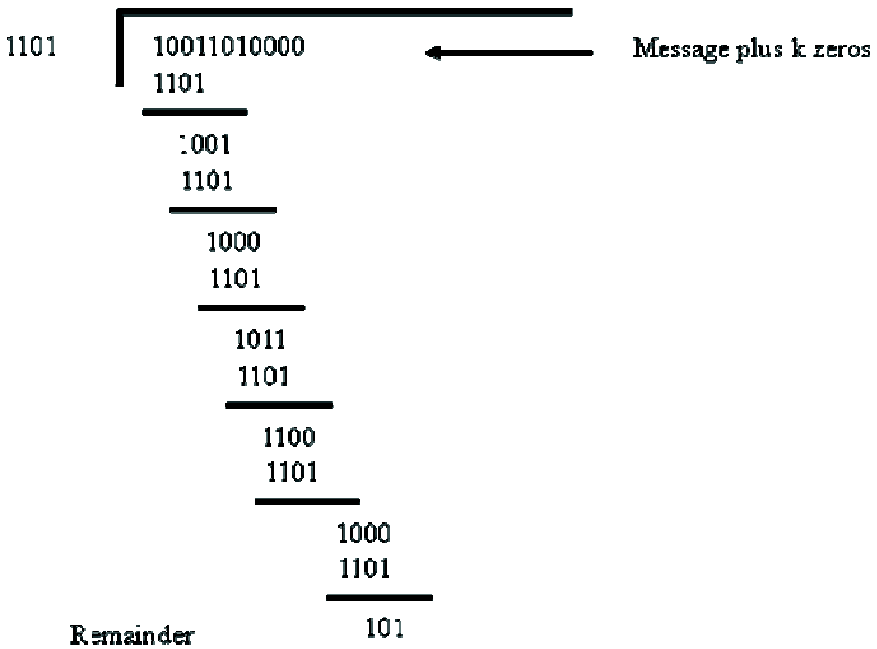
Than the modified message M(x+k) is sent along with the k-bits divisor code to the receiver through channel. The receiver will divide this M(x+k) bits with G(k) bits, if the remainder is zero receiver can say there is no error in the message. Finally the original message M(x) is separated from the modified message M(x+k).

Let us take assume an example for simple decimal numbers, if you want to send some number say 10 and divisor code is 3. First, make all legal messages divisible by 3. For that you need to multiply by 4 to get 40 and add 2 to make it divisible by 3 = 42. When the data is received and divided by 3, and if there is no remainder, it means there is no error. If no error, divide by 4 and separate it by 2 to get sent message. If we receive 43, 44, 41, 40, we can say there is an error. But if 45 is received, we will not be able to recognize as an error.

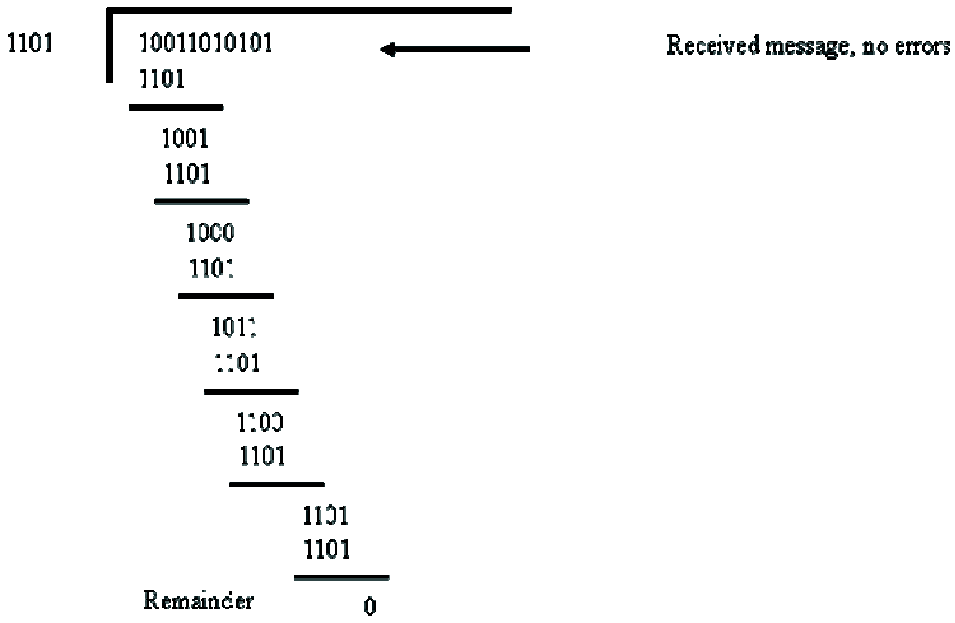
We can represent n-bit message as an n-1 degree polynomial; e.g., M=10011010 corresponds to M(x) = x7+ x4 + x3 + x1.

Add k bits of extra data to an n-bit message.
Let k be the degree of some divisor polynomial G(k); e.g., G(k) = x3+ x2 + 1.

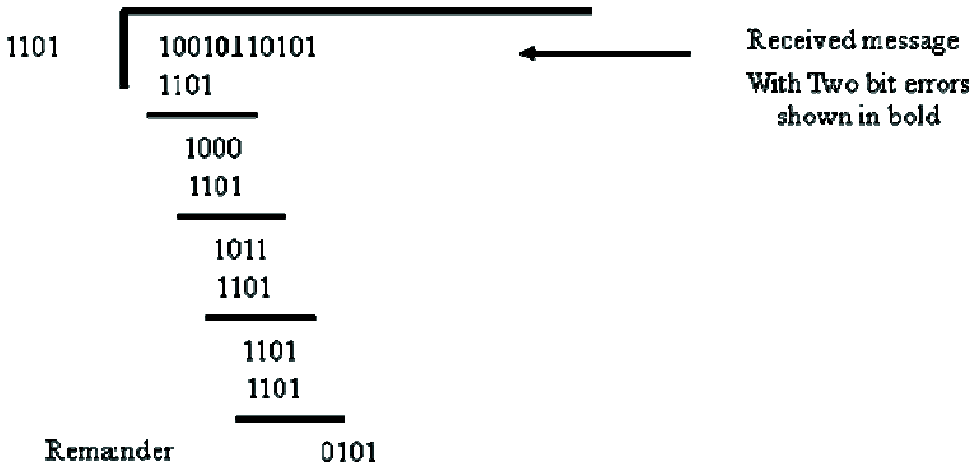
Multiply M(x) = x7+ x4 + x3 + x1 by x^k; for our example, we get
x10 + x7 + x6 + x4 (10011010000);divide result by G(k) (1101);



Send 10011010000 + 101 = 10011010101, since this must be exactly divisible by G(k);



Now, assume if receiver will receive a message with errors, for example receiver has received a message 10010110101.



Cyclic codes have favorable properties in that they are well suited for detecting burst errors. CRCs are particularly easy to implement in hardware, and are therefore commonly used in digital networks and storage devices such as hard disk drives.

Error correction

Mainly, we have two error correction mechanisms one is Automatic Repeat request and another approach is of using some error correction codes like hamming code.

Automatic Repeat Request

It is an error control method for data transmission that makes use of error-detection codes, acknowledgment and/or negative acknowledgment messages, and timeouts to get reliable data transmission. Generally, when the sender does not receive the acknowledgment before the timeout occurs, it retransmits the frame until it is either correctly received or the error persists beyond a predetermined number of retransmissions. Three types of ARQ protocols are Stop-and-wait ARQ, Go-Back-N ARQ, and Selective Repeat ARQ, these mechanisms we will study further in this unit.

Error-correcting codes

Any error-correcting code can be used for error correction. An error-correcting code is a system of adding redundant data, or parity data, to a message, such that it can be recovered by a receiver even when a number of errors were introduced, either during the process of transmission, or on storage. Since the receiver does not have to request the sender for retransmission of the data, a back-channel is not required in forward error correction, and it is therefore suitable for simplex communication such as broadcasting. Error-correcting codes are often used in lower layers of OSI like data link layer and physical layer.

Error-correcting codes can be classified into two type's convolutional codes which processed on a bit-by-bit basis and block codes that processed on a block-by-block basis. Convolutional codes are suitable for implementation in hardware. However, block codes are error correction in data communication. Hamming code is an example of block codes. Hamming codes are code words formed by adding redundant check bits, or parity bits, to a data word. The Hamming distance between two code words is the number of bits in which two code words differ. For an example 10001001 and 10110001 bytes has a Hamming distance of 3. The minimum Hamming distance for a code is the smallest Hamming distance between all pairs of words in the code. The minimum Hamming distance for a code, D(min), determines its error detecting and error correcting capability. Hamming codes can *detect* D(min) - 1 errors and correct (D(min) - 1)/2 errors.

☞ Check Your Progress 1

1. What are the sub-layers of data link layer? Explain.

   ……

   ……

   ……

   ……

2. List the services of data link layer.

   ……

   ……

   ……

   ……

3. What is parity bit method? Explain its use with the help of an example.

   ……

   ……

   ……

4. Explain the use of Automatic Repeat Request in error correction.

   ……

   ……

   ……

3.4 FLOW AND ERROR CONTROL

Packets can be lost and/or corrupted during transmission due to Bit level errors and loss due to congestion. We use checksums to detect bit level errors, and to maintain reliability into the data transmission stage we use *acknowledgements* and *timeouts* to signal lost or corrupt frame. An acknowledgement (ACK) is a packet sent by one host in response to a packet it has received. A timeout is a signal that an ACK to a packet that was sent has not yet been received within a specified timeframe. In this section we will discuss several retransmission strategies, which are also considered as a flow control and error control mechanism.

Stop and Wait

The sender allows one message to be transmitted, checked for errors and an appropriate ACK (Positive Acknowledgement) or NAK (Negative Acknowledgement) returned to the sending station. No other data messages can be transmitted until the receiving station sends back a reply, thus the name STOP and WAIT is derived from the originating station sending a message, stopping further transmission and waiting for a reply. This scheme is also shown in figure 2 given below.

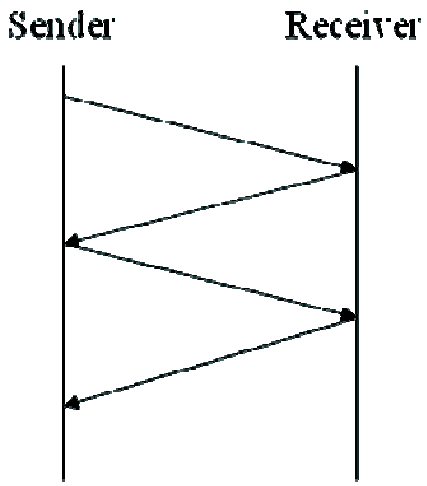


Figure 2: Stop and Wait Protocol

Its major drawback is the idle line time that results when the stations are in the waiting period. If the ACK is lost then the sending station retransmits the same message to the receiver side. The redundant transmission could possibly create a duplicate frame. A typical approach to solve this problem is the provision for a sequence number in the header of the message. The receiver can then check for the sequence number to determine if the message is a duplicate. The Stop and Wait mechanism requires a very small sequence Number, since only one message is outstanding at any time. The sending and receiving station only use a one bit alternating sequence of 0 and 1 to maintain the relationship of the transmitted message and its ACK/NAK status.

Sliding Window

Here data and control frames flow from sender to receiver in a more continuous manner and several frames can be outstanding at any one time as depicted in figure 3. Allow multiple outstanding (un-ACKed) frames. Upper bound on un-ACKed frames, called window. Sender needs to buffer data so that if data is lost, it can be resent.
Receiver needs to buffer data so that if data is received out of order, it can be held until all packets are received in Flow control Next,
How can we prevent sender overflowing receiver's buffer?
Receiver tells sender its buffer size during connection setup.

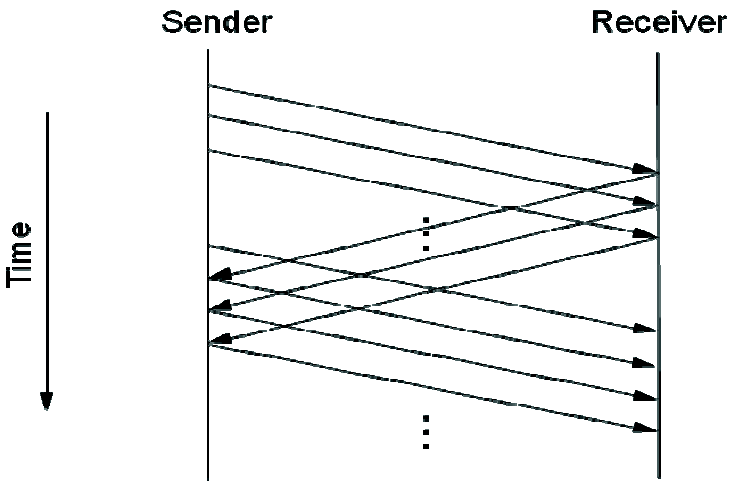


Figure 3: Simple Sliding Window Scheme

The transmitting station maintains a sending window that maintains the number of frames it is permitted to send to the receiving station and the receiving station also maintains a receiving window that performs complementary functions. The two sides use the window to coordinate the flow of frames between each other. The window wrap around is used to reuse the same set of numbers for different frames. There are sliding window techniques:

1. Go Back N
2. Selective Repeat

Go Back N

This is a sliding window technique as shown in figure 4. It allows data and control messages to be
transmitted continuously without waiting for its acknowledgement from the receiver. In the event of error detection at the receiving side, the erroneous message is retransmitted, as well as all other frames that were transmitted after the erroneous message.

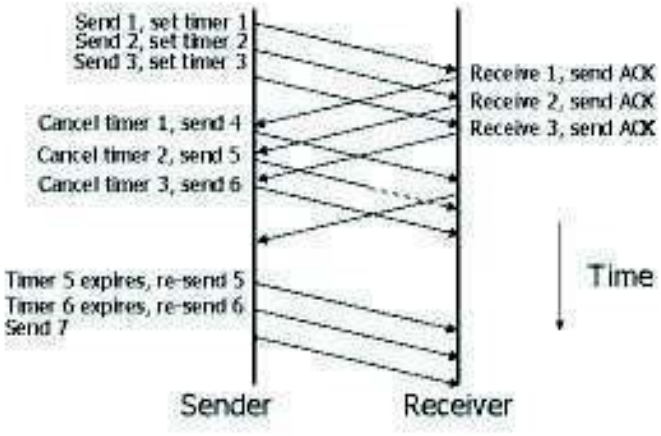


Figure 4: Go Back N Scheme

Sender has to buffer all unacknowledged packets, because they may require retransmission. Receiver may be able to accept out-of-order packets, but only up to its buffer limits. The sender needs to set timers in order to know when to retransmit a packet that may have been lost

Selective Repeat

This method provides for a more refined approach. In contrast to the Go back N, the only messages retransmitted are those for which negative acknowledgement is received. In this the sending process continues to send a number of frames specified by a window size even after a frame loss. Unlike Go-Back-N, the receiving process will continue to accept and acknowledge frames sent after an initial error; this is the general case of the sliding window protocol with both transmit and receive window sizes greater than 1.

The receiver process keeps track of the sequence number of the earliest frame it has not received, and sends that number with every acknowledgement (ACK) it sends. If a frame from the sender does not reach the receiver, the sender continues to send subsequent frames until it has emptied its window. The receiver continues to fill its receiving window with the subsequent frames, replying each time with an ACK containing the sequence number of the earliest missing frame. Once the sender has sent all the frames in its window, it re-sends the frame number given by the ACKs, and then continues where it left off.

Now if we compare Selective Repeat behaves in the same way like Go-Back-N , it accepts when the receiver receives a frame which is out of sequence, it sends a SREJ(Selective Reject) message. Sender retransmits only the rejected packet and continues with other packets. Here in Selective Repeat method the both the Sender's and Receiver's buffer size are equal to the window size.

In the following figure 5, you can see that the difference between Go Back N and Selective Repeat, because of the buffer frame 5 and Frame 6 are stored and selectively the reject message is sent only for frame 4 (which was lost in transmission) however in Go back N the reject message is sent for all 4, 5 and 6 frames.

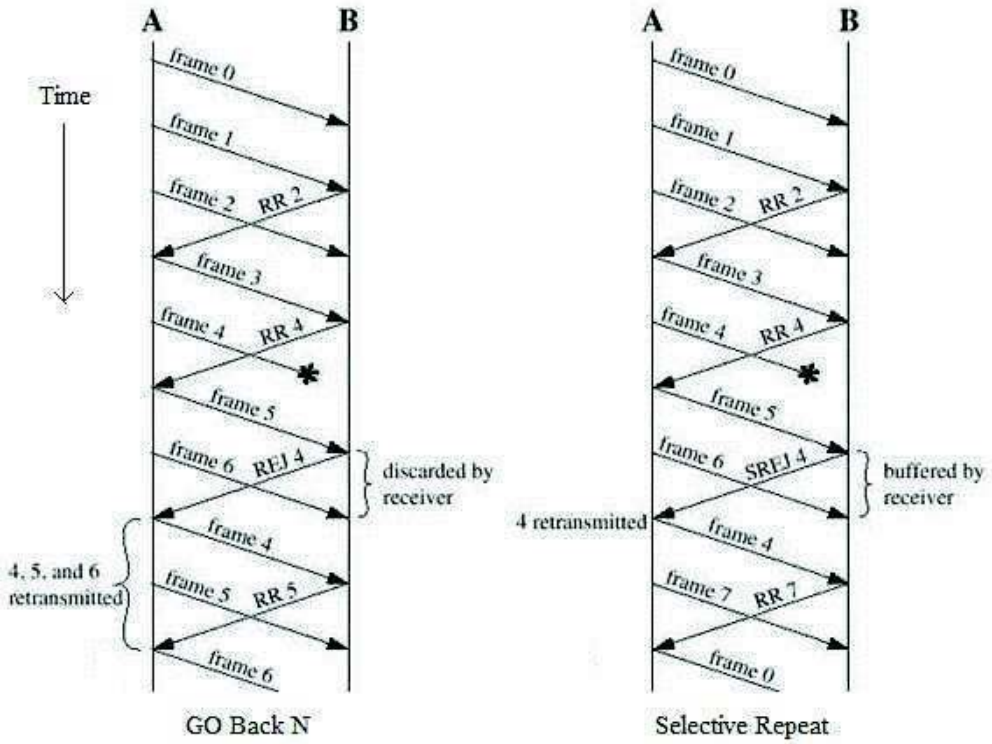


Figure 5: Comparison between the Go Back N and Selective Repeat method

Studies reveal that the selective repeat mechanism produces greater throughput than the Go Back N. Selective Repeat mechanism requires additional logic to maintain the sequence of the recent message and merge it into the proper place as the queue at the receiver end.

☞ Check Your Progress 2

1. Explain the importance of Sliding Window protocol. Also, List the types of sliding window techniques.

   ……

   ……

   ……

   ……

2. Discuss the working of selective Repeat method. Also, compare it with GO Back N.

   ……

   ……

   ……

   ……

3.5 MEDIUM ACCESS CONTROL (MAC) SUBLAYER

In any broadcast network, key issue is how to determine who gets to use the channel when there is competition for it. The protocols used to determine who goes next on a multi-access channel belong to a sub-layer of a Data Link Layer called MAC sublayer.

3.5.1 Contention Based Media Access Protocols

Contention is what happens at a staff meeting when several people start to speak at the same time. In contention protocol, no one controls usage of the communication channel.

All workstations on a contention network share a common transmission channel. Messages are broadcasted on that channel and may be overheard by all attached workstations. A workstation responds only to message with its address. Message intended for other nodes are ignored.

Message to be transmitted are converted to packets and are sent when ready, without verifying the availability of the channel. When transmission of a station overlaps with that of another, collision occurs. Colliding packets with their messages are destroyed.

3.5.2 Random Access Protocols

In random access approach, any station is not superior to another station and none is assigned the control over another. A station with a frame to be transmitted can use the link directly based on a procedure defined by the protocol to make a decision on whether or not to send.

Pure ALOHA

It is based on simple principles that if you have data to send, send the data immediately. If the message collides with another transmission, after some random time wait, we can resend it message. In this, all frames from any station are of fixed length size and produce frames with equal frame lengths. A station that has data can transmit at any time, after transmitting a frame, the sender waits for an acknowledgment for an amount of time. If ACK was not received, sender assumes that the frame or ACK has been destroyed and resends that frame after it waits for a random amount of time.

Slotted ALOHA

Slotted ALOHA is an improvement over pure ALOHA, which has discrete timeslots. A station is allowed to send the message only at the beginning of a timeslot, due to time the possibility of collisions are reduced. If a station misses the beginning of a slot, it has to wait until the beginning of the next time slot. A central clock or station informs all stations about the start of an each slot.

Channel utilization or efficiency or Throughput is the percentage of the transmitted frames that arrive successfully (without collisions) or the percentage of the channel bandwidth that will be used for transmitting frames without collisions.

The throughput (S) for pure ALOHA is $S = G \times e^{-2G}$. The maximum throughput is $S_{max} = 0.184$ when G= (1/2). Where, G is equal to the traffic load. In case of Slotted ALOHA the throughput is $S = G \times e^{-G}$ and the maximum throughput is $S_{max} = 0.368$ when G = 1. The following figure 6 shows the different between pure and slotted ALOHA based on the traffic load and throughput.

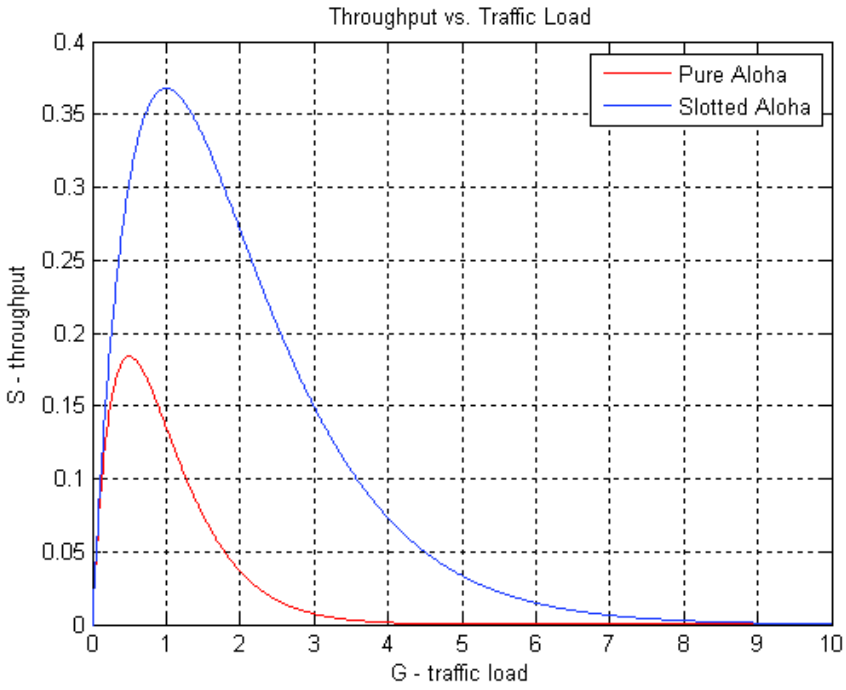


Figure 6: Different between pure and slotted ALOHA (source wikipedia.org)

CSMA/CD

Before discussing about CSMA/CD (Carrier Sense Multiple Access with Collision Detection), let us first discuss about simple CSMA. Carrier Sense Multiple Access

(CSMA) is a MAC layer protocol in which a node verifies the absence of other traffic before transmitting on a shared transmission medium. Here, the Carrier Sense means the fact that a transmitter uses feedback from a receiver before trying to send any message. If a carrier is sensed, the station waits for the transmission in progress to finish before initiating its own transmission. And the Multiple Access means that multiple stations are sending and receiving on the same medium. Based on different situations of medium like medium busy or idle different CSMA protocols has been designed like non-Persistent CSMA, 1-Persistent CSMA and p-Persistent CSMA. All these types of CSMA have inefficiency in term of collision detection. Assume that a collision has occurred, than the channel is unstable until colliding packets have been fully transmitted. A standards and rules need to be created for stations like when they could send data and when they could not.

This standard in CSMA is Carrier Sense Multiple Access with Collision Detection, referred to as CSMA/CD.

To avoid collision, CSMA/CD compel stations to “listen” to the channel before sending in order to make sure that no other host on the wire is sending. When the channel is not busy, station may send its data. The sender will then continue to listen, to make sure that sending the data have not caused a collision. If a collision is heard, senders will send a jam signal over the network. This jam signal indicates to all other devices on the network segment that there has been a collision, and they should not send data onto the channel. After sending the jam signal, each of the senders will wait a random amount of time before beginning the entire process over. CSMA/CD (Carrier Sense Multiple Access with Collision Detection) While reducing channel wastage. It is widely used for bus topology LANs (IEEE 802.3, Ethernet).

3.5.3 Polling based MAC Protocols

Polling involves the channel control of all workstations in a network. The primary workstation which acts like a teacher going down the rows of the class room asking each student for homework. When one student has answered, the next is given a chance to respond. A polling network contains two classes of workstations, the primary workstation and the multiple secondary workstations connected to it. A buffer that can temporarily store messages is associated with each secondary workstation. When a workstation has information to transmit, the data is passed to the buffer. The frames are held until the central controller polls the workstation.

Following are two possibilities for the path of a message from some to destination workstation:

- All messages may be required to pass to the central workstation, which route them to their destination.
- Messages may be sent directly.

Polling technique can be said to maintain a tight control over the network resources than do contention based protocols.

Token Passing

The network continuously circulates a special bit pattern known as a token among all the nodes in the network.

Each token contains network information, comprising of a header, a data field and a trailer. Any node willing to send a frame has to grab a token first. After a node has captured a token it transmits its frame. The frame is relayed by all intermediate nodes till it reaches destination, when it is copied. Now let us talk about some standards.

3.5.4 IEEE Standard 802.3 and Ethernet

It uses CSMA/CD mechanism Expand (carrier Seen Multiple Access/Collision Detect). When station wants to transmit, it listens to the cable. If the cable is busy, the station waits until it goes idle, otherwise it transmits immediately. If two or more stations simultaneously begin transmitting on an idle cable they will collide. All colliding stations then terminate their transmissions, wait a random time and repeat the whole process all over again.

3.5.5 IEEE Standard 802.4 Token Bus

Token bus combines features of Ethernet and token ring (discussed in the next section). It combines the physical configuration of Ethernet (bus topology) and collision free (predictable delay) feature of token ring. Token bus is a physical bus that operates as logical ring using tokens.

It is a linear cable onto which the stations are attached. When the logical ring is initialised, the highest numbered station may send the first frame after it is done, it passes permission to its immediate neighbour by sending the neighbour a special control frame called a token.

The token propagates around the logical ring with only the token holder being permitted to transmit frames. Since only one station at a time holds the token, collisions do not occur.

3.5.6 IEEE Standard 802.5 Token Ring

In a token ring, the token circulates around the ring whenever all stations are idle. When a station wants to transmit a frame, it is required to seize the token and remove it from the ring before transmitting. This action is done by inverting a single bit in the 3-byte token which instantly changes it into the first 3 bytes of a normal data frame. Because there is only one token, only one station can transmit at a given instant, thus solving the channel access problem.

3.5.7 Address Resolution Protocol (ARP)

We have seen that IP address makes the addressing uniform on the Internet. Routing of packets is done using the IP addresses of the packet. However, communication in a local network is broadcast, which is done using physical address. Therefore, when the packet reaches the destined network, there must be a process of obtaining the physical address corresponding to its IP address, of a computer in order to finally deliver the datagram to the destined computer. The physical address corresponding to an IP address is resolved by using address resolution protocol (ARP). ARP maps given IP address to a physical address as shown in the Figure 7. It takes host's IP address as input and gives its physical address as output.



Figure 7: ARP maps the IP address to the physical address

ARP assumes that every host knows its IP address and physical address. Any time a host needs to know the physical address of another host on the network, it creates an ARP packet that includes the IP address X of the destination host asking–Are you the one whose IP address is X? If yes, please send back your physical address. This packet is then broadcasted over the local network. The computer, whose IP address matches X, sends an ARP reply packet, with its physical address. All the other hosts ignore the broadcast. Next time the host needs to send a datagram to the same destination, it need not broadcast an ARP query datagram; instead it can look up in its ARP cache. If the mapping is not found in the cache, then only the broadcast message is sent.

3.5.8 Reverse Address Resolution Protocol (RARP)

This protocol performs the job exactly opposite to ARP. It maps a physical address to its IP address as shown in Figure 8. Where is this needed? A node is supposed to have its IP address stored on its hard disk. However, there are situations when the host may not have hard disk at all, for example a diskless workstation. But also, when a host is being connected to the network for the first time, at all such times, and a host does not know its IP address. In that case, RARP find out the IP address, this process is shown in Figure 8.

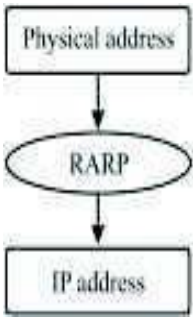


Figure 8: RARP maps the physical address to the IP address

☞ Check Your Progress 3

1. Compare the Throughput of pure and slotted ALOHA.

   ……

   ……

   ……

2. Explain the need of RARP.

   ……

   ……

3.6 SUMMARY

After studying this unit, we are sure that you understood the services and protocol of data link layer. Essentially it provides the functional means to transfer data between network entities and might provide the means to detect and possibly correct errors that may occur in the physical layer. We have briefly discussed various methods used for error-detection and corrections are – Parity bit, cyclic redundancy check, Hamming code, etc. In this unit you have studied some flow control and error control mechanism to ensure the reliability of communication. In this unit you have studied sliding window mechanisms mainly used for flow control at data link layer. As you know that the key issue is how to determine who gets to use the channel when there is

competition for it. In this unit, we have studied the protocols used to determine who goes next on a multi-access channel. In the end of this unit we have studied address resolution protocols to map between IP addresses and the physical addresses of the machines.

3.7 REFERENCES/FURTHER READING

1. *Computer Networks*, A. S. *Tanenbaum* 4th Edition, Practice Hall of India, New Delhi. 2003.
2. *Introduction to Data Communication & Networking,* 3rd Edition, Behrouz Forouzan, Tata McGraw Hill.
3. *Computer Networking*, J.F. Kurose & K.W. Ross, A Top Down Approach Featuring the Internet, Pearson Edition, 2003.
4. *Communications Networks*, Leon Garcia, and Widjaja, Tata McGraw Hill, 2000.
5. *Data and Computer Communications,* Willian Stallings, 6th Edition, Pearson Education, New Delhi.
6. www. wikipedia.org
7. Larry L. Peterson, *Computer Networks*: A Systems Approach, 3rd Edition (The Morgan Kaufmann Series in Networking).

3.8 SOLUTIONS/ANSWERS

☞ Check Your Progress 1

1. Data link layer is divided into two sublayers LLC and MAC. **Logical Link Control** (LLC) concerned with providing a reliable communication part between two devices. It is also involved with flow control and sequencing. The LLC is non-architecture-specific and is the same for all IEEE defined LANs. **Medium Access Control (MAC)** focuses on methods of sharing a single transmission medium.

2. Following are services provided by data link layer:
   - i) **Framing:** Encapsulation of network layer data packets into frames, and Frame synchronization
   - ii) **Flow Control:** Flow control deals with how to keep the fast sender from overflowing a slow receiver by buffering and acknowledgement procedures. This flow control at data link layer is provided in addition to the one provided on the transport layer.
   - iii) **Error detection and correction codes:** Various methods used for error-detection and corrections are – Parity bit, cyclic redundancy check, checksum, Hamming code, etc.
   - iv) Multiple access protocols for channel-access control
   - v) Physical addressing (MAC addressing)
   - vi) Quality of Service (QoS) control

3. Parity bit method is very simple error detection method in the digital communication. A binary digit called “parity” is used to indicate whether the number of bits with “1” in a given set of bits is even or odd. The parity bit is then attached to original bits. Assume sender want to send some bit streams like 001 0101 and 101 0011. If we are using even parity bit method, we will add “0” with

the bit steam having even number of 1's otherwise add "1". So our bit steams will be changed after adding parity bit as **1**001 0101and **0**101 0011. At the receiver again the number of 1's are counted in the original message, if the parity bit is mismatched we can say an error has occurred in the message. Just like the even parity we may have odd parity bit method.

4. It is an error control method for data transmission that makes use of error-detection codes, acknowledgment and/or negative acknowledgment messages, and timeouts to get reliable data transmission. Generally, when the sender does not receive the acknowledgment before the timeout occurs, it retransmits the frame until it is either correctly received or the error persists beyond a predetermined number of retransmissions. Three types of ARQ protocols are Stop-and-wait ARQ, Go-Back-N ARQ, and Selective Repeat ARQ, these mechanisms we will study further in this unit.

☞ Check Your Progress 2

1. In Sliding Window data and control frames flow from sender to receiver in a more continuous manner and several frames can be outstanding at any one time. Allow multiple outstanding (un-ACKed) frames. Upper bound on un-ACKed frames, called window. Sender needs to buffer data so that if data is lost, it can be resent. Receiver needs to buffer data so that if data is received out of order, it can be held until all packets are received Flow control. The transmitting station maintains a sending window that maintains the number of frames it is permitted to send to the receiving station and the receiving station also maintains a receiving window that performs complementary functions. The two sides use the window to coordinate the flow of frames between each other. The window wrap around is used to reuse the same set of numbers for different frames.

   There are sliding window techniques:

   Go Back N
   Selective Repeat

2. This method provides for a more refined approach. In contrast to the Go back N, the only messages retransmitted are those for which negative acknowledgement is received. In this the sending process continues to send a number of frames specified by a window size even after a frame loss. Unlike Go-Back-N, the receiving process will continue to accept and acknowledge frames sent after an initial error; this is the general case of the sliding window protocol with both transmit and receive window sizes greater than 1.

   The receiver process keeps track of the sequence number of the earliest frame it has not received, and sends that number with every acknowledgement (ACK) it sends. If a frame from the sender does not reach the receiver, the sender continues to send subsequent frames until it has emptied its window. The receiver continues to fill its receiving window with the subsequent frames, replying each time with an ACK containing the sequence number of the earliest missing frame. Once the sender has sent all the frames in its window, it re-sends the frame number given by the ACKs, and then continues where it left off.

   Now if we compare Selective Repeat behaves in the same way like Go-Back-N , it except when the receiver receives a frame which is out of sequence, it sends a SREJ(Selective Reject) message. Sender retransmits only the rejected packet and continues with other packets. Here in Selective Repeat method the both the

   Sender's and Receiver's buffer size are equal to the window size.

☞ Check Your Progress 3

1. Throughput is the percentage of the transmitted frames that arrive successfully (without collisions) or the percentage of the channel bandwidth that will be used for transmitting frames without collisions. The throughput (S) for pure ALOHA is $S = G \times e^{-2G}$. The maximum throughput is $S_{max} = 0.184$ when G= (1/2). Where, G is equal to the traffic load. In case of Slotted ALOHA the throughput is $S = G \times e^{-G}$ and the maximum throughput is $S_{max} = 0.368$ when G = 1.
2. RARP maps a physical address to its IP address. Where is this needed? A node is supposed to have its IP address stored on its hard-disk. However, there are situations when the host may not have hard disk at all, for example a diskless workstation. But also when a host is being connected to the network for the first time, at all such times, a host does not know its IP address. In that case RARP find out the IP address.

UNIT 4 INTERNETWORKING DEVICES

Structure **Page Nos**

4.0 Introduction 58
4.1 Objectives 58
4.2 Internetworking Devices 58
- 4.2.1 Network interface card
- 4.2.2 Modem (modulator/demodulator)
- 4.2.3 Repeaters
- 4.2.4 Hubs
- 4.2.5 Bridges
- 4.2.6 Switch
- 4.2.7 Gateway

4.3 Summary 69
4.4 References/Further Readings 69
4.5 Solutions/Answers 70

4.0 INTRODUCTION

In this unit, you will learn on various internetwork devices such as NIC adapters, routers, hubs, switches, modems, gateway and other related devices. A network is consists of a larger number of the communication devices. The simplest device that is used in the communication is the NIC adapter which is attached with the every computer in a network. If you want to build a LAN, you will need to have computers, hubs, switches, network adapters, UTP/STP cables, routers, internal/external modems, connectors, cable testers and clipping tool. This unit explains some of mostly used network devices.

4.1 OBJECTIVES

After going through this unit, you should be able to know:

- Understand various network devices
- Functions of various network devices
- Merits and limitations of various network devices
- Difference between layer 2 and layer 3 switching, and
- Network gateway and its importance.

4.2 INTERNETWORKING DEVICES

Computer network can be established by using various network devices such as such as cables, Network Interface Cards (NICs), Modems, Repeaters, Hubs, Bridges, Switches, and Gateways. The following are various internetwork devices that are used in building LAN/WAN.

4.2.1 Network Interface Card

A network card or network interface controller or network adapter or simply NIC (network interface card) is a piece of computer hardware designed to allow computers to communicate over a computer network as shown in Figure 1. It access to a networking medium and often provides a low-level addressing system through the use of MAC addresses. It allows users to connect to each other, either by using cables or wirelessly.



Figure 1: A Network Interface Card (NIC)

Early network interface controllers were commonly implemented on expansion cards that plugged into a computer bus; the low cost and ubiquity of the Ethernet standard means that most new computers have a network interface built into the motherboard.

The network controller implements the electronic circuitry required to communicate using a specific physical layer and data link layer standard such as Ethernet, Wi-Fi, or Token Ring. This provides a base for a full network protocol stack, allowing communication among small groups of computers on the same LAN and large-scale network communications through routable protocols, such as IP.

The NIC may use one or more of four techniques to transfer data:

- **Polling** is where the CPU examines the status of the peripheral under program control.
- **Programmed I/O** is where the microprocessor alerts the designated peripheral by applying its address to the system's address bus.
- **Interrupt-driven I/O** is where the peripheral alerts the microprocessor that it is ready to transfer data.
- **Direct memory access** is where an intelligent peripheral assumes control of the system bus to access memory directly. This removes load from the CPU but requires a separate processor on the card.

4.2.2 Modem (Modulator/Demodulator)

Modem is a device that converts digital and analog signals as shown in the Figure 2. At the source, modems convert digital signals to a form suitable for transmission over analog communication facilities (public telephone lines). At the destination, modems convert the signal back to a digital format.

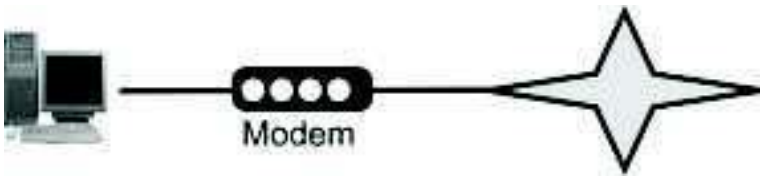


Figure 2: Modem

CSU / DSU

CSU/DSU (Channel Service Unit/Data Service Unit) is a hardware device about the size of an external modem that converts digital data frames from the communications technology used on a local area network (LAN) into frames appropriate to a wide-area network (WAN) and vice versa. A common type of device is also shown in the Figure 3. For example, if you have a Web business from your own home and have leased a digital line (perhaps a T-1 or fractional T-1 line) to a phone company or a gateway at an Internet service provider, you have a CSU/DSU at your end and the phone company or gateway host has a CSU/DSU at its end.

The Channel Service Unit (CSU) receives and transmits signals from and to the WAN line and provides a barrier for electrical interference from either side of the unit. The CSU can also echo loop back signals from the phone company for testing purposes. The Data Service Unit (DSU) manages line control, and converts input and output between RS-232C, RS-449, or V.35 frames from the LAN and the time-division multiplexed (TDM) DSX frames on the T-1 line. The DSU manages timing errors and signal regeneration. The DSU provides a modem-like interface between the computer as Data Terminal Equipment (DTE) and the CSU.

Channel service unit/data service unit (CSU/DSU) is a piece of data communications equipment that performs the following functions:

- Acts as a transceiver
- Connects data terminating equipment to dedicated circuits such as T1 and T3.
- A CSU/DSU performs multiplexing and de-multiplexing on T1 and T3 circuits.
- May have the ability to add and drop channels from a T1 or T3.
- Modern CSU/DSU's split the arriving data stream into multiple voice channels and/or multiple data channels.
- Here is a picture of the back of an external, stand-alone CSU/DSU for a T1

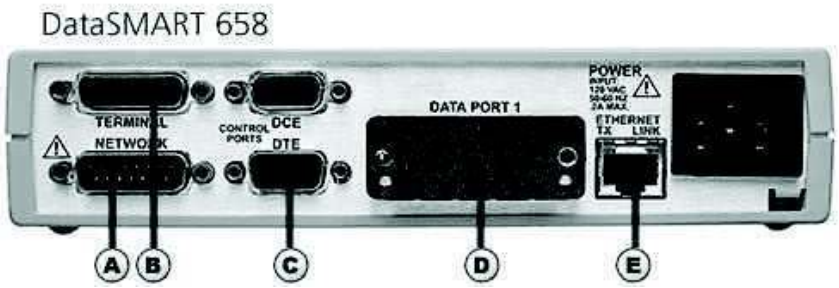


Figure 3: A CSU/DSU Device

4.2.3 Repeaters

A **repeater** is an electronic device that receives a signal and retransmits it at a higher level or higher power, or onto the other side of an obstruction, so that the signal can cover longer distances without degradation, an example is shown in the Figure.4. Because repeaters work with the actual physical signal, and do not attempt to interpret the data being transmitted, they operate on the Physical layer, the first layer of the OSI model.



Figure 4: Repeater

In telecommunication, the term **repeater** has the following standardized meanings:
An analog device that amplifies an input signal regardless of its nature (analog or digital).

A digital device that amplifies, reshapes, retimes, or performs a combination of any of these functions on a digital input signal for retransmission.

4.2.4 Hubs

A hub (concentrator) contains multiple ports as shown in Figure 5, which is used to connect devices in a star topology. When a packet arrives at one port, it is copied to all the ports of the hub. But when the packets are copied, the destination address in the frame does not change to a broadcast address. It does this in a rudimentary way; it simply copies the data to all of the Nodes connected to the hub (broadcast).



Figure 5: Hub

Advantages and Disadvantages of Hub

Following are some advantages and disadvantages of Hubs:

Advantages:

- Improves performance, especially for bursty traffic and large file transfers
- Enables optimum performance of PCI computers
- Offers ease of use: Fast Ethernet hubs require no hardware or software settings; just plug them in
- Leverages your knowledge of Ethernet and investment in management tools and applications

Disadvantages:

- Total bandwidth remains fixed; as network traffic grows, performance suffers

- The network manager cannot manage network load—for example, by segmenting the network into multiple collision domains or restricting certain types of traffic to certain ports
- Does not reduce collisions
- Requires Category 5 UTP cabling for each 100BaseTX connection

4.2.5 Bridges

The main network device found at the data link layer is a bridge. This device works at a higher layer than the repeater and therefore is a more complex device. It has some understanding of the data it receives and can make a decision based on the frames it receives as to whether it needs to let the information pass, or can remove the information from the network. This means that the amount of traffic on the medium can be reduced and therefore, the usable bandwidth can be increased.

Bridges are store and forward devices to provide error detection; a common type of bridge is shown in the Figure 6. They capture an entire frame before deciding whether to filter or forward the frame, which provides a high level of error detection because a frame's CRC checksum can be calculated by the bridge. Bridges are highly susceptible to broadcast storms. A broadcast storm occurs when several broadcasts are transmitted at the same time. It can take up huge bandwidth.

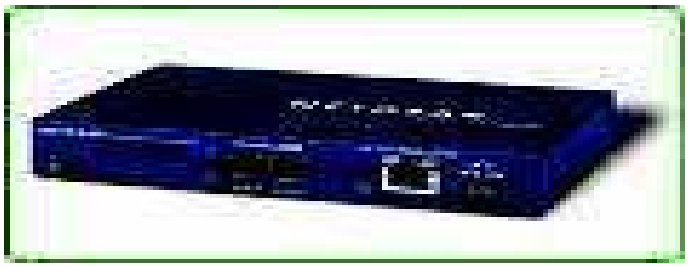


Figure 6: Bridge

Advantages and Disadvantages of Bridges

Following are some advantages and disadvantages of Bridges:

Advantages:

- Reliability
- Manageability
- Scalability

Disadvantages:

- A bridge cannot filter out broadcast traffic.
- It introduces 20 to 30 % latency.
- Only 2 networks can be linked with a bridge

☞ Check Your Progress 1

1. Explain the meaning of repeaters in analog and digital system

……

……

2. What are the advantages and disadvantages of bridges?

……

……

……

4.2.6 Switch

A switch is a data-link layer network device that forwards frames using MAC addresses in the header of frames. Common types of switches are shown in the Figure 7. It is used to improve network performance by: -

- Segmenting the network and creating separate collision domains.
- Reducing competition for bandwidth.

In a switch frame, forwarding is handled by specialized hardware called "Application Specific Integrated Circuit" (ASIC). ASIC technology allows a silicon chip to be programmed to perform specific functions much faster than that of a chip programmed by software.



Figure 7: Switch

Following are the Steps of Switch Functioning

1. **Learning**

   When switch starts, the MAC address table has no entry. When a node transmits data on its wire the MAC address of the node is learned by Switch Port connected to that node. In this way all the MAC addresses are learned by respective ports and these entries remain in the cache for a specific time. If during this specific time no new frame arrives from a node MAC address entry for that node is dropped from cache.

2. **Forwarding & Filtering**

   When a MAC address for a port is learnt, packets addressed to that MAC address are forwarded only to the port associated with it, using one of the Switching Methods.

3. **Loop Avoidance**

   Switches and Bridges use Spanning Tree Protocol (STP), specified by IEEE 802.1d, to prevent loops.

Switching Methods

- **Store & Forward:** in this method the switch receives complete frame. CRC (Cyclic Redundancy Check), source address and destination address are checked.

- **Cut Through:** In this method forwarding starts as soon as destination address of the frame is received in header. Also known as WIRE SPEED.
- **Fragment Free (Modified Cut Through):** In this method forwarding starts as soon as first 64 bytes of the frame are received as fragmentation occurs usually in first 64 bytes.

Advantages and Disadvantages of Switch

Following are some advantages and disadvantages of switches:

Advantages:

- Reduces the number of Broadcast domains
- Supports VLAN's (virtual local area network (VLAN) is a logical grouping of hosts on one or more LANs that allows communication to occur between hosts as if they were on the same physical LAN.) that can help in Logical segmentation of ports [physical ports]. Splitting up the broadcast domain.
- Intelligent device [compared to Hub's] which can make use of CAM table for Port to MAC mapping
- Compared to Bridges, Switches are more H/w oriented therefore operations are less CPU intense [Basic operations]
- The cost to number of ports ratio is best i.e. for a cheaper cost you get switches with more number of ports available than Routers.

Disadvantages:

- Not as good as a router in limiting Broadcasts
- Communication between VLAN's need inter VLAN routing [Router], but these days there are a number of Multilayer switches available in the market.
- Handling Multicast packets needs quite a bit of configuration and proper designing.

Layer 2 Switch

Layer 2 switching uses the media access control address (MAC address) from the host's network interface cards (NICs) to decide where to forward frames. Layer 2 switching is hardware based, which means switches use application-specific integrated circuit (ASICs) to build and maintain filter tables (also known as MAC address tables). One way to think of a layer 2 switch is as a multi-port bridge.

- Layer 2 switching provides the following:Hardware-based bridging (MAC)
- Wire speed
- High speed
- Low latency
- Low cost

Layer 2 switching is highly efficient because there is no modification to the data packet, only to the frame encapsulation of the packet, and only when the data packet is passing through dissimilar media (such as from Ethernet to FDDI). Layer 2 switching is used for workgroup connectivity and network segmentation (breaking up collision domains). This allows a flatter network design with more network segments than traditional 10BaseT shared networks. Layer 2 switching has helped develop new components in the network infrastructure.

- Server farms — Servers are no longer distributed to physical locations because virtual LANs can be created to create broadcast domains in a switched internetwork. This means that all servers can be placed in a central location, yet a certain server can still be part of a workgroup in a remote branch.

- Intranets — Allows organization-wide client/server communications based on a Web technology.

These new technologies allow more data to flow off from local subnets and onto a routed network, where a router's performance can become the bottleneck.

Limitations

Layer 2 switches have the same limitations as bridge networks.
Bridged networks break up collision domains, but the network remains one large broadcast domain. Similarly, layer 2 switches (bridges) cannot break up broadcast domains, which can cause performance issues and limits the size of your network. Broadcast and multicasts, along with the slow convergence of spanning tree, can cause major problems as the network grows. Because of these problems, layer 2 switches cannot completely replace routers in the internetwork.

Layer 3 Switch

A Layer 3 switch is a high-performance device for network routing. Layer 3 switches actually differ very little from routers. A Layer 3 switch can support the same routing protocols as network routers do. Both inspect incoming packets and make dynamic routing decisions based on the source and destination addresses inside. Both types of boxes share a similar appearance.

Layer 3 switches were conceived as a technology to improve on the performance of routers used in large local area networks (LANs) like corporate intranets. The key difference between Layer 3 switches and routers lies in the hardware technology used to build the unit. The hardware inside a Layer 3 switch merges that of traditional switches and routers, replacing some of a router's software logic with hardware to offer better performance in some situations.

Layer 3 switches often cost less than traditional routers. Designed for use within local networks, a Layer 3 switch will typically not possess the WAN ports and wide area network features a traditional router will always have.
Layer 3 switches can be placed anywhere in the network because they handle high-performance LAN traffic and can cost-effectively replace routers. Layer 3 switching is all hardware-based packet forwarding, and all packet forwarding is handled by hardware ASICs.

Functions of Layer 3 switch

1. Determine paths based on logical addressing

2. Run layer 3 checksums (on header only)

3. Use Time to Live (TTL)

4. Process and respond to any option information

5. Update Simple Network Management Protocol (SNMP) managers with Management Information Base (MIB) information

6. Provide Security

The benefits of layer 3 switching include the following

- Hardware-based packet forwarding
- High-performance packet switching
- High-speed scalability
- Low latency
- Lower per-port cost
- Flow accounting
- Security
- Quality of service (QoS)

ATM Switch

ATM Switches as shown in Figure 8 provide high-speed transfer between both LANs and WANs. Asynchronous Transfer Mode (ATM) is a network technology adopted by the telecommunication sector. It is a high-performance, cell-oriented switching and multiplexing technology that utilises fixed-length packets to carry different types of traffic. The data transfer takes place in the form of cells or packets of a fixed size (53 bytes).

The cell used with ATM is relatively small compared to units used with older technologies. The small constant cell size allows ATM equipment to transmit video, audio, and computer data over the same network, and assures that no single type of data hogs the line.

ATM technology is used for both local and wide area networks (LANs and WANs) that support real-time voice and video as well as data. ATM is widely used as a backbone technology in carrier networks and large enterprises, but never became popular as a local network (LAN) topology. ATM is highly scalable and supports transmission speeds of 1.5, 25, 100, 155, 622, 2488 and 9953 Mbps.

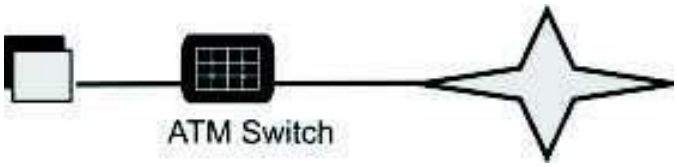


Figure 8: ATM Switch in the middle

Router

Router is a networking device which forwards data packets along networks by using headers and forwarding/routing tables to determine the best path to forward the packets. Common types of modern routers are shown here in Figure 9. Routers work at the Internet layer of the TCP/IP model or layer 3 of the OSI model. Routers also provide interconnectivity between like and unlike media. A router is connected to at least two networks, commonly two LANs or WANs or a LAN and its ISP's network. Some DSL and cable modems, for home use, have been integrated with routers to allow multiple home computers to access the Internet.



Figure 9: Two Modern Routers

Introducing Routing

Once we create an internetwork by connecting your WANs and LANs to a router, we shall need to configure logical network addresses, such as IP addresses, to all hosts on the internetwork so that they can communicate across that internetwork.

The term ***routing*** is used for taking a packet from one device and sending it through the network to another device on a different network. Routers don't really care about hosts—they only care about networks and the best path to each network. The logical network address of the destination host is used to get packets to a network through a routed network, and then the hardware address of the host is used to deliver the packet from a router to the correct destination host.

If your network has no routers, then it should be apparent that you are not routing. Routers route traffic to all the networks in your internetwork. To be able to route packets, a router must know, at a minimum, the following:

- Destination address
- Neighbor routers from which it can learn about remote networks
- Possible routes to all remote networks
- The best route to each remote network
- How to maintain and verify routing information
- The router learns about remote networks from neighbor routers or from an administrator

As it is already discussed that IP routing is basically of three types: static routing, default routing and dynamic routing.

Static Routing

Static routing is the process in which the system network administrator would manually configure network routers with all the information necessary for successful packet forwarding. The administrator constructs the routing table in every router by putting in the entries for every network that could be a destination.

Default Route

A default route is often called the 'route of last resort'. It is the last route tried by a router when all other routes fail because it has the fewest number of network bits matching and is therefore less specific. We use *default routing* to send packets with a remote destination network not in the routing table to the next-hop router. You should only use default routing on stub networks—those with only one exit path out of the network. To configure a default route, you use wildcards in the network address and mask locations of a static route. In fact, you can just think of a default route as a static route that uses wildcards instead of network and mask information.
The syntax for Default routing is : *ip route 0.0.0.0 0.0.0.0 <next hop or exit interface*

Dynamic Routing

Dynamic routing is when protocols (Routing Protocols) are used to find networks and update routing tables on routers. This is easier than using static or default routing, but it'll cost in terms of router CPU processes and bandwidth on the network links. The chief advantages of dynamic routing over static routing are scalability and adaptability. A dynamically routed network can grow more quickly and larger, and is able to adapt to changes in the network topology brought about by this growth or by the failure of one or more network components. Chief among the disadvantages is an increase in complexity.

4.2.7 Gateway

In a communications network, gateway is a network node equipped for interfacing with another network that uses different protocols.

A gateway may contain devices such as protocol translators, impedance matching devices, rate converters, fault isolators, or signal translators as necessary to provide system interoperability. It also requires the establishment of mutually acceptable administrative procedures between both networks. A protocol translation/mapping gateway interconnects networks with different network protocol technologies by performing the required protocol conversions.

A gateway is a network point that acts as an entrance to another network. On the Internet, gateway is a node or stopping point node or a host (end-point) node. Both the computers of Internet users and the computers that serve pages to users are host nodes, while the nodes that connect the networks in between are gateways. For example, the computers that control traffic between company networks or the computers used by internet service providers (ISPs) to connect users to the internet are gateway nodes.

In the network for an enterprise, a computer server acting as a gateway node is often simultaneously acting as a proxy server and a firewall server. A gateway is often associated with both a router, which knows where to direct a given packet of data that arrives at the gateway, and a switch, which furnishes the actual path in and out of the gateway for a given packet.

On an IP network, clients should automatically send IP packets with a destination outside a given subnet mask to a network gateway. A subnet mask defines the IP range of a private network. For example, if a private network has a base IP address of 192.168.0.0 and has a subnet mask of 255.255.255.0, then any data going to an IP address outside of 192.168.0.X will be sent to that network's gateway. While forwarding an IP packet to another network, the gateway might or might not perform Network Address Translation.

Most computer operating systems use the terms described above. Microsoft Windows, however, describes this standard networking feature as Internet Connection Sharing, which acts as a gateway, offering a connection between the Internet and an internal network. Such a system might also act as a DHCP server. Dynamic Host Configuration Protocol (DHCP) is a protocol used by networked devices (clients) to obtain various parameters necessary for the clients to operate in an Internet Protocol (IP) network. By using this protocol, system administration workload greatly decreases, and devices can be added to the network with minimal or no manual configurations.

☞ **Check Your Progress 1**

1. Explain the advantages of using switch. Also discuss its disadvantages.

   ……

   ……

   ……

   ……

2. What is network gateway? Explain

   ……

   ……

   ……

   ……

4.3 SUMMARY

In this unit, various internetwork components used in a computer network are explained. Some of the components such as NIC, Modem, Repeater, Hub, Switch and their functions along with merits and limitations are clearly discussed. After completing this unit you can understand the importance of various internetworking devices particularly at the network layer. You have also studied the different switching and routing methods in this unit. The block of this course has presented the details of transport layer and application layer.

4.4 REFERENCES/FURTHER READINGS

1) *Computer Networks*, A. S. *Tanenbaum* 4th Edition, Practice Hall of India, New Delhi. 2003.

2) *Computer Networking*, J.F. Kurose & K.W. Ross, A Top Down Approach Featuring the Internet, Pearson Edition, 2003.

3) *Introduction to Data Communication & Networking,* Behrouz Forouzan, Tata McGraw Hill, 1999.

4) *Communications Networks*, Leon Garcia, and Widjaja, Tata McGraw Hill, 2000.

5) *Data and Computer Communications,* Willian Stallings, 6th Edition, Pearson Education, New Delhi.

6) www.wikipedia.org

4.5 SOLUTIONS/ANSWERS

☞ Check Your Progress 1

1. In telecommunication, the term **repeater** has the following standardized meanings:

   - An analog device that amplifies an input signal regardless of its nature (analog or digital).
   - A digital device that amplifies, reshapes, retimes, or performs a combination of any of these functions on a digital input signal for retransmission.

2. Following are some advantages and disadvantages of Bridges:

   Advantages:

   - Reliability
   - Manageability
   - Scalability

   Disadvantages:

   - A bridge cannot filter out broadcast traffic.
   - It introduces 20 to 30 % latency.
   - Only 2 networks can be linked with a bridge

☞ Check Your Progress 2

1. Following are some advantages and disadvantages of switches:

   Advantages:

   - Reduces the number of Broadcast domains
   - Supports VLAN's (virtual local area network (VLAN) is a logical grouping of hosts on one or more LANs that allows communication to occur between hosts as if they were on the same physical LAN.) that can help in Logical segmentation of ports [physical ports]. Splitting up the broadcast domain.
   - Intelligent device [compared to Hub's] which can make use of CAM table for Port to MAC mapping
   - Compared to Bridges, Switches are more H/w oriented therefore operations are less CPU intense [Basic operations]
   - The cost to number of ports ratio is best i.e. for a cheaper cost you get switches with more number of ports available than Routers.

   Disadvantages:

   - Not as good as a router in limiting Broadcasts
   - Communication between VLAN's need inter VLAN routing [Router], but these days there are a number of Multilayer switches available in the market.
   - Handling Multicast packets needs quite a bit of configuration & proper designing.

2. In a communications network, gateway is a network node equipped for interfacing with another network that uses different protocols.

A gateway may contain devices such as protocol translators, impedance matching devices, rate converters, fault isolators, or signal translators as necessary to provide system interoperability. It also requires the establishment of mutually acceptable administrative procedures between both networks. A protocol translation/mapping gateway interconnects networks with different network protocol technologies by performing the required protocol conversions.

UNIT 2 TRANSPORT LAYER

Structure **Page No.**

2.0 Introduction 34
2.1 Objective 34
2.2 Addressing 35
2.3 Reliable delivery 35
2.4 Flow control 38
2.5 Connection Management 38
2.6 Multiplexing 40
2.7 Congestion Control 40
2.8 Quality of Cervices (QoS) 42
2.9 TCP window Management 43
2.10 Ports 44
2.11 Summary 46
2.12 References/Further Reading 46
2.13 Solution /Answers 47

2.0 INTRODUCTION

The transport layer supports two protocols in TCP/IP protocol suite. One is Transmission Control Protocol (TCP). TCP is connection oriented that provides reliable end-to-end transmission. Another protocol is User Datagram Protocol (UDP). UDP is simple and provides well sequenced transport function when reliability and serving are less important than size and speed. Transport layer services are implemented by transport protocols used between two transport entities. Transport layer services are similar to the data link services. Data link layer is designed to provide its services within a single network, while the transport layer provides services across an inter network made up of many networks. There are seven categories of services provided by the transport layer. These services are End to end delivery, Addressing, Reliable delivery, Flow control, Connection management, Multiplexing and Congestion Control.

2.1 OBJECTIVE

After going through this unit, you should be able to:

- Know the Functions and Services of transport layer
- Understand the Working of transport layer
- Understand the TCP Window management
- Know the different transport layer design issues

End to End Delivery

As shown in figure 1, network layer answers the end to end delivery of individual packets from a machine to another machine in different network, but does not see any relationship between those packets. It treats each as an independent entity. Further, the packet needs to be delivered to the last participating entity, i.e. a process engaged in data exchanged. But the transport layer makes sure that the entire message (not a single packets receives) is delivered to a process that is the end (last) entity participating in message exchange. So it provides process-to-process or end-to-end delivery of an entire message.

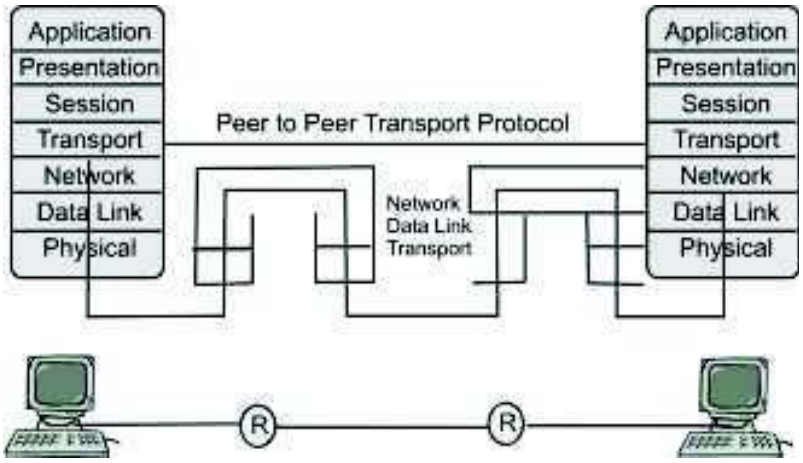


Figure 1: End to End delivery of packets

2.2 ADDRESSING

Transport layer interact with the functions of session layer. Many protocols combine session, presentation and application level protocols into a single package called an application. In these cases delivery to the session layer functions is, in effect delivery to the application. So communication occurs not just from end machine to end machine but from end application to end application. Data generated by an application on one machine must be received not just by other machines but by the correct application on that machine.

In most cases, we end up with the communication between many to many entities, called service access points as shown in figure 2 given below. To ensure accurate delivery from service access point to service access point we used another level of addressing in addition to the network and data link level.

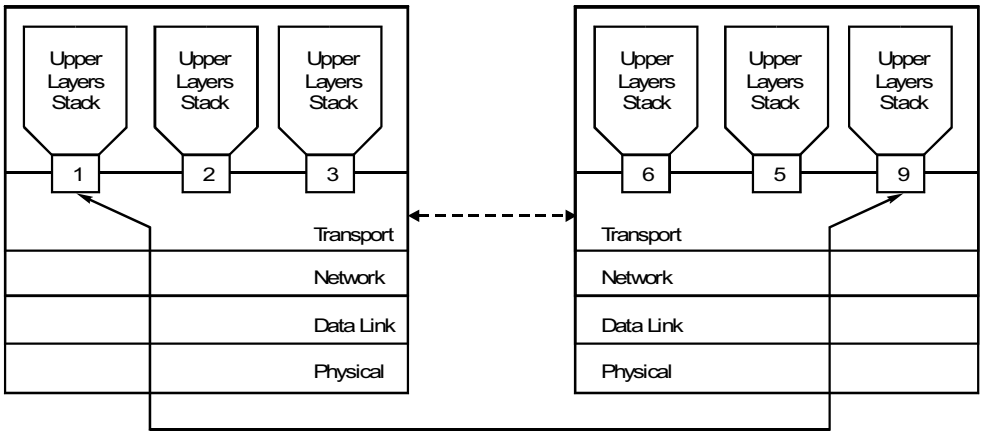


Figure 2: Service Access Points

2.3 RELIABLE DELIVERY

It is responsible for reliable delivery of data by providing the following methods also shown in figure 3:

- Error control
- Sequence control
- Loss control

- Duplication control

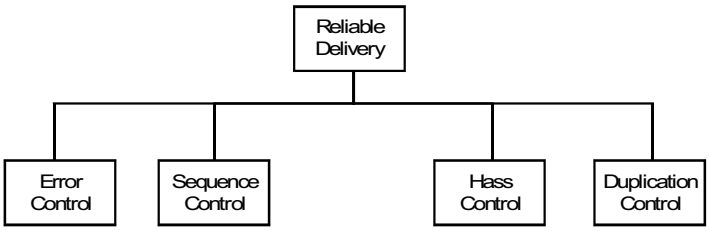


Figure 3: Methods of reliable delivery of packets

a) **Error Control:** When transferring data the primary goal of reliability if occur control. Data must be delivered to their destination. Exactly as they originated from the source. The reality of physical data transport are that while 100 per cent error free delivery is probably impossible, transport layer protocols are designed to come as close as possible.

Mechanisms full errors handling at this layer are based on error detection and retransmission. With the error handling, performed using algorithms implemented in software such as checksum "error detection and correction".

b) **Sequence Control:** Second aspect of reliability implemented at the transport layer is sequence control. On the sending end, the transport layer is responsible for ensuring that data with received from the upper layers are usable by the lower layers. ON the receiving end it is responsible for ensuring that the various pieces of a transmission are correctly reassembled.

Segmentation: When the size of the data units received from the upper layer is too long for the network layer datagram and data link layer frame to handle, the transport layer divides it into smaller usable blocks. This dividing process is called segmentation.

Concatenation: When the sizes of the data units belonging to a single session are so small that several can fit together into a single data queue are frame, the transport protocol combines them into a single data unit. This combining process is called concatenation.

Sequence Number: Most transport layer services add sequence number at the end of each segment.

If a longer data unit has been segmented the sequence number indicate the reassembly.

If several shorter units have been concatenated the numbers indicate the end of each subunit and allow them to be separated accurately at the destination.

c) **Loss Control:** The third aspect of reliability covered by the transport layer is loss control as depicted in figure 4. The transport layer ensures all pieces of the transmission arrive at the destination, not just some of them. When data have been segmented for delivery, some segments may be lost in transmit. Sequence number allows the receiver's transport layer protocol to identify any missing segment and request in delivery.

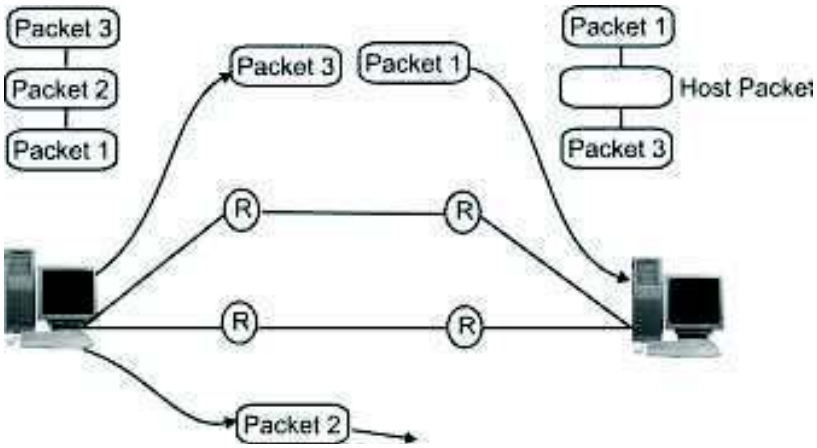


Figure 4: Loss Control

d) **Duplication Controls:** The fourth aspect of reliability by the transport layer is duplication control as shown in figure 5. Transport layer functions must guarantee that no places of data arrive at the receiving system duplicated. As they allow identification of last packets, sequence no. allows the receiver to identify and discard duplicate segments.

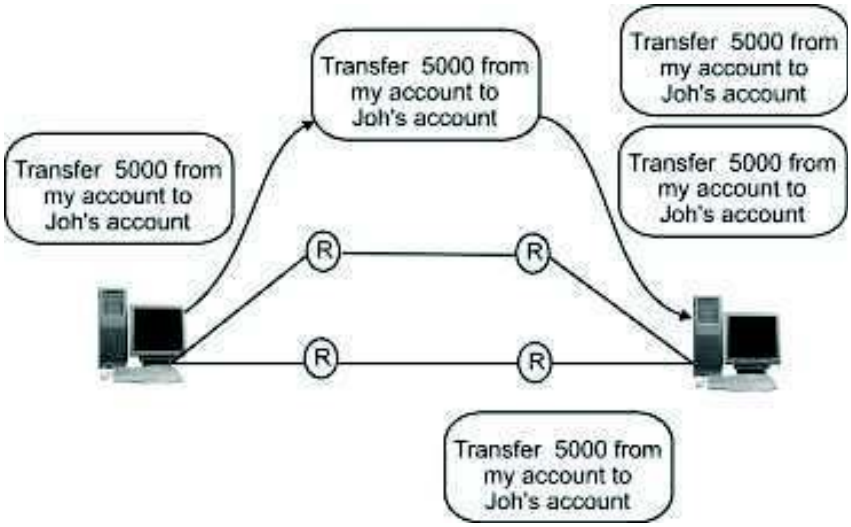


Figure 5: Duplication Control

☞ Check Your Progress 1

1. Which layer ensures the process-to-process or end-to-end delivery of an entire message? Explain.

   ……

   ……

   ……

2. List the methods or mechanism provided by Transport layer for reliable delivery of data.

   ……

   ……

2.4 FLOW CONTROL

Like the data link layer transport layer is responsible for flow control. Flow control is performed end to end rather than across a single link. Transport layer flow control uses a sliding window protocol. The window at the transport layer can vary in size to accommodate buffer occupancy as depicted in figure 6 given below.

Sliding window is used to make data transmission more efficient as well as to control the flow of data so that the receiver does not become overwhelmed. Sliding window used at the transport layer are usually byte oriented rather than frame oriented.

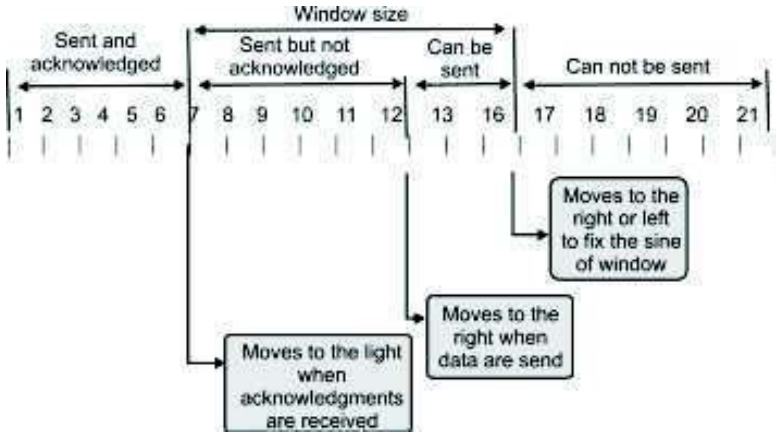


Figure 6: Sliding Window for Flow Control

Some points about sliding window at the transport layer are as follows:

1. Sender does not have to send a full window's worth of data.
2. An acknowledgement can expand the size of the window based on the sequence number of the acknowledged data segment.
3. The size of the window can be increased as decreased by the receiver.
4. The receiver can send acknowledgement at anytime.

2.5 CONNECTION MANAGEMENT

End to end delivery can be accomplished in two ways connection oriented and connectionless. The connection oriented mode is most commonly used from both two modes. A connection oriented protocol establishes a virtual circuited or pathway through the internal between sender and receiver. All of the packets belonging to a message are then sent over this same path. Using a single pathway for the entire message facilitates the acknowledgement process and retransmission of damaged and lost frames connection oriented services is generally considered reliable.

Connection Oriented transmission has three stages:

1. Connection establishment.
2. Data transfer
3. Connection termination.

Connection Establishment: Before communicating device can send data to the other, the initializing device must first determine the availability of the other to exchange data and a pathway must be found through the network by which the data can be sent.

This step is called connection establishment. Connection establishment requires three actions called three way handshake as shown in figure 7 given below.

- The computer requesting the connection sends a connection request packet to the intended receiver.
- The responding computer returns a confirmation packet to the requesting computer.
- The requesting computing returns a packet acknowledging the confirmation.

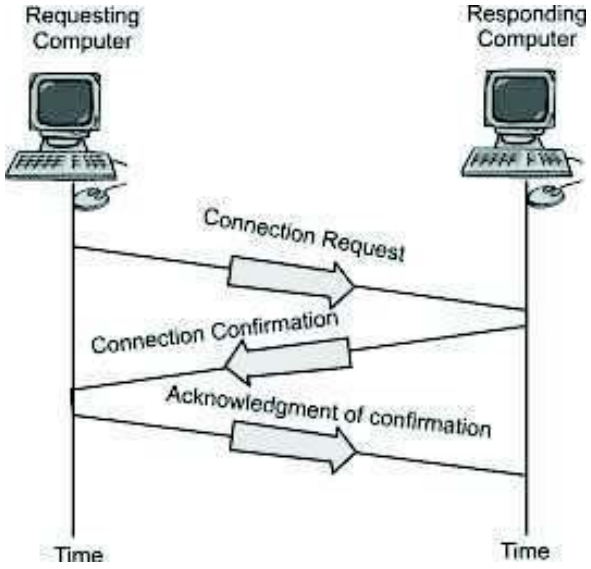


Figure 7: Connection Establishment

Connection Termination: Once all of the data have been transferred, the connection must be terminated.

Connection termination also requires three way handshake as shown in figure 8:

- Requesting computer sends a disconnection request packet.
- Responding computers confirms the disconnection request.
- The requesting computer acknowledges the confirmation.

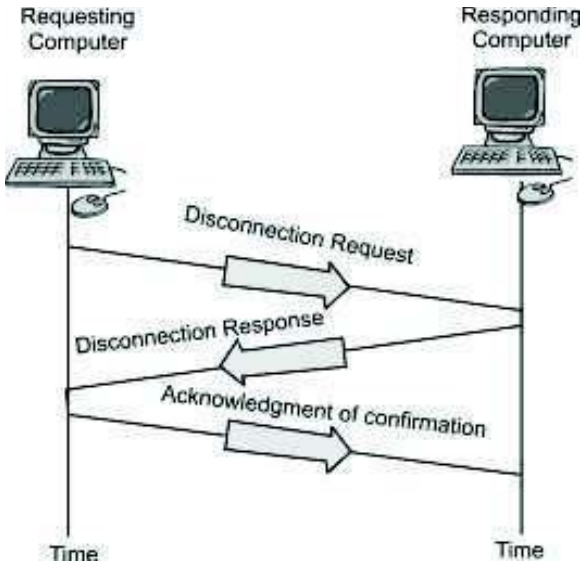


Figure 8: Connection Termination

2.6 MULTIPLEXING

To improve transmission efficiency, the transport layer has the option of multiplexing. Multiplexing at this layer occurs in two ways as shown in figure 9:

1. **Upward** -meaning that multiple transport layer connections use the same network connection

2. **Downward**- (meaning that one transport layer connection uses multiple network connections.

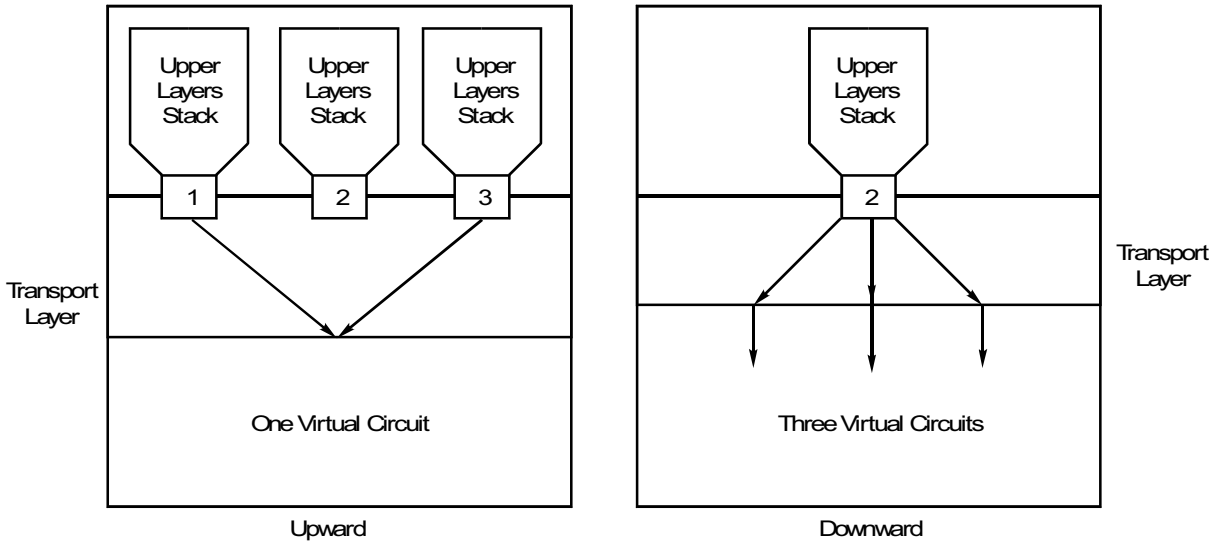


Figure 9: Upward and downward Multiplexing

Upward Multiplexing: Transport layer uses virtual circuits based on the services of the lower three layers. The underlying network charge for each virtual circuit connection. To make well cost-effective use of an established circuit, the transport layer sends several transmissions based for the same destination along the same path by upward multiplexing.

Downward Multiplexing: It allows the transport layer to split a single connection among several different paths to improve throughput (speed of delivery) as shown in figure 9. This option is useful when the underlying networks have low or slow capacity. For example, some network layer protocols have restriction on the sequence number that can be handled .X.25 uses a three bit numbering code, so sequence number are restricted to the range of 0 to 7. In this case throughput can be unacceptably low. To counteract this problem the transport layer can use more than one virtual circuit at the network layer to improve throughput by sending several data segment at once delivery is faster.

2.7 CONGESTION CONTROL

Congestion control refers to techniques and mechanisms that can either prevent congestion, before it happens, or remove congestion, after it has happened. Congestion control mechanism is divided into two broad categories:

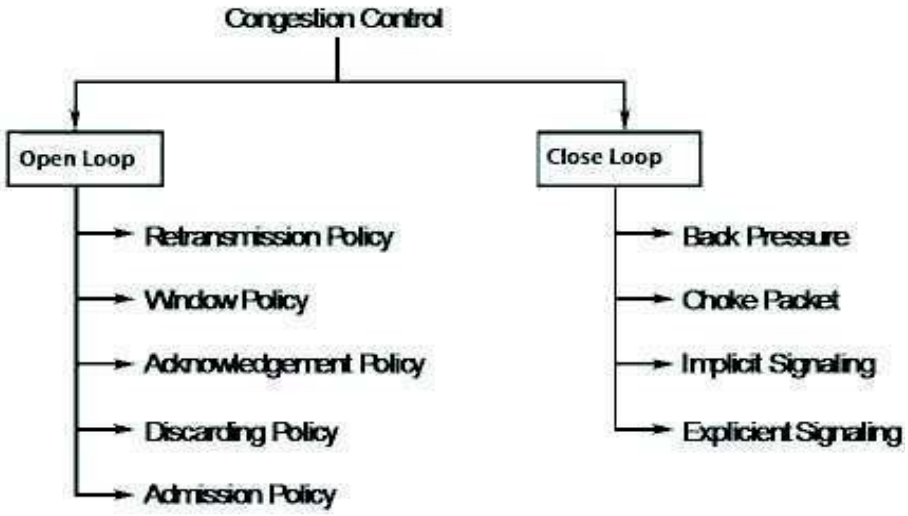


Figure 10:

Open Loop: Polices are applied to prevent congestion before it happens.

Closed Loop: Mechanism try to alleviate congestions after it happens.

Congestion Control Policy

TCP's general policy for handling congestion is based on 3 phases:

- Slow start,
- Congestion avoidance, and
- Congestion detection

The sender starts with a very slow rate of transmission but increases the rate rapidly to reach a threshold. When threshold is reached, the data rate is reduced to avoid congestion.

Congestion Avoidance: Additive Increase

In the congestion avoidance algorithm, the size of the congestion window increases additively until congestion is detected.

Congestion Detection: Multiplicative Decrease

An implementation reacts to congestion detection in one of the following ways:

1. If detection is by time-out, a new slow-start phase starts.
2. If detection is by three (acknowledgements) Acks, a new congestion Avoidance phase starts.

Slow Start: Exponential Increase

At the beginning of the connection, set the congestion window size to the maximum segment size. For each segment that is acknowledged, increase the size of the congestion window by one maximum size until you reach a threshold of half the allowable window size. This is some timer called slow start which is totally misleading because the process is not slow at all.

The size of the congestion window increase exponentially. The sender sends one segment receives one acknowledgement, increases the size to two segments, sends two segments, receives ack for two segments; increase the size to four and so on.

In other words, after receiving the 3rd ACK, the size of the window has been increased to eight segments (i.e. $2^3 = 8$). To avoid congestion before it happen one must slow down this exponential growth. After the size reaches the threshold, the size is increased one segment for each acknowledgement even if as ACK is for several segment.

Multiplicative Decrease

If the congestion occurs the congestion windows size must be decreased. The only way the sender causes that connection has occur through a lost segment. If the sender does not receive an acknowledgement for a segment before it transmission timer matured, it assumes that there is congestion.

The strategy says if a time out occurs, the threshold must be set to half of the congestion window size and the congestion or size should start from one again.

2.8 QUALITY OF SERVICES (QOS)

A stream of packet, from a source to a destination is called a flow. In connection oriented network, all the packets belonging to a flow the same route. But in connection less network they may follow different routes. The need of each flow can be characterized by four primary parameters:

1. Reliability: Reliability is the ability of a system to perform and maintain its functions in normal conditions as well as under unexpected conditions.

2. Delay is defined as the time interval elapsed between the departures of data from the source to its arrival at the destination.

3. Jitter**:** Jitter refers to the variation in time between packets arriving at the destination.

4. Bandwidth refers to the data rate supported by a network connection or interface.

Techniques to Improve QOS

1. **Over Provisioning:** An easy solution is to provide so much router capacity, buffer space and bandwidth that the packets just fly through costly.

2. **Buffering:** Flows can be buffered on the receiving side before being delivered. Buffering those does not affected the reliability of bandwidth and increases the delay but it smoothes out the jitter.

3. **Scheduling:** Packets from different flows arrive at a switch or router for processing. A good scheduling technique treats the different flows in a fair and appropriate manner.

   Here we discuss some scheduling techniques to improve the quality of service such as:

   i) FIFO Queuing

   ii) Priority Queuing

   iii) Weighted Fair Queuing

FIFO Queuing

In FIFO Queuing packets wait in a buffer (Queue) until the node is ready to process them. If the average rate is higher than the average processing rate, the queue will fill up and new packets will be discarded.

Priority Queuing

In this packets are first assigned to **priority class**. Each priority class has its own Queue. The packets in the highest priority Queue are processed first. But if there is a continuous flow in high-priority Queue, the packets in the low priority Queues will never have a chance to be processes.

Weighted Fair Queuing

In this method, the packets are still assigned to different classes and admitted to different queues. The queues, however, are weighted based on the priority of the queues; higher priority means a higher weight.

The system processes packets in each queue in a round-robin fashion with the number of packets selected from each queue based on the corresponding weight.

For example: If the weights are 3, 2 and 1, three packets are processed from the first queue, two from the second queue and one from the third queue.

If the system does not impose priority on the classes, all weights can be equal. In thus way, we have fair queuing with priority.

4. Traffic Shaping
   a) Leaky bucket
   b) Token bucket

2.9 TCP WINDOW MANAGEMENT

TCP uses two buffers and one window, to control the flow of data coming from the sending application program. The application program creates data and writes it to the buffer. The sender imposes a window on this buffer and sends the segments as long as the size of the window is not zero. The TCP receiver has buffer also. It receives data, checks them, and stores them in buffer to be consumed by the receiving application program.

A sliding window is used to make transmission more efficient as well as to control the flow of data so that the destination does not become over whelmed with data. TCP's sliding window is byte oriented

TCP's Sliding Window

The source does not have to send a full window's worth of data. The size of the window can be increased or decreased by the destination. The destination can send an acknowledgement at any time.

Silly Window Syndrome

Silly Window Syndrome is another problem that can degrade the TCP performance. This problem occurs when the sender transmits data in large blocks, but an interactive application on the receiver side reads data 1 byte at a time as shown in figure 11.

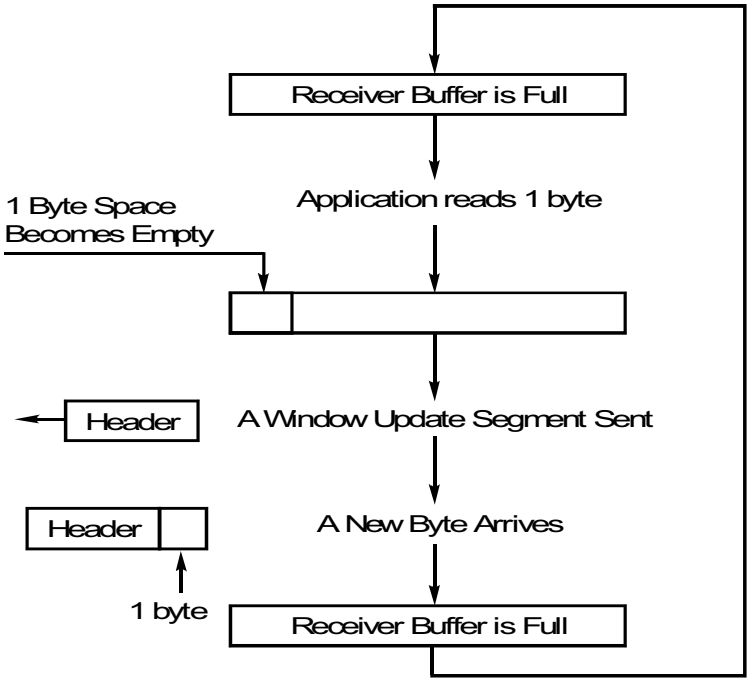


Figure 11: Silly Window Syndrome

1. Initially the receiver's buffer is full so it sends a window size 0 to block the sender.

2. But the interactive application reads one byte from the buffer, so one byte space becomes empty.

3. The receiving TCP sends a window update to the sender informing that it can send 1 byte.

4. The sender sends 1-new byte.

5. The buffer is full again and the window size is 0. The behavior can continue forever's.

Solution to Silly Window Syndrome

A solution to silly window syndrome was suggested as follows:
It was suggested that the **receiver should not send** a window update for 1 byte. Instead it must wait until it has a substantial amount of buffer space available and then sends the window update. To be specific, the receiver should not send a window update until it can handle the maximum window size, it has advertised at the time of establishing a connection or its buffer is half empty, whichever is smaller.

The sender can also help to improve the situation. It should not send tiny segments instead it must wait and send a full segment or at least one containing half of the receivers buffers size.

2.10 PORTS

In computer networking of connection-based communication port is like a medium through which, an application establish a connection with another application by binding a socket by a port number. Addressing the information and the port number, accompanied the data transfer over the network.

The Ports are used by TCP and UDP to deliver the data to the right application, are identified by a 16-bit number present in the header of a data packet. Ports are typically used to map data to a particular process running on a client. If we consider a letter (data packet) sent to a particular apartment (IP) with house no. (Port no), at this time the port no. is the most important part for the delivery of the letter. In order for the delivery to work, the sender needs to include a house number along with the address to ensure the letter gets to the right destination.

Do you know?

TCP and UDP ports are 16 bit number

The TCP and UDP protocols use ports to map incoming data to a particular process running on a computer.

- Port is represented by a positive (16-bit) integer value
- Some ports have been reserved to support common/well known services:
  - FTP 21/TCP
  - TELNET 23/TCP
  - SMTP 25/TCP
  - LOGIN 513/TCP
- User level process/services generally use port number value >= 1024

Types of Port

1. Well known port (0 to 1023)-They are controlled by IANA
2. Registered Port (1024-49159)
3. Dynamic port (49152-65535)

If we consider the client-server architecture, a server application binds a socket to a specific port number in connection-based communication. It registered the server with the system where all the data destined for that port.

Do you know?

Port number permits unique identification of several simultaneous processes using TCP/UDP

Now we are aware of the importance of the port number. In the same order there are some ports which are predefine and called reserved ports. Some of them are given in Table 1given below:

Table 1: Reserved Port Numbers.

Service	Port no.
ECHO	7
DAYTIME	13
FTP	21
TELNET	23
SMTP	25
FINGER	79
HTTP	80
POP3	110

Do you know?

If we consider the range of the port numbers, there are 0 to 65,535 ports available.

Tips

The port numbers ranging from 0 - 1023 are reserved ports or we can say that are restricted ports. All the 0 to 1023 ports are reserved for use by well-known services such as FTP, telnet and http and other system services. These ports are called well-known ports.

☞ Check Your Progress 2

1. List three stages of Connection Oriented transmission.

   ……

   ……

   ……

   ……

2. Compare and contrast between Upward and Downward Multiplexing.

   ……

   ……

   ……

   ……

2.11 SUMMARY

Transport layer is mainly responsible for end to end reliable delivery, segmentation and concatenation. Two main protocols that operate on transport layer are TCP and UDP.TCP provides reliable connection oriented service while UDP provides unreliable connectionless service. The data link and transport layer perform many of the same duties .The data link layer function in a single network, while the transport layer operates across an internet. Flow control at the transport layer is handles by three walled sliding window. Multiplexing can be downward of upward in transport layer. Connection establishment and termination can be done by using three way handshakes. Transport layer works on port address. To know more about the Transport layer and its protocols (TCP and UDP), student may please refer to the course material of BCS-61TCP/IP Programming or BCS-54 Network Programming.

2.12 REFERENCES/FURTHER READING

1. Computer Networks, A. S. Tanenbaum 4th Edition, Practice Hall of India, New Delhi. 2003.

2. Introduction to Data Communication & Networking, 3rd Edition, Behrouz Forouzan, Tata McGraw Hill.

3. Douglas E. Comer, *Internetworking with TCP/IP Vol.1: Principles, Protocols, and Architecture* (4th Edition).

4. James F. Kurose, *Computer Networking*: A Top-Down Approach Featuring the Internet (3rd Edition).

5. Larry L. Peterson, *Computer Networks*: A Systems Approach, 3rd Edition (The Morgan Kaufmann Series in Networking).

6. W. Richard Stevens, *The Protocols (TCP/IP Illustrated, Volume 1)*.

7. www.wikipedia.org

8. William Stallings, *Data and Computer Communications*, Seventh Edition.

2.13 SOLUTION /ANSWERS

☞ Check Your Progress 1

1. Transport layer makes sure that the entire message (not a single packets receives) is delivered to a process that is the end (last) entity participating in message exchange. So it provides process-to-process or end-to-end delivery of an entire message.

2. Transport Layer is responsible for reliable delivery of data, it provide following methods to provide reliable delivery of data.
   - Error control
   - Sequence control
   - Loss control
   - Duplication control

☞ Check Your Progress 2

1. Connection Oriented transmission has following three stages:
   i) Connection establishment.
   ii) Data transfer
   iii) Connection termination.

2. **Upward Multiplexing:** Transport layer uses virtual circuits based on the services of the lower three layers. The underlying network charge for each virtual circuit connection. To make well cost-effective use of an established circuit, the transport layer sends several transmissions based for the same destination along the same path by upward multiplexing. **Downward Multiplexing:** Allows the transport layer to split a single connection among several different path to improve throughput (speed of delivery). This option is useful when the underlying networks have low or slow capacity. For example, some network layer protocols have restriction on the sequence number that can be handled .X.25 uses a three bit numbering code, so sequence number are restricted to the range of 0 to 7. In this case throughput can be unacceptably low. To counteract this problem, the transport layer can use more than one virtual circuit at the network layer to improve throughput by sending several data segment at once delivery is faster.

UNIT 1 NETWORK LAYER

Structure **Page No.**

1.0 Introduction 5
1.1 Objectives 5
1.2 Switching 6
1.3 Routing Algorithm 7
1.3.1 Classification of Routing Algorithms
1.3.2 Non-Adaptive Routing Algorithm (Static Routing)
1.3.3 Dynamic Routing Algorithm (Adaptive)
1.3.4 Comparison Link State Versus Distance Vector Routing
1.4 Congestion Control 15
1.4.1 Algorithm For Congestion Control
1.5 Network Addressing 20
1.5.1 Classful Addressing
1.5.2 NetID and HostID
1.6 Fragmentation 26
1.7 Error Messaging Services 28
1.7.1 ICMP (Internet Control Message Protocol)
1.7.2 IGMP(Internet Group Message Protocol)
1.8 Summary 32
1.9 Further Reading 32
1.10 Solution/Answers 33

1.0 INTRODUCTION

As you know, the network layer is one of the important layers of OSI model. It is responsible for different tasks of networking, but mainly its role is to determine addresses and finding a route between a source and destination node or between two intermediate devices. It establishes and maintains a logical connection between these two nodes, either a connectionless or a connection oriented communication. The basic purpose of the network layer is to provide a network to network communication capability in contrast to machine to machine common provided by data line 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 determined based on static tables that are “wired into” the network and rarely changed. If too many packets are present in the subnet at the same time, they will get in one another’s way forming bottlenecks. The controlling such congestion also belongs to the network layer. The quality of service also depends on network layer issue.

In this unit, we will study the fundamental Issues of network layer. These issues are designing interface between the host and the network, the routing methods, congestion control methods and Internetworking issues. In this unit we will study how routing is done at network layer using adaptive and non adaptive algorithm. We will also discuss the Network addressing. Further, some Error reporting protocols ICMP and IGMP on network layer will be discussed.

1.1 OBJECTIVES

After going through this unit, you should be able to:

- Know the basic issues of network layer
- Understand the different switching methods used at network layer
- Know the routing mechanisms

- Understand the congestion control methods
- Differentiate between adaptive and non adaptive algorithm
- Know process of Error reporting protocols at network layer

1.2 SWITCHING

As you have studied earlier in block 1, unit 3, that Switching is used to determine the path to be used for forwarding the information to the receiver. You also know that the Switching methods are mainly divided into Circuit, Message and Packet switching. In this section, we will explore the other switching mechanism like virtual circuit and datagram.

Virtual circuit is a connection oriented communication service that is delivered by means of packet mode communication. After a connection or virtual circuit is established between two nodes or application processes, a bit stream may be delivered between the nodes. It is similar to the circuit switching only in virtual circuit permanent/physical connection are not established. If router fails all virtual circuits that pass through, the failed router are terminated.

Datagram is opposite of Virtual circuit, it is connection less service. A datagram or packet needs to be self-contained without any dependency on earlier data-transfer because there is no connection of fixed duration between the two communicating nodes as shown in figure 1. Following Table 1 show the difference between circuit switching, virtual circuit and datagram:

Table 1: Difference Between Circuit Switching, Virtual Circuit and Datagram

S.No.	**Circuit Switching**	**Virtual Circuit**	**Datagram**
1.	Dedicated path between sender and receiver	Non-dedicated between sender and receiver.	Non-dedicated.
2.	Connection oriented Highly Reliable	Connection oriented	Connection less
3.	Data transfer in continuous form.	Data transfer in packets	Data transfer in packets.
4.	Bandwidth is fixed.	Not fined dynamic.	Dynamic
5.	E.g. telephone services	Subnet	Internet
6.	Data transfer in voice form.	Data transfer usually in text from.	Text form
7.	Call setup delay is maximum.	Delay Moderate	Call set up negligible.
8.	Transmission delay minimum	Moderate	Maximum transmission delay
9.	Unused bandwidth is just wasted	Moderate	No wasted.

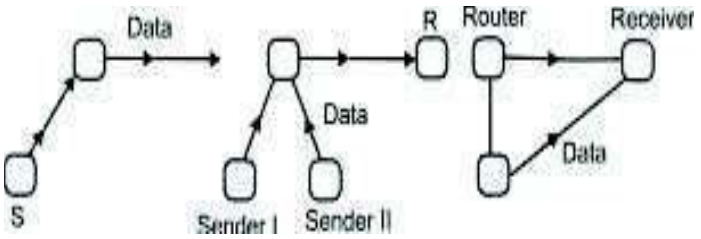


Figure 1: Circuit switching, Virtual circuit and Datagram (left to right)

1.3 ROUTING ALGORITHM

The main function of the network layer is routing packets from the source machine to the destination machine. So, the algorithm that choose the routes and the data structure that they use are a major area of network layer design.

It is that part of the network layer responsible for deciding which output line an incoming packet should be transmitted on.

Desired Properties of a Routing Algorithm

1. **Correction:** The routing should be done properly and correctly so that the packets may reach their proper destination.

2. **Simplicity:** The routing should be done in a simple manner so that the over head is as low as possible.

3. **Robustness:** Once a major network becomes operative, it may be expected to run continuously for years without any failure.

4. **Stability:** The routing algorithm should be stable under all possible circumstances.

5. **Fairness:** Every node connected to the network gets a fair change of transmitting their packets. This is generally done on a first come first serve basis.

6. **Optimality:** The routing algorithms should be optimal in terms of throughput and minimizing mean packet delays. Here there is a trade off and one has to choose depending on his suitability.

1.3.1 Classification of Routing Algorithms

Routing algorithm may be classified as follows:

1. Adaptive Algorithm

2. Non-adaptive algorithms

Adaptive algorithms use such dynamic information as current topology, load delay etc to select routes.

Non adaptive algorithms, routes never changes once initial routes have been selected. Also, called static routing.

Adaptive Routing Algorithm (Dynamic Routing)

It changes their routing decision to reflect changes in the topology and in traffic as well. These get their routing information from adjacent routers or from all routers. Routing decision may be changed when network topology and/or traffic load changes. The optimization parameters are the distance, number of hops and estimated transit time.

Adaptive routing algorithms can be further classified as follows:

1. **Isolated:** Each router makes its routing decisions using only the local information it has on hand. Specifically, routers do not even exchange information with their neighbors.

2. **Centralized:** A centralized node makes all routing decision specifically the centralized node has access to global information.

3. **Distributed:** Algorithm that uses a combination of local and global information.

Isolated: In this method, the node decides the routing without seeking information from other node. The disadvantage is that the packet may be sent through a congested route resulting in a delay.
Some of the examples of this type of algorithm for routing are :

- **Hot Potato Routing:** Form of routing in which the nodes of a network have no buffer to store packets in before they are moved on to their final predetermined destination.
- In normal routing situation, when multiple packets contend for a single outgoing channel, packets that are not buffered are dropped to avoid congestion.
- Backward Learning: In this method the routing tables at each node gets modified by information from the incoming packets. Backward learning routing algorithm used for routing traffic that makes decisions by assume that a can optimally reach B through C.

Centralized Routing

Advantage: Only one node is required to keep the information.

Disadvantage: If the central node goes down the entire network is down, i.e. single point of failure.

Distributed: It receives information from its neighboring nodes and then takes the decision about which way to send the packet.

Disadvantages: If in between the interval it receives information and sends the packet, something changes then packet may be delayed.

Optimality Principle

Optimality principle is a general statement about optimal routes regardless of network topology or traffic.



Figure 2: An example of Optimality Principle

With reference to Figure 2 above, Optimality principle states that if router I is on the optimal path from router 'I' to router 'K' then the optimal path from 'J' to 'K' also falls along the same route.

To prove the above statement we can say, if there was a better way from J to K, then you could use that with the path from I to J for a better path from I to K, so your starting point (the path from I to K was optimal) is contradicted.

1.3.2 Non-Adaptive Routing Algorithm (Static Routing)

These algorithms do not take their routing decisions on measurements and estimates of the current traffic and topology. Instead the route to be taken from one node to the other is computed in advance. This is also known as static routing.

1. **Shortest Path Routing:** According to this algorithm build a graph of the subnet, with each node of graph representing a router and each arc of the graph representing communication line. To choose a route between a pair of routers, just finds the shortest path between them on the graph.

   Two ways of measuring distance

   a) Distance in terms of link delay.

   b) Measuring path length in number of hops.

      In the most general case, the labels on the arc could be computed as a function of the bandwidth, average traffic communication cost, mean queue length measured delay, and other factors.

   c) **Flooding:** According to this algorithm every incoming packet is sent out on every outgoing line except the cone it arrived on.

      Flooding generates vast number of duplicate packets unless some measures are taken to damp the process. One such measure is to have a hop counter contained in the header of each packet, which is decremented at each hop with the packet being discarded when the counter reaches zero. The hop counter should be initialized to the length of path from source to destination. If the sender does not know how long the path is, it can be initialize the counter to the worst case, namely, the full diameter of the subnet. An alternative technique for damming the flood is to keep track of which packets have been flooded, to avoid sending them out a second time.

   d) **Selective Flooding**

      In this algorithm the routers do not send every incoming packet out on every line, only on those line that are going approximately in right direction.

      Though flooding is not practical in most application, but (yes) is does have some uses.

      i) **In military applications:** Tremendous robustness of flooding is highly desirable.

      ii) **In distributed data base application:** It is sometimes necessary to update all the data bases concurrently, in which case flooding can be useful.

Disadvantages:

1. Duplicacy

2. Infinite looping.

3. **Flow-Based Routing:** This algorithm considers two strategies in account to decide the route.

   a) Topology.
   b) Load for routing
   Previous static algorithm only considers topology in account not the load for routing. The basic idea behind the analysis is that for a given line, if capacity and average flow is known, it is possible to compute the mean packet delay on that line from queuing theory. The routing problem then reduces to finding the routing algorithm that produces the minimum average delay for the subnet.

   For this technique, certain information must be known in advance.

   c) Topology

   d) Traffic Matrix Fij

   e) Line capacity matrix Cij

   Now $T = \frac{1}{\mu_c - \lambda}$

   Where T = Mean delay

   $\frac{1}{\mu}$ = Mean packet size

   □ = Mean flow in packet/sec. (No of arrival frame on particular line).

   C = Capacity

   Two common methods are used to calculate the shortest path between two routers.

1. Distance vector routing (Bellman ford routing algorithm and the Ford Fulkerson algorithm).
2. Link state routing (based on Dikastra's algorithm).

1.3.3 Dynamic Routing Algorithm (Adaptive)

Routing algorithms can be classified based on inter-domain and intra-domain as shown in figure 3 given below.

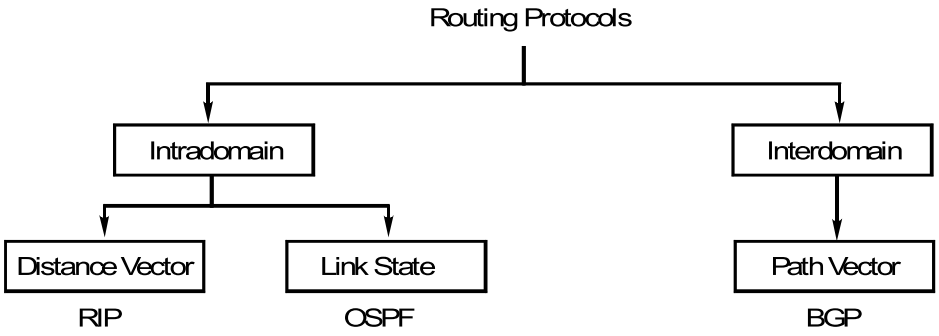


Figure 3: Classification of routing algorithms

RIP □ Routing information protocol

OSPF □ Open shortest path first

BGP □ Border gateway protocol.

1. **Distance vector routing**

Distance Vector Routing: According to this algorithm, each router maintains a table (vector) giving the best known distance to each destination and which line to use to get there. These tables are updated by exchanging information with the neighbors. This algorithm is also called **Bellman-Ford** or the **Ford-Fulkerson Algorithm**.

In distance vector routing, each router maintains a routing table indexed by, and containing one entry for each router in the subnet. This entry contains two parts

a) The preferred outgoing line to use for that destination.

b) An estimate of the time as distance to that destination.

The router is assumed to know the distance to each of its neighbors. For example, consider a subnet as given below in figure 4.

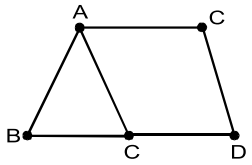


Figure 4: Subnet Diagram

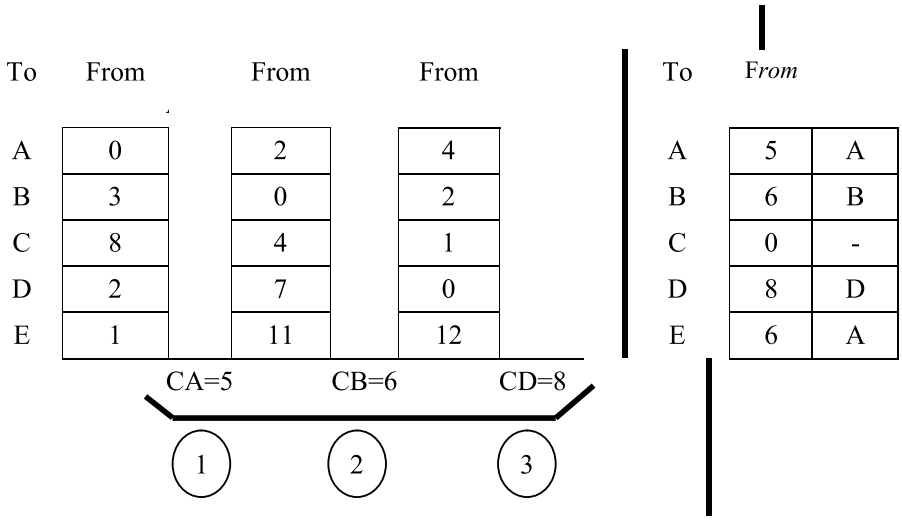


Figure 5: An example of Distance Vector Routing

Part (a) shows a subnet.

Part (b) the first 3 column shows the delay received from neighbor of router 'C' is A and B and D.

For example : as shown in the figure 5, 'A' claims to have 3 msec delay to B 8 msec delay to 'C' and so on. Similarly 'B' claims to have 2 msec delay to A, 4 msec delay to 'C' and so on.

'C' has estimated his delay to neighbour A, B, D as 5, 6, 8 respectively (CA = 5, CB = 6, CD = 8).

Now (4) column shows how router 'C' decides his new route to router 'E'. There are three ways

a) If 'C' follows line 'A' then delay is CE = CA ® AE = 5 + 1 = 6 msec.

c) If ‘C’ follows line ‘B’ then delay is CE = CB ® BE = 6 + 11 = 17 msec.

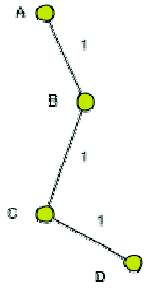
d) If ‘C’ follows line ‘D’ then delay is CD ® DE ® 8 + 12 = 20 msec.

Min delay time is via neighbor route ‘A’ so from C to E line is chosen ‘A’ in column (4)

The same calculations is performed for all destination, with the new routing table as (4)

Problem of Distance Vector Routing

1.**Count to Infinity Problem**-Distance vector routing has a serious drawback in its receptivity. In particular, it reacts rapidly to good news, but slowly to bad news. Following illustration shows an imagined network and denotes the distances from router A to every other router in the figure 6. Until now everything works fine.



B	C	D
1	2	3

Figure 6: Count to Infinity Problem

The illustration shows that link (A, B) is broken. Router B observed it, but in its routing table he sees, that router C has a route to A with 2 hops.

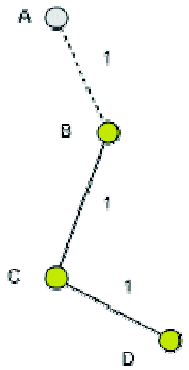
The problem is, that router B doesn't know that C has router B as successor in his routing table on the route to A.

That occurs followed count-to-infinity problem. B actualizes his routing table and takes the route to A over router C.

In the next figure 7; we can see the new distances to A. In C's routing the route to A contains router B as next hop router, so if B has increase his costs to A, C is forced to do so. Router C increases his cost to A about B + 1 = 4.

Now we see the consequence of the distributed Bellman-Ford protocol: Because router B takes the path over C to A, it updates its routing table and so on!
At the end, this problem is going to immobilize the whole network.

1.



B	C	D
3	2	3

B	C	D
3	4	3

B	C	D
5	4	5

Figure 7: Count to Infinity Problem illustration

2. **Hierarchical Routing:** As networks grow in size, the router routing tables grow proportionally. Not only the router memory consumed but also the more CPU time is needed to scan them and more bandwidth is needed.

The problem can be solved to some extent by using Hierarchical routing. In this routers are divided into regions as depicted in figure 8, with each router knowing all the details how to route packets to destination within it own region, but nothing about the internal structure of other regions.

For huge network, a two level hierarchy may be insufficient, it may be necessary to group the regions into clusters, the clusters into zones, the zones into groups and so on.

For example : When different networks are connected together, it is natural to regard each one as a separate region in order to free the routers in one network from having to know the topological structure of other ones.

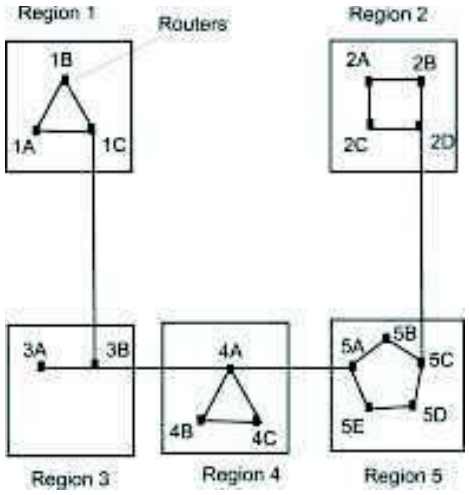


Figure 8: Hierarchical Routing

Full Table for 1A (Dest=Destination)

Dest	Line	Hops
1A		
1B	1B	1
1C	1C	1
2A	1B	2
2B	1B	3
2C	1B	3
2D	1B	4
3A	1C	3
3B	1C	2
4A	1C	2
4B	1C	4
4C	1C	4
5A	1C	4
5B	1C	5
5C	1B	5
5D	1C	6
5E	1C	5

Hierarchical Table for 1A

Dest	Line	Hops
1A		
1B	1B	1
1C	1C	1
2	1B	2
3	1C	2
4	1C	3
5	16	4

The full routing table as shown above for 1A has 17 entries but when routing is done hierarchically there are only 7 entries (e entries for local routers 4 entries for regions which are considered as single router. All the traffic for region 2 goes by 1B-2A line but rest of traffic goes by 1C-3B line.

The basic concept of link-state routing is that every node constructs a map of the connectivity to the network, in the form of a graph, showing which nodes are connected to which other nodes. Each node then independently calculates the next best logical path from it to every possible destination in the network. The collection of best paths will then form the node's routing table.

Disadvantages:

1. There is a penalty to be paid in the form of increased path length. For example the best route from 1A to 5C is via region 2, but with hierarchical routing all traffic to region 5 goes via region 3, because that is better for most destinations in region 5.

2. If single network become very large then multilevel hierarchy can be used.

   The presence of congestion means that the load is (temporarily) greater than the resources can handle.

3. Link state Routing

   Link State routing protocols (an adaptive routing algorithm) do not view networks in terms of adjacent routers and hop counts, but they build a comprehensive view of the overall network which fully describes the all possible routes along with their costs. Using the SPF (Shortest Path First) algorithm, the router creates a "topological database" which is a hierarchy reflecting the network routers it knows about. It then puts it's self on the top of this hierarchy, and has a complete picture from it's own perspective.

The complete working of algorithm can be divided into Five Steps:

1. Discover your neighbors and learn their addresses.

   In this process send “Hello”, packet on each point-to-point line. After receiving the hello packet Destination, node replies with its address.

2. Measure the cost (delay) to each neighbor.

   Send an “ECHO” packet over the line. Destination is required to respond to “ECHO” packet immediately. Measure the time required for this operation.

3. Construct a packet containing all this information

   The information tables are creating having all details of neighboring nodes.

4. Send this packet to all other routers.

   Use selective flooding. Sequence numbers prevent duplicate packets from being propagated. Lower sequence numbers are rejected as obsolete

5. Compute the shortest path to every other router.

   Dijkstra’s Shortest Path algorithm is used to determine the shortest path to each destination.

   When a router using a Link State protocol, such an OSPF (Open Shortest Path First) knows about a change on the network, it will broadcast this change instantly, there for flooding the network with this information. The information routers require to build their databases is provided in the form of Link State advertisement packets (LSAP). Routers do not advertise their entire routing tables; instead each router advertises only its information regarding immediately adjacent routers.

1.3.4 Comparison Link State Versus Distance Vector Routing

- Link state has big memory requirements
- In link state shortest path computations require many CPU circles
- Link state, If network is stable little bandwidth is used; react quickly to topology changes
- In link state announcements cannot be “filtered”. All items in the database must be sent to neighbors
- In link state all neighbors must be trusted
- In link state authentication mechanisms can be used to avoid undesired adjacencies
- In link state no split horizon techniques are possible

  Even though Link State protocols work more efficiently, problem can arise. Usually problems occur cause of changes in the network topology (links go up-down), and all routers don't get updated immediately cause they might be on different line speeds, there for, routers connected via a fast link will receive these changes faster than the others on a slower link.

Different techniques have been developed to deal with these problem and these are:

1. Dampen update frequency
2. Target link-state updates to multicast
3. Use link-state area hierarchy for topology
4. Exchange route summaries at area borders
5. Use Time-stamps Update numbering & counters
6. Manage partitions using a area hierarchy

1.4 CONGESTION CONTROL

Congestion: When too many packets are in a subnet or a part of subnet, performance degrades as depicted in figure 9. This situation is called congestion.

Factors Causing the Congestion

1. Many input lines demanding the same output lines.
2. Slow receiver fast sender.
3. Low bandwidth lines can also cause congestion.
4. Congestion itself (duplicacy).
5. Traffic is bursty.

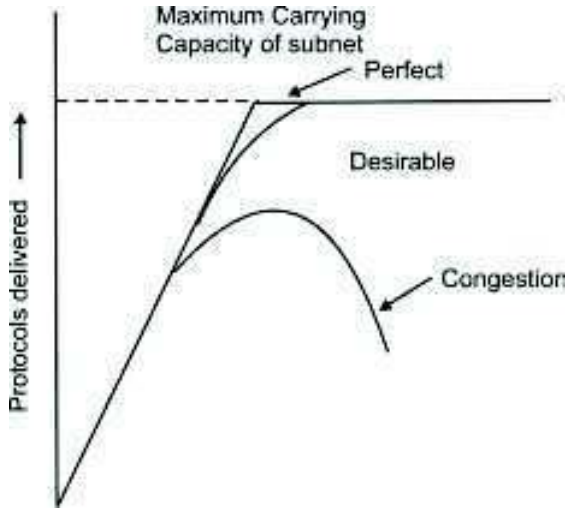


Figure 9: capacity of subnet and congestion

Congestion control principles are divided into two categories:

a) Open loop: In open loop solution the good designs are being developed to solve the problem so that congestion does not occur at first place once the system is setup and running, no mid pores connection is made. In open loop control, tools are included to decide when to accept new traffic, when to discard packets and which ones. And making scheduling decisions at various points in the network. The decisions are offline decision is not based on current state of network close loop solution.

   The concept of feedback loop is used in closed loop solution. This approach has three parts, when apply to the congestion control.

   i) Monitor the system to detect when and where congestion occurs.

   ii) Pass the information to places where the action can be taken.

   iii) Adjust system operation to correct the problem.

1.4.1 Algorithm for Congestion Control

Two major criterion of congestion control are

1. To decrease load

2. To increase capacity

Traffic Shaping (Congestion Control Policy in ATM)

One of the main causes of congestion is that traffic is often bursty. If hosts could be made to transmit a uniform rate, congestion would be less common. Another open loop method to help manage congestion is forcing the packet to be transmitted at a more predictable rate. This approach to congestion management is widely used in ATM networks and is called traffic shaping.

Traffic shaping is about regulating the average rate (and burstiness) of data transmission.

Leaky Bucket Algorithm

Imagine a bucket with a small hole in the bottom as depicted in figure 10. No matter at what rate enters the bucket, the outflow is at a constant rate, when there is any

water is bucket and zero when the bucket is empty. Also once the bucket is full, any additional water entering it spills over the sides and is lost.

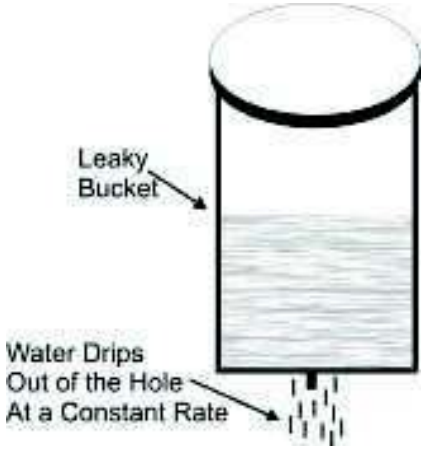


Figure 10: Normal Leaky bucket

The same idea can be applied to packets conceptually; each host is connected to the network by an interface containing a leaky bucket, i.e. a finite internal queue. If a packet arrives at the queue when it is full, the packet is discarded. This arrangement can be built into the hardware interface. It was first proposed by turner and is called leaky bucket algorithm as given figure 11.

Leaky bucket algorithm can be understood as "The leaky bucket consists of finite queue when a packet arrives, if there is room on the queue it is appended to the queue, otherwise it is discarded. At every clock tick, one packet is transmitted".

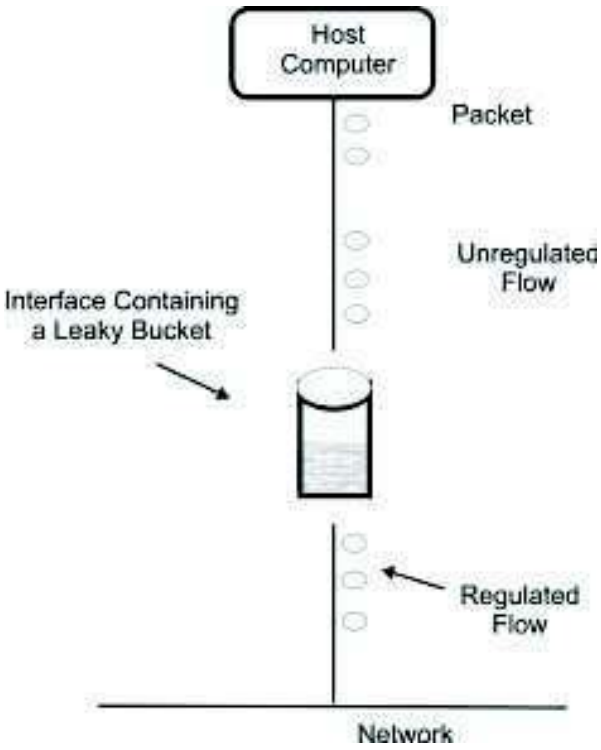


Figure 11: Leaky bucket algorithm

Advantage

This algorithm smoothens the bursts and greatly reduces the chances of congestion.

Disadvantages

1. When the queue is full, packets are discarded.

2. Sometimes it is necessary to speed up the output which is not possible in leaky bucket algorithm.

Token Bucket Algorithm: The leaky bucket algorithm has a rigid output pattern at the average rate, no matter how bursty the traffic is.

In many applications, it is better to allow the output to speed up somewhat when large bursts arrive so a more flexible algorithm is needed, preferably one that never losses data one such algorithm is token bus algorithm.

In token bucket algorithm, the leaky bucket holds 'tokens' generated by a clock at the rate of one token every DT sec.

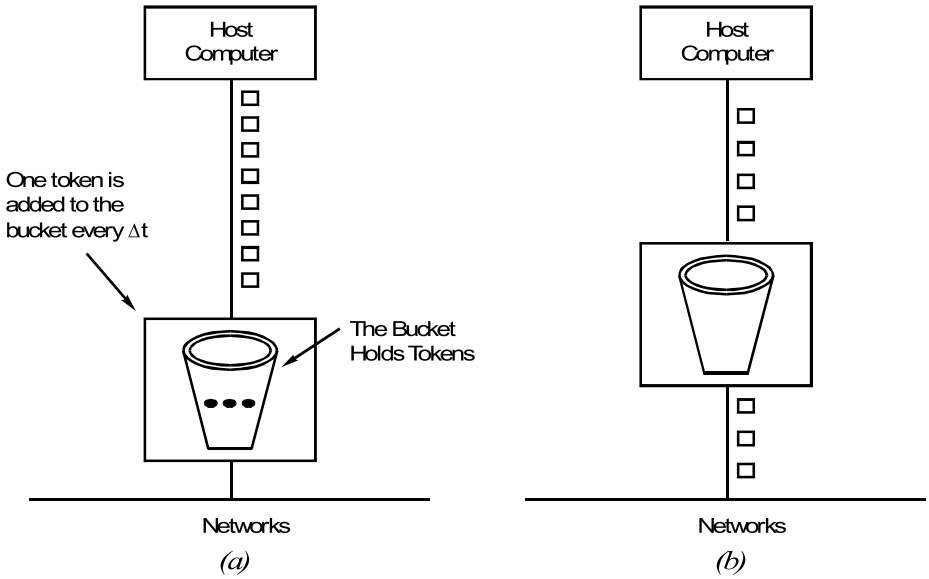


Figure 12: Token Bucket Algorithm

In above figure 12, we see a bucket holding three tokens, with five packets waiting to be transmitted. For a packet to be transmitted, it must capture and destroy one token. Three of fine packets have gone through, but other two are waiting for tokens to be generated.

The token bucket algorithm provides a different kind of traffic shaping than the leaky bucket algorithm. Bust of up to n. packets can be sent at once, allowing some burstiness in the output stream and giving faster response to sodden bursts of input.

A token bucket algorithm throws away tokens when the bucket fills up but never discards packets.

Table 1: Token Bucket V/S Leaky Bucket

S.No.	Token Bucket	Leaky Bucket
1.	The algorithm shaping is quite different.	In this there is a trade off between memory and bandwidth and packet life time.
2.	It allows saving up to maximum size of 'n' i.e. burst can be send of size 'n' at once.	It has constant traffic depending on the Leakage.
3.	Token bucket discarded token when bucket fills up.	This discards the packets when bucket fills

☞ Check Your Progress 1

1. The shortest path in routing can refer to……..

   a) The least expensive path
   b) The least distant path
   c) The path with the smallest number of hops
   d) Any or a combination of above

   ……

   ……

2. In distance vector routing, each router receives vector from

   a) Every router in the network
   b) Every router less than two units away
   c) A table stored by the software
   d) Its neighbor only

   ……

   ……

3. In link state routing, flooding allows changes to be recorded by

   a) All router
   b) Neighbor router only
   c) Some routers
   d) All networks

   ……

   ……

4. In which type of switching, do all the datagrams of a message follow the same channel of a path

   a) Circuit switching
   b) Datagram packet switching
   c) Virtual circuit packet switching
   d) Message switching

   ……

   ……

5. Which type of switching uses the entire capacity of a dedicated link

   a) Circuit switching
   b) Datagram packet switching
   c) Virtual circuit packet switching
   d) Message switching

   ……

   ……

1.5 NETWORK ADDRESSING

IP address versions

IP became the official protocol for the internet in 1983.As the internet has evolved, so has the IP. There have been six versions since its inception. Three versions are main.

1. Version 4(IPv4)
2. Version 5(Ipv5)
3. Version 6(Ipv6)

IP addressing

The identifier used in the IP layer of the TCP/IP protocol suite to identify each device connected to the Internet is called the Internet address or IP address.

OR

An IP address (Ipv4) is a 32-bit address that uniquely and universally defines the connection of a host or a router to the Internet.

There are **three** common ways in which IP addresses can be represented.

1. There is the binary notation which uses the base two number system to represent numbers.
2. There is the decimal notation which uses the base ten number system to represent numbers.
3. There is the hexadecimal notation which uses the base sixteen number system to represent numbers.

Do you know?

An IP address is a 32-bit(4-bytes) address.

Example 1

Assume, IGNOU's IP address is 142.190.23.180. This IP address consists of four bytes. The first byte has the value of 142. The second byte has the value of 190. The third byte has the value of 23, and the fourth byte has the value of 18.

Do you know?

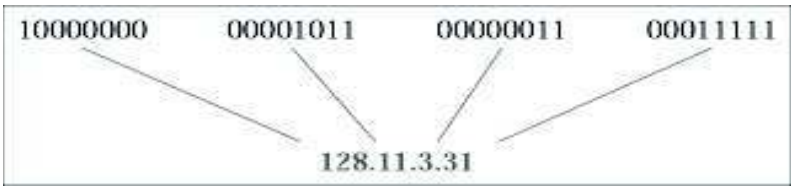
IP addresses are unique.

IP addresses are unique in the sense that each address defines one, and only one, connection to the Internet. Two devices on the Internet can never have the same address.

Do you know?

The address space of IPv4 is 2^{32} or 4,294,967,296.

Notations of IP addresses



Example 1

Change the following IP addresses from binary notation to dotted-decimal notation.

a) 10000001 00001011 00001011 11101111
b. 11000001 10000011 00011011 11111111
c. 11100111 11011011 10001011 01101111
d. 11111001 10011011 11111011 00001111

Solution

We replace each group of 8 bits with its equivalent decimal number and add dots for separation:

a) 129.11.11.239
b) 193.131.27.255
c) 231.219.139.111
d) 249.155.251.15

Example 2

Change the following IP addresses from dotted-decimal notation to binary notation.

a) 111.56.45.78
b) 221.34.7.82
c) 241.8.56.12
d) 75.45.34.78

Solution

We replace each decimal number with its binary equivalent:

a) 01101111 00111000 00101101 01001110
b) 11011101 00100010 00000111 01010010
c) 11110001 00001000 00111000 00001100
d) 01001011 00101101 00100010 01001110

Example 3

Find the error, if any, in the following IP addresses:

a) 111.56.045.78
b) 221.34.7.8.20
c) 75.45.301.14
d) 11100010.23.14.67

Solution

a) There are no leading zeroes in dotted-decimal notation (045).
b) We may not have more than four numbers in an IP address.
c) In dotted-decimal notation, each number is less than or equal to 255; 301 is outside this range.
d) A mixture of binary notation and dotted-decimal notation is not allowed.

Example 4

Change the following IP addresses from binary notation to hexadecimal notation.

a) 10000001 00001011 00001011 11101111
b) 11000001 10000011 00011011 11111111

Solution

We replace each group of 4 bits with its hexadecimal equivalent (see Appendix B). Note that hexadecimal notation normally has no added spaces or dots; however, 0X (or 0x) is added at the beginning or the subscript 16 at the end to show that the number is in hexadecimal.

a) 0X810B0BEF or 810B0BEF16

b) 0XC1831BFF or C1831BFF16

Classless Inter-Domain Routing allocates address space to Internet service providers and end users on any address bit boundary, instead of on 8-bit segments. CIDR notation is a syntax of specifying IP addresses and their associated routing prefix. It appends to the address a slash character and the decimal number of leading bits of the routing prefix, e.g., 192.0.2.0/24 for IPv4

1.5.1 Classful Addressing

IP addresses, when started a few decades ago, used the concept of classes. This architecture is called classful addressing. In the mid-1990s, a new architecture, called classless addressing, was introduced and will eventually supersede the original architecture. However, part of the Internet is still using classful addressing, but the migration is very fast.

Occupation of the address space

Class	Number of Addresses	Percentage
A	2^{31} = 2,147,483,648	50%
B	2^{30} = 1,073, 741, 824	25%
C	2^{29} = 536,870,912	12.5%
D	2^{28} = 268,435,456	6.25%
E	2^{28} = 268,435,456	6.25%

Table 13: Addresses per Class

When IP addresses were first started, they used the concept of classes. The range of IP addresses were divided into five classes: As shown in the figure 13 and 14, A, B, C, D, and E. Class A used up 50% of the address space, class B used up 25%, class C used up 12.5%, class D used up 6.25%, and class E also used up 6.25%.



Figure 14: Address Spaces of IPv4 classes

The way you recognize which class an IP address belongs to is by analyzing the first byte. If the number in the first byte is between 0-127, then the IP address is in the Class A range as shown in figure 15. If it is between 128-191 it is in Class B. If it is between 192-223 it is in the Class C range. If it is between 224-239 it is in the Class D range, and if it is between 240-255, then it belongs to Class E.

Class in decimal notations

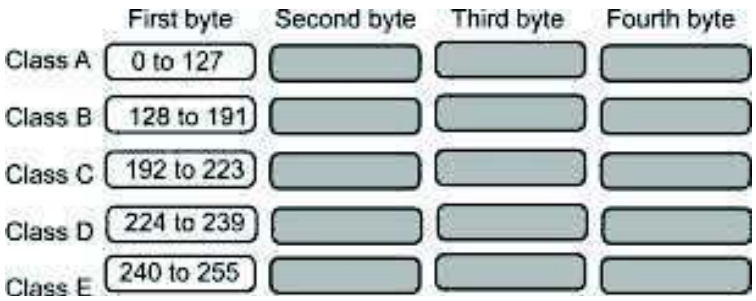


Figure 15: Classes of IPv4in decimal notation

Example

If IGNOU's IP address is 140.192.23.180. Looking at this address we can see that the first byte is 140. Since 140 is between the numbers 128-191, we know that it is in the Class B range.

Class in binary notation

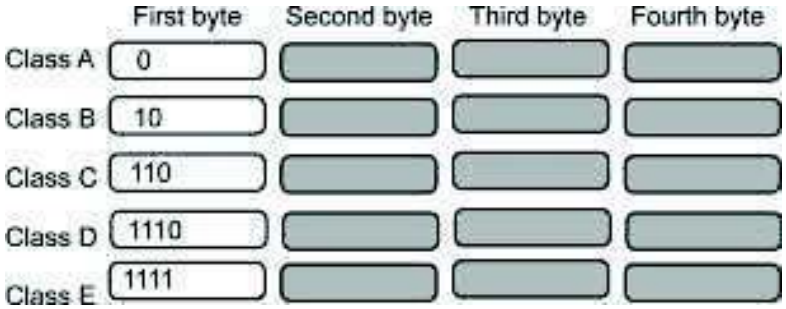


Figure 16: Classes of IPv4in binary notation

According to the figure 16, we can device a mechanism as given figure 17 for finding the address class in binary notation like:

If first left most bit is 0 then it is class A
If first bit is 1 and second bit is 0 then it is class B
If first two bits are 1 and third bit is 0 then it is class C
If first three bits are 1 and fourth bit is 0 then it is class D
If all the four bits are 1 then it is class E

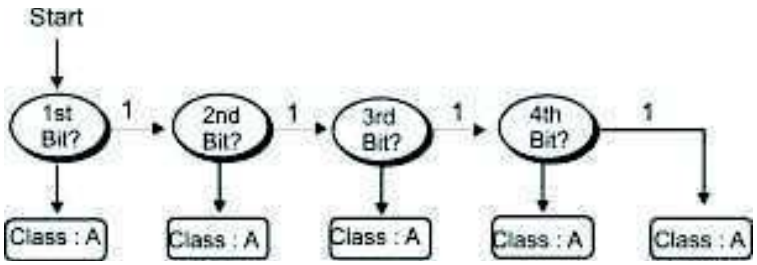


Figure 17: Finding the address class in binary notation

Example 5

How can we prove that we have 2,147,483,648 addresses in class A?

Solution

In class A, only 1 bit defines the class. The remaining 31 bits are available for the address. With 31 bits, we can have 2^{31} or 2,147,483,648 addresses.

Example 6

Find the class of each address:

a) 00000001 00001011 00001011 11101111
b) 11000001 10000011 00011011 11111111
c) 10100111 11011011 10001011 01101111
d) 11110011 10011011 11111011 00001111

Solution

a) The first bit is 0. This is a class A address.
b) The first 2 bits are 1; the third bit is 0. This is a class C address.
c) The first bit is 1; the second bit is 0. This is a class B address.
d) The first 4 bits are 1s. This is a class E address

Example 7

Find the class of each address:

a) 227.12.14.87
b) 193.14.56.22
c) 14.23.120.8
d) 252.5.15.111
e) 134.11.78.56

Solution

a) The first byte is 227 (between 224 and 239); the class is D.
b) The first byte is 193 (between 192 and 223); the class is C.
c) The first byte is 14 (between 0 and 127); the class is A.
d) The first byte is 252 (between 240 and 255); the class is E.
e) The first byte is 134 (between 128 and 191); the class is B.

Example 8

In Example 5 we showed that class A has 2^{31} (2,147,483,648) addresses. How can we prove this same fact using dotted-decimal notation?

Solution

The addresses in class A range from 0.0.0.0 to 127.255.255.255. We need to show that the difference between these two numbers is 2,147,483,648. This is a good exercise because it shows us how to define the range of addresses between two addresses. We notice that we are dealing with base 256 numbers here. Each byte in the notation has a weight. The weights are as follows

$$256^3, 256^2, 256^1, 256^0$$

Now to find the integer value of each number, we multiply each byte by its weight:

Last address: $127 \times 256^3 + 255 \times 256^2 +$

$$255 \times 256^1 + 255 \times 256^0 = 2{,}147{,}483{,}647$$

First address: = 0
If we subtract the first from the last and add 1 to the result (remember we always add 1 to get the range), we get 2,147,483,648 or 2^{31}.

1.5.2 NetID and HostID

An IP address is divided into a network ID (netid) and a host ID (hostid) as depicted in figure 18. The lengths of the netid vary depending on the class the IP address belongs to. In class A the netid occupies the first byte and the hostid occupies the remaining three bytes. In class B the netid occupies the first two bytes and the hostid occupies the remaining two bytes. In class C the first three bytes define the netid and the last remaining byte defines the hostid. Class D and E are not divided into netid and hostid. The following figure 18 shows how the netid and hostid are divided in Classes A, B, and C.

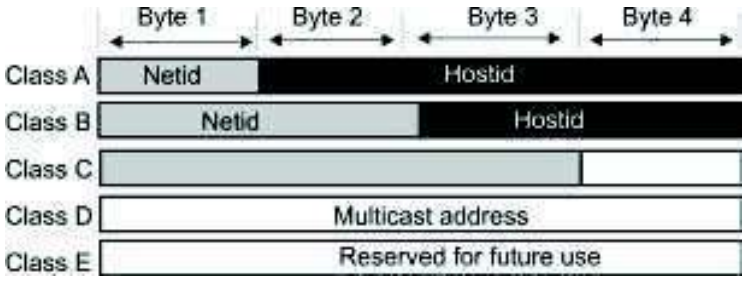


Figure 18: Network ID and a host ID of IPv4 classes

Do you know?

Class D addresses are used for multicasting; there is only one block in this class and Class E addresses are reserved for future purposes; most of the block is wasted.

In classful addressing the netid and hostid are easily distinguishable by looking at the IP address. First you have to determine which class the IP address belongs to and from there you can tell which part is the netid and which part is the hostid. If it is in Class A, then the first byte represents the netid and the last three represent the hostid, and so on.

Disadvantages of class full addressing

It wastes a lot of IP addresses and since the Internet keeps growing larger, we can't afford to throw away IP addresses.

That is why a new addressing scheme was devised. It is called classless addressing because it doesn't use the classes which were used in classful addressing.

Do you know?

Millions of class A and class B addresses and are wasted in class full addressing

Do you know?

The number of addresses in class C is smaller than the needs of most organizations.

Example 9

Given the network address 17.0.0.0, find the class, the block, and the range of the addresses.

Solution

The class is A because the first byte is between 0 and 127. The block has a netid of 17. The addresses range from 17.0.0.0 to 17.255.255.255.

Example 10

Given the network address 132.21.0.0, find the class, the block, and the range of the addresses.

Solution

The class is B because the first byte is between 128 and 191. The block has a netid of 132.21. The addresses range from 132.21.0.0 to 132.21.255.255.

Example 11

Given the network address 220.34.76.0, find the class, the block, and the range of the addresses.

Solution

class is C because the first byte is between 192 and 223. The block has a netid of 220.34.76. The addresses range from 220.34.76.0 to 220.34.76.255.

☞ Check Your Progress 2

1. Which IP address class has few hosts per network

   a) Class A
   b) Class B
   c) Class C
   d) Class D

   ……

   ……

2. Which of the following is true about IP addresses

   a) It is divided into exactly two classes
   b) It contains a fixed length host-id
   c) It was established as a user friendly interface
   d) It is 32 bits long

   ……

   ……

3. Which of the following is class C host address

   a) 230.0.0.0
   b) 130.4.4.6
   c) 200.1.2.3
   d) 30.4.5.6

   ……

   ……

1.6 FRAGMENTATION

Each network imposes some maximum size on its packets. A problem appears when a large packet wants to travel through a network whose maximum packet size is too small. One solution is to make sure the problem does not occur in the first place. In other words, the internet should use a routing algorithm that avoids sending packets through networks that cannot handle them. However, this solution is no solution at all. What happens if the original source packet is too large to be handled by the destination network? The routing algorithm can hardly bypass the destination.

Basically, the only solution to the problem is to allow gateways to break up packets into **fragments,** sending each fragment as a separate internet packet. However, converting a large object into small fragments is considerably easier than the reverse process.

Two opposing strategies exist for recombining the fragments back into the original packet.

1. Transparent Fragmentation
2. Non Transparent Fragmentation

Transparent Fragmentation

The first strategy is to make fragmentation caused by a "small-packet" network transparent to any subsequent networks through which the packet must pass on its way to the ultimate destination. This option is shown in Figure 19 (a). In this approach, the small-packet network has gateways that interface to other networks. When an oversized packet arrives at a gateway, the gateway breaks it up into fragments. Each fragment is addressed to the same exit gateway, where the pieces are recombined. In this way, passage through the small-packet network has been made transparent. Subsequent networks are not even aware that fragmentation has occurred.

Do you know?

ATM networks have special hardware to provide transparent fragmentation of packets into cells and then reassembly of cells into packets. In the ATM world, fragmentation is called segmentation

Transparent fragmentation is straightforward as shown in figure 19 a.

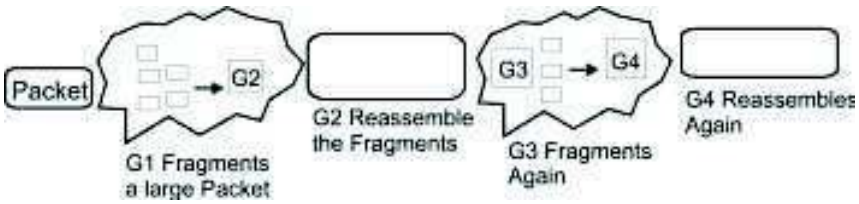


Figure 19 a: Transparent fragmentation

Drawback of transparent fragmentation

1. The exit gateway must know when it has received all the pieces, so either a count field or an "end of packet" bit must be provided.

2. All packets must exit via the same gateway. By not allowing some fragments to follow one route to the ultimate destination and other fragments a disjoint route. some performance may be lost.

3. A last problem is the overhead required to repeatedly reassemble and then re-fragment a large packet passing through a series of small packet networks.

Tips

ATM requires transparent fragmentation.

Non Transparent Fragmentation

The nontransparent fragmentation strategy refrains from recombining fragments at any intermediate gateways. Once a packet has been fragmented, each fragment is treated as though it were an original packet. All fragments are passed through the exit gateway (or gateways), as shown in Figure 19 (b). Recombination occurs only at destination host. IP works this way.

Non Transparent fragmentation also has some problems. For example, it requires every host to be able to do reassembly. Yet another problem is that when the large packet is fragmented the total overhead increases, because each fragment must have a header.

An advantage of this method is that multiple exit gateways can now be used and higher performance can be achieved

When a packet is fragmented, the fragments must be numbered in such a way that the original data stream can be reconstructed. One way of numbering the fragments is to use a tree. If packet 0 must be split up, the pieces are called 0.0, 0.1, 0.2 etc. If these fragments themselves must be fragmented later on, the pieces are numbered 0.0.0, 0.0.1, 0.0.2, ….0.1.2 etc. If enough fields have been reserved in the header for the worst case and no duplicates generated anywhere, this scheme is sufficient to ensure that all the pieces can be correctly reassembled at the destination, no matter what order they arrive in.
However, if even one network loses or discards packets, end-to-end retransmissions are needed, with unfortunate effects for the numbering system. Suppose that a 1024-bit packet is initially fragmented into four equal-sized fragments, 0.0, 0.1, 0.2 and 0.3. Fragment 0.1 is lost, but the other parts arrive at the destination. Eventually, the source times out and retransmits the original packet again. Only this time the route taken passes through a network with a 512-bit limit, so two fragments are generated. When the new fragment 0.1 arrives at the destination, the receiver will think that all four pieces are now accounted for a reconstruct the packet incorrectly.

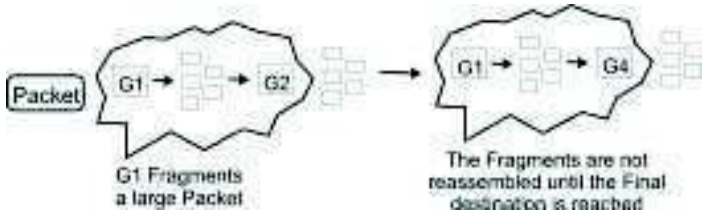


Figure 19 b: Nontransparent fragmentation

1.7 ERROR MESSAGING SERVICES

1.7.1 ICMP (Internet Control Message Protocol)

The Internet Control Message Protocol (ICMP) is a helper protocol that supports IP with facility for Error reporting Simple queries.

ICMP messages are sent in following situations

- when a datagram cannot reach its destination,
- when the gateway does not have the buffering capacity to forward a datagram,
- When the gateway can direct the host to send traffic on a shorter route.

Do you know?
ICMP is considered an integral part of IP as shown in figure 20.

The Internet Protocol (IP) is not designed to be absolutely reliable. The purpose of these control messages is to provide feedback about problems in the communication environment, not to make IP reliable. There are still no guarantees that a datagram will be delivered or a control message will be returned. Some datagrams may still be undelivered without any report of their loss. The higher level protocols that use IP must implement their own reliability procedures if reliable communication is required. The ICMP messages typically report errors in the processing of datagrams. To avoid the infinite regress of messages about messages etc., no ICMP messages are sent about ICMP messages.

TIPS
ICMP provides error reporting, flow control and first-hop gateway redirection.

Position of ICMP in the network layer



Figure 20: Protocols of Internet layer (TCP/IP)

ICMP header format

It consists of following fields

- Type
- Code
- ICMP header checksum
- Data

00 01 02 03 04 05 06 07	08 09 10 11 12 13 14 15	16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Type	Code	ICMP header checksum
Data		

The fields can be described as follows

1. **Type.** It is of 8 bits. It specifies the format of the ICMP message.
2. **Code.** It is of 8 bits. It further qualifies the ICMP message.
3. **ICMP Header Checksum.** It is of 16 bits. It is checksum that covers the ICMP message. This is the 16-bit one's complement of the one's complement sum of the ICMP message starting with the Type field. The checksum field should be cleared to zero before generating the checksum.
4. **Data.** It is of variable length. It contains the data specific to the message type indicated by the Type and Code fields

Types of ICMP messages

Each ICMP message contains three fields that define its purpose and provide a checksum. They are

- TYPE,
- CODE, and
- CHECKSUM fields (described above).

The TYPE field identifies the ICMP message, the CODE field provides further information about the associated TYPE field, and the CHECKSUM provides a method for determining the integrity of the message.

Tips
ICMP message are sent as packet so these are also called ICMP packet

ICMP messages are divided into two categories

- Error-reporting messages.
- Query messages.

Tips
ICMP messages are identified by "type" numbers

The **error-reporting** messages report problems that a router or a host (destination) may encounter. The **query messages** get specific information from a router or another host. For example, this can be used by the hosts to discover the routers present in their network. The host would send a ICMP query asking for routers to respond. The outers present in the network will respond with an ICMP reply message. The host would get information about the router from this reply.

Examples of error reporting messages

- Destination unreachable
- Source quench
- Time exceeded
- Parameter problem
- Redirection

Example of query messages

- Echo request and reply
- Timestamp request and reply
- Address mask request and reply
- Router solicitation and advertisement

1.7.2 IGMP(Internet Group Message Protocol)

IGMP is a protocol that manages group membership. The IGMP protocol gives the multicast routers information about the membership status of hosts (routers) connected to the network.

Tips

Internet Group Management Protocol (IGMP) is the protocol used to support multicasting.

Position of IGMP in the network layer



Types of messages in IGMP

IGMP has three types of messages:

- the query,
- the membership report,
- and the leave report.

There are two types of query messages as shown in figure 21, general and special

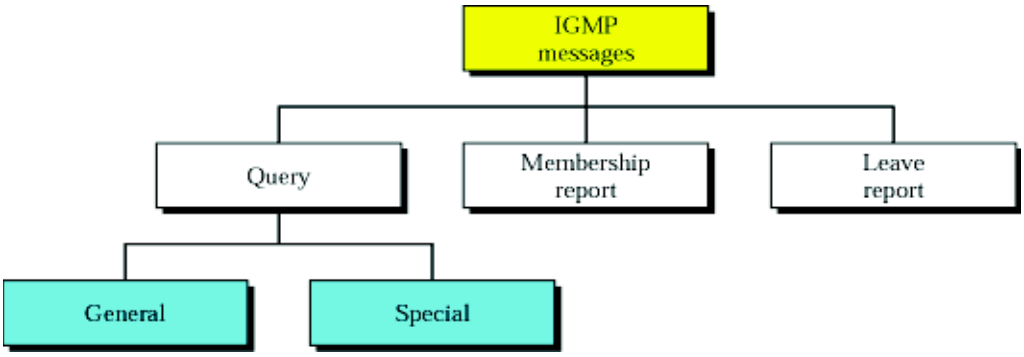


Figure 21: Types of query messages in IGMP

IGMP frame format

It consists of following fields

- Type
- Maximum response time
- Checksum
- Group addresses

00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Type								Code								IGMP Checksum															
Identifier																															
Group Address																															
Access Key																															

The fields can be described as follows

1. **Type.** It is of 8 bits. Following types of IGMP messages are possible

2. **Code.** It is of 8 bits. In a Create Group Request message, this field indicates if the new host group is to be public or private. In all other Request messages, this field is set to zero.

   In a Reply message, the *Code* field specifies the outcome of the request.

3. **IGMP Checksum.** It is of 16 bits. The checksum is the 16-bit one's complement of the one's complement sum of the IGMP message starting with the IGMP Type. For computing the checksum, the checksum field should first be cleared to 0. When the data packet is transmitted, the checksum is computed and inserted into this field. When the data packet is received, the checksum is again computed and verified against the checksum field. If the two checksums do not match then an error has occurred.

4. **Identifier.** It is of 32 bits. In a confirm Group Request message, the identifier field contains zero. In all other Request messages, the identifier field contains a value to distinguish the request from other requests by the same host. In a Reply message, the identifier field contains the same value as in the corresponding Request message.

5. **Group Address.** It is of 32 bits. In a Create Group Request message, the group address field contains zero. In all other Request messages, the group address field contains a host group address. In a Create Group Reply message, the group address field contains either a newly allocated host group address (if the request is granted) or zero (if denied). In all other Reply messages, the group address field contains the same host group address as in the corresponding Request message.

6. **Access Key.** This field is of 64 bits. In a Create Group Request message, the access key field contains zero. In all other Request messages, the access key field contains the access key assigned to the host group identified in the Group Address field (zero for public groups). In a Create Group Reply message, the access key field contains either a non-zero 64-bit number (if the request for a private group is granted) or zero. In all other Reply messages, the access key field contains the same access key as in the corresponding Request.

Do you know?

IGMP is defined in RFC 1112.

1.8 SUMMARY

In this unit, we studied various design issues of network layer. Network layer provides best route from source to destination using adaptive routing algorithm like distance vector routing and link state routing. A serous drawback of distance vector routing is count to infinity problem. It is also responsible for congestion control using leaky bucket and token leaky bucket algorithm. The four main protocols that operates on network layer are ARP, RARP, ICMP, IGMP. Network layer mainly works on IP address. IP addresses are 32bits.IP addresses have been divided into five classes namely A,B,C,D,E.ICMP and .ICMP and IGMP are error reporting protocols.ARP and RARP are used for address translation.

1.9 REFERENCES/FURTHER READING

1. Introduction to Data Communication & Networking, 3rd Edition, Behrouz Forouzan, Tata McGraw Hill.

2. Computer Networks, A. S. Tanenbaum 4th Edition, Practice Hall of India, New Delhi. 2003.
3. Douglas E. Comer, Internetworking with TCP/IP Vol.1: Principles, Protocols, and Architecture (4th Edition).
4. James F. Kurose, Computer Networking: A Top-Down Approach Featuring the Internet (3rd Edition).
5. Larry L. Peterson, Computer Networks: A Systems Approach, 3rd Edition (The Morgan Kaufmann Series in Networking).
6. www. wikipedia.org
7. W. Richard Stevens, The Protocols (TCP/IP Illustrated, Volume 1).
8. William Stallings, Data and Computer Communications, Seventh Edition.

1.10 SOLUTION/ANSWERS

☞ Check Your Progress 1

1. D
2. D
3. A
4. C
5. A

☞ Check Your Progress 2

1. C
2. D
3. C

UNIT 3 APPLICATION LAYER

Structure **Page No.**

3.0 Introduction 48
3.1 Objectives 49
3.2 Client Server Architecture 49
3.3 Domain Name Server (DNS) 49
3.3.1 The DNS name space
3.3.2 Resource records
3.3.3 Name servers
3.3.4 Remote login (telnet)
3.3.5 The telnet application
3.3.6 The telnet protocol
3.4 Remote Login (Telnet) 54
3.4.1 The FTP application
3.4.2 The FTP protocol
3.5 File Transfer Protocol (FTP) 55
3.6 Network Management 56
3.6.1 Configuration management
3.6.2 Reconfiguration
3.6.3 Documentation
3.6.4 Fault management
3.6.5 Reactive fault management
3.6.6 Proactive fault management
3.6.7 Performance management
3.6.8 Security management
3.6.9 Accounting management
3.6.10 SNMP protocol
3.7 Word Wide Web and Client Server Applications 61
3.7.1 Architectural overview (www)
3.8 Electronic Mail 70
3.8.1 Architecture and services
3.8.2 The user agent
3.9 Summary 73
3.10 References/Further Readings 73
3.11 Solutions / Answers 73

3.0 INTRODUCTION

The application layer contains a variety of protocols that are commonly needed by users. One widely-used application protocol is HTTP (Hyper Text 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. The application layer enables the user, whether human or software, to access the network. It provides user interfaces and support for services such as electronic mail, remote file access and transfer, shared database management, and other types of distributed information services. Application-layer functions typically include identifying communication partners, determining resource availability, and synchronizing communication. When identifying communication partners, the application layer determines the identity and availability of communication partners for an application that has data to transmit. When determining resource availability, the application layer must decide whether sufficient network or the requested communication exists. This unit covers the details of all important functions and

services of Application layer including different protocols available in Application layer.

3.1 OBJECTIVES

After going through this unit, you will be able to:

- define and know the functions of Application layer
- Understand the working of Domain Name Server
- Know the features and services of Telnet and FTP
- Understand the Network Management issues
- Understanding of the WWW Client/Server architecture
- Know the Email Architecture and Services

3.2 CLIENT SERVER ARCHITECTURE

Client/server architecture means one or more computers is/are acting as client one computer is behaving as serve which preferably of high configuration. In simple words client is defined as requester of services and server is defined as provider of services. Client and Servers are nothing but programs, which generally run on different machines/computers.

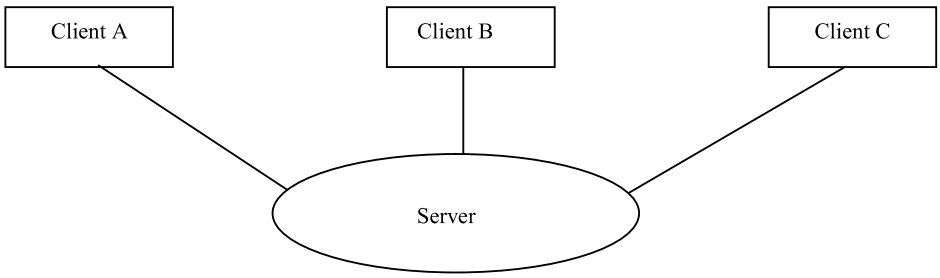


Figure 1: Client/Server Architecture

When client wants some information or wants to perform some task by server it has to contact the server as shown in figure1. Therefore, the client initially needs the address of the server, and tries to contact the server using that address. Server is a program which is always running and waiting to serve clients. Server welcomes the client (if free) as the client contacts it. After contacting the server, the server welcomes that client (if free), the client informs about its own address to servers so that server can reply the client also.

When both clients and servers exchange important information to each other connection establishes. Once with connection or link established, both client and server could exchange any information with each other. We will read more details of client/server architecture further in this unit during discussion about Word Wide Web.

3.3 DOMAIN NAME SERVER (DNS)

Programs theoretically could refer to hosts, mailboxes, and other resources by their network (e.g., IP) addresses, these addresses are hard for people to remember. Also, sending e-mail to *ravi@128.111.45.46* means that if Ravi's ISP or organization moves the mail server to a different machine with a different IP address, his e-mail address has to change. Consequently, ASCII names were introduced to decouple machine names from machine addresses. In this way, Ravi's address might be something like ravi@art.ucsb.edu. Nevertheless, the network itself understands only numerical addresses, so some mechanism is required to convert the ASCII strings to network addresses.

Way back in the ARPANET, there was simply a file, hosts.txt that listed all the hosts and their IP addresses. Every night, all the hosts would fetch it from the site at which it was maintained. For a network of a few hundred large timesharing machines, this approach worked reasonably well.

However, when thousands of minicomputers and PCs were connected to the net, everyone realized that this approach could not continue to work forever. For one thing, the size of the file would become too large. However, even more important, host name conflicts would occur constantly unless names were centrally managed, something unthinkable in a huge international network due to the load and latency. To solve these problems, **DNS** (the **Domain Name System**) was invented.

The essence of DNS is the invention of a hierarchical, domain-based naming scheme and a distributed database system for implementing this naming scheme. It is primarily used for mapping host names and e-mail destinations to IP addresses but can also be used for other purposes. Very briefly, the way DNS is used as follows. To map a name onto an IP address, an application program calls a library procedure called the **resolver**, passing it the name as a parameter. The resolver sends a UDP packet to a local DNS server, which then looks up the name and returns the IP address to the resolver, which then returns it to the caller. Armed with the IP address, the program can then establish a TCP connection with the destination or send it UDP packets.

3.3.1 The DNS Name Space

Managing a large and constantly changing set of names is a nontrivial problem. In the postal system, name management is done by requiring letters to specify (implicitly or explicitly) the country, state or province, city, and street address of the addressee. Conceptually, the Internet is divided into over 200 top-level **domains**, where each domain covers many hosts. Each domain is partitioned into sub domains, and these are further partitioned, and so on. All these domains can be represented by a tree, as shown in Figure 2. The leaves of the tree represent domains that have no sub domains (but do contain machines, of course). A leaf domain may contain a single host, or it may represent a company and contain thousands of hosts.

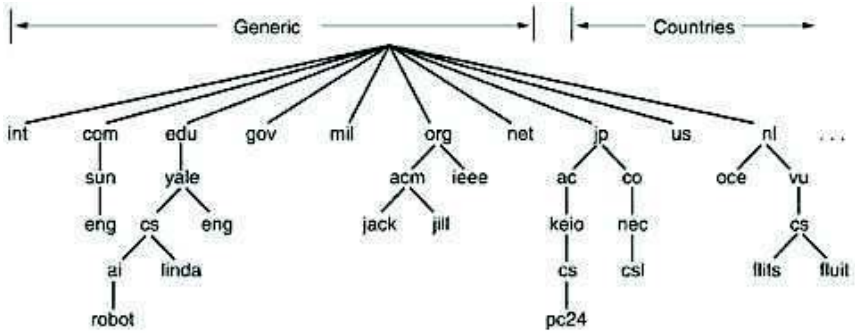


Figure 2: A portion of the Internet domain name space

The top-level domains come in two flavors: **generic** and **countries**. The original generic domains were com (commercial), edu (educational institutions), gov (the Government), int (certain international organizations), mil (the U.S. armed forces), net (network providers), and org (nonprofit organizations). The country domains include one entry for every country, as defined in ISO 3166.

In November 2000, ICANN approved four new, general-purpose, top-level domains, namely, biz (businesses), info (information), name (people's names), and pro (professions, such as doctors and lawyers). In addition, three more specialized top-level domains were introduced at the request of certain industries. These are aero (aerospace industry), coop (co-operatives), and museum (museums). Other top-level domains will be added in the future.

As an aside, as the Internet becomes more commercial, it also becomes more contentious. Take pro, for example. It was intended for certified professionals.

In general, getting a second-level domain, such as name-of-company.com, is easy. It merely requires going to a registrar for the corresponding top-level domain (com in this case) to check if the desired name is available and not somebody else's trademark. If there are no problems, the requester pays a small annual fee and gets the name. By now, virtually every common (English) word has been taken in the com domain. Try household articles, animals, plants, body parts, etc. Nearly all are taken.

Each domain is named by the path upward from it to the (unnamed) root. The components are separated by periods (pronounced "dot"). Domain names can be either absolute or relative. An absolute domain name always ends with a period (e.g., eng.sun.com.), whereas a relative one does not. Relative names have to be interpreted in some context to uniquely determine their true meaning. In both cases, a named domain refers to a specific node in the tree and all the nodes under it.

Domain names are case insensitive, so edu, Edu, and EDU mean the same thing. Component names can be up to 63 characters long, and full path names must not exceed 255 characters.

In principle, domains can be inserted into the tree in two different ways. For example, cs.yale.edu could equally well be listed under the US country domain as cs.yale.ct.us. In practice, however, most organizations in the United States are under a generic domain, and most outside the United States are under the domain of their country. There is no rule against registering under two top-level domains, but few organizations except multinationals do it (e.g., sony.com and sony.nl).

Each domain controls how it allocates the domains under it. For example, Japan has domains ac.jp and co.jp that mirror edu and com. The Netherlands does not make this distinction and puts all organizations directly under nl. Thus, all three of the following are university computer science departments:

1. cs.yale.edu(Yale University, in the United States)

2. cs.vu.nl(Vrije Universiteit, in The Netherlands)

3. cs.keio.ac.jp(Keio University, in Japan)

To create a new domain, permission is required of the domain in which it will be included. For example, if a VLSI group is started at Yale and wants to be known as vlsi.cs.yale.edu, it has to get permission from whoever manages cs.yale.edu.

Similarly, if a new university is chartered, say, the University of Northern South Dakota, it must ask the manager of the edu domain to assign it unsd.edu. In this way, name conflicts are avoided and each domain can keep track of all its subdomains. Once a new domain has been created and registered, it can create subdomains, such as cs.unsd.edu, without getting permission from anybody higher up the tree.

Naming follows organizational boundaries, not physical networks. For example, if the computer science and electrical engineering departments are located in the same building and share the same LAN, they can nevertheless have distinct domains.

3.3.2 Resource Records

Every domain, whether it is a single host or a top-level domain, can have a set of **resource records** associated with it. For a single host, the most common resource record is just its IP address, but many other kinds of resource records also exist. When a resolver gives a domain name to DNS, what it gets back are the resource records associated with that name. Thus, the primary function of DNS is to map domain names onto resource records.

A resource record is a *five-tuple*. Although they are encoded in binary for efficiency, in most expositions, resource records are presented as ASCII text, one line per resource record. The format we will use is as follows:

- **Domain_name, Time_to_live, Class Type, Value**

The Domain_name tells the domain to which this record applies. Normally, many records exist for each domain and each copy of the database holds information about multiple domains. This field is thus the primary search key used to satisfy queries. The order of the records in the database is not significant.

The Time_to_live field gives an indication of how stable the record is. Information that is highly stable is assigned a large value, such as 86400 (the number of seconds in 1 day). Information that is highly volatile is assigned a small value, such as 60 (1 minute).

The third field of every resource record is the Class. For Internet information, it is always IN. For non-Internet information, other codes can be used, but in practice, these are rarely seen.

The Type field tells what kind of record this is. The most important types are listed in Table. 1.2.2.

Table 1: The principal DNS Resource Record

Type	Meaning	Value
SOA	Start of Authority	Parameters for this zone
A	IP address of a host	32-Bit integer
MX	Mail exchange	Priority, domain willing to accept e-mail
NS	Name Server	Name of a server for this domain
CNAME	Canonical name	Domain name
PTR	Pointer	Alias for an IP address
HINFO	Host description	CPU and OS in ASCII
TXT	Text	Uninterpreted ASCII text

3.3.3 Name Servers

In theory at least, a single name server could contain the entire DNS database and respond to all queries about it. In practice, this server would be so overloaded as to be useless. Furthermore, if it ever went down, the entire Internet would be crippled.

To avoid the problems associated with having only a single source of information, the DNS name space is divided into non-overlapping **zones**. One possible way to divide

the name space is shown in figure 3. Each zone contains some part of the tree and also contains name servers holding the information about that zone.

Normally, a zone will have one primary name server, which gets its information from a file on its disk, and one or more secondary name servers, which get their information from the primary name server. To improve reliability, some servers for a zone can be located outside the zone.

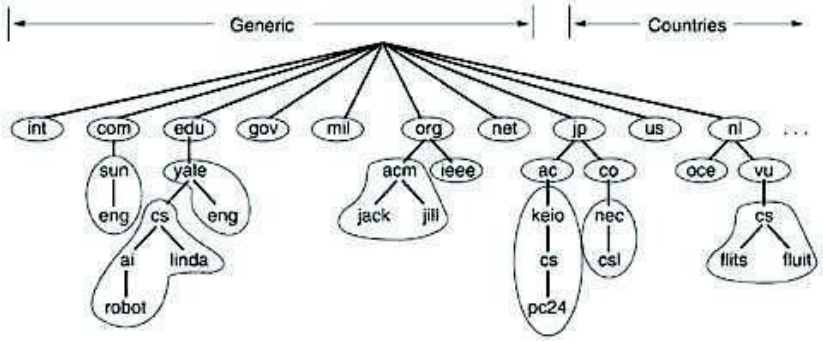


Figure 3: Part of the DNS name space showing the division into zones.

Where the zone boundaries are placed within a zone is up to that zone's administrator. This decision is made in large part based on how many name servers are desired, and where.

☞ Check Your Progress 1

1. Select the right choice.

   a) Which Layer is not present in TCP/IP model?
   (A) Application Layer (B) Internet Layer
   (C) Transport Layer (D) Presentation Layer

   ……

   ……

   b) Let most segment of a name inn DNS represents
   (A) Individual Network. (B) Individual computer.
   (C) Domain name (D) Network type.

   ……

   ……

   c) Address 192.5.48.3 belongs to
   (A) Class A. (B) Class B.
   (C) Class C. (D) Class D.

   ……

   ……

2. Discuss the DNS Name Space?

   ……

   ……

3. Write a note on Name Server?

3.4 REMOTE LOGIN (TELNET)

Telnet permits a user to connect to an account on a remote machine. A client program running on the user's machine communicates using the Telnet protocol with a server program running on the remote machine.

3.4.1 The Telnet Application

The user (say Ravi) has an account on both the local and remote machines.
For example, 'Ravi' on farkas.mcmaster.ca types telnet optlab.cas.mcmaster.ca at his user prompt. Here, farkas.mcmaster.ca is the client and optlab.cas.mcmaster.ca is the server. The Telnet client would perform a methotrexate() call to determine the IP address of optlab.cas.mcmaster.ca. Then the client would create a socket to communicate with the Telnet server. The server prompts 'Ravi' for a login identifier - the name of the user's account on the remote server followed by a password. The Telnet users interact with the remote machine in the same way they would interact with their local machine. The client relays Ravi's keystrokes to the remote server, and the remote server displays them on its pseudo terminal (which is actually the display screen on the client machine).

3.4.2 The Telnet Protocol

The Telnet client program performs two important functions - **interacting** with the user terminal on the local host and **exchanging messages** with the Telnet server. The client connects to port 23 on the remote machine, which is the port number reserved for Telnet servers. The TCP connection persists for the duration of the login session. The client and the server maintain the connection, even when the user interrupts the transfer of data, for example by hitting *cntl-C*. Since Telnet is designed to work over two hosts on different platforms, the protocol assumes that the two hosts run a Network Virtual Terminal (NVT).The TCP connection is set up across these two NVT terminals. The NVT is a very simple character device with a keyboard and a printer - data typed by the user on the keyboard is translated by the client software into NVT format and sent via its NVT terminal to the server, and data received in NVT format from the server is translated by the client into the local machine format and output to the printer.

The NVT terminals on the two hosts exchange data in the 7-bit U.S. variant of the ASCII format, with each character sent as an octet with the first bit set to19-21-10. Some control information, such as end-of-line indication, is transmitted as the character sequence CR (carriage return) followed by an LF (linefeed). Each Telnet control message starts with the special octet (Interpret as Command (IAC)) octet of all 1s) to ensure that the recipient interprets the subsequent octets as a command. Otherwise, each octet is interpreted as data (e.g., a user keystroke). Sending control messages on the same connection as the data is referred to as *inband signaling*. The initial control messages between the client and the server are used to exchange information about their capabilities (Telnet option negotiation). For example, the client may indicate the type and speed of its terminal, and whether data is to be sent one character or one line at a time. After the capabilities exchange, the server instructs the client to send a login identifier and password. Once the authentication completes, the user interacts directly with the remote machine. The client application relays user keystrokes to the server, and the server relays the output back to the client, using inband signaling, with the interpretation that commands follow the IAC octet of all ones. Telnet cannot rely on the conventional data stream alone to carry such control sequences between client and server.

3.5 FILE TRANSFER PROTOCAL (FTP)

FTP allows a user to copy files to and from a remote machine. The client program also sends commands to the server program to coordinate the copying of files between the two machines on behalf of the user.

3.5.1 The FTP Application

The FTP client connects to the remote machine which prompts the user to enter a login identifier and a password. However, some users may not have their own accounts on the remote machine. To grant access to a broad set of users, many FTP servers have a special account (e.g. *anonymous*) that does not require password information. Instead, the user logs in using *guest* or his email address as password. The FTP server coordinates access to a collection of files in various directories. In case of anonymous FTP, the server typically has a special directory, with one or more subdirectories, that can be accessed by the client. The user logged into the FTP server can traverse through the directories of files on the remote machine, and send or receive files. This is typically done via the command-line interface. The interface may also allow the client to send or receive multiple files with a single command. Recent FTP client applications provide a menu-based graphical user interface. For instance, a Web browser allows users to specify the desired file as an URL (e.g.,ftp://ftp.optlab.cas.mcmaster.ca/midterm.pdf). In this case, the web browser connects to the FTP server as an anonymous user and sends a sequence of FTP commands to fetch the requested file.

3.5.2 The FTP protocol

FTP differs from other applications such as Telnet since it uses separate TCP connections for control and data. Recall that in Telnet both control information and data are sent over the same TCP connection using in-band signaling. The two TCP connections in FTP are:

1. **The control connection** is established in the normal client-server fashion. In this case, the server does a *passive open* (is listening) on port 21 for FTP, and waits for the client connection. The client does an *active open* (the2nd handshake in a TCP connection) to establish the control connection. The client uses an ephemeral port number (above 1023) for the control connection. This control connection stays up for the entire time that the client communicates with this server. This connection is used for commands from the client to the server and for the server's replies. The IP type of service for the control connection should be to minimize delay in passing these commands over the TCP connection.

2. **A data connection** is created each time a file is transferred between the client and the server. The IP type of service for the data connection should be to maximize throughput since this connection is file transfer, and we want to send this entire file over a high bandwidth line. The specification of FTP includes more than 30 different commands, which are transmitted over the control connection in NVT ASCII format. The commands are not case-sensitive and may have arguments; each command ends with a two character sequence of a carriage return (CR) followed by a line feed (LF). It must be emphasized here that these commands are different from the commands typed by the user at the interface provided by the client. Transferring a single file for instance requires only a single user-level command (e.g., *put* or *get*), but this single command triggers the client to send a set of FTP commands to the server. The FTP server responds to each command with a three-digit reply code (for the FTP client) and an optional text message (for the user).

The control connection persists over a sequence of FTP commands, as the client and the server continue their dialogue. The typical interaction starts with a command that identifies the user on the server machine followed by another command to send the user password. The arguments for these commands are gleaned from the user's input (his account name and password). The server uses this information to verify whether the user has an authorized account on the remote machine, and in the case of anonymous FTP decides on the set of19-21-3directories to which the anonymous guest has access. The next set of commands depends on the user request to send, receive, or view the files in a present directory.

The actual file (data) transfer uses a separate TCP connection established by the host sending the data. For instance, if the user wants to retrieve the file *midterm.pdf* from the remote server, the server initiates the creation of the TCP data connection. In case, the user wants to put a file into the remote machine, it is the client who initiates the creation of the TCP connection. The data connection is usually established on port 20 on the server machine. In the former case (when the file is to be retrieved from the server), the server does not know the destination port for the FTP client. So before sending the command to retrieve the file, the client instructs its operating system to allocate a port number (above 1023) for such a transaction. This information is given to the server via the control connection. The data connection is created (using the usual TCP 3 way handshake), and the server writes the contents of the file, and closes the connection. The client reads the bytes from its socket upto the end of file (EOF) character. Also, unlike Telnet, FTP does not require the data transfer to 7 bit ASCII characters (NVT format); it actually permits a wide range of data types including binary files. The client requests the form of data transfer using the control connection. In practice, each data transfer requires a separate TCP connection. In contrast, the control connection can persist across multiple data transfers.

3.6 NETWORK MANAGEMENT

We can define **network management** as monitoring, testing, configuring, and troubleshooting network components to meet a set of requirements defined by an organization. These requirements include the smooth, efficient operation of the network that provides the predefined quality of service for users. To accomplish this task, a network management system uses hardware, software, and humans. In this chapter, first we briefly discuss the functions of a network management system. Then we concentrate on the most common management system, the Simple Network Management Protocol (SNMP).

We can say that the functions performed by a network management system can be divided into **five broad categories**: configuration management, fault management, performance management, security management, and accounting management.

3.6.1 Configuration Management

A large network is usually made up of hundreds of entities that are physically or logically connected to one another. These entities have an initial configuration when the network is set up, but can change with time. Desktop computers may be replaced by others; application software may be updated to a newer version; and users may move from one group to another. The **configuration management** system must know, at any time, the status of each entity and its relation to other entities. Configuration management can be divided into two subsystems: reconfiguration and documentation.

3.6.2 Reconfiguration

Reconfiguration, which means adjusting the network components and features, can be a daily occurrence in a large network. There are three types of reconfiguration: hardware reconfiguration, software reconfiguration, and user-account reconfiguration. Hardware reconfiguration covers all changes to the hardware. For example, a desktop computer may need to be replaced. A router may need to be moved to another part of the network. A sub network may be added or removed from the network. All these need the time and attention of network management. In a large network, there must be specialized personnel trained for quick and efficient hardware reconfiguration. Unfortunately, this type of reconfiguration cannot be automated and must be manually handled case by case.

Software reconfiguration covers all changes to the software. For example, new software may need to be installed on servers or clients. An operating system may need updating. Fortunately, most software reconfiguration can be automated. For example, updating an application on some or all clients can be electronically downloaded from the server.

User-account reconfiguration is not simply adding or deleting users on a system. You must also consider the user privileges, both as an individual and as a member of a group. For example, a user may have read and write permission with regard to some files, but only read permission with regard to other files. User-account reconfiguration can be, to some extent, automated. For example, in a college or university, at the beginning of each quarter or semester, new students are added to the system. The students are normally grouped according to the courses they take or the majors they pursue.

3.6.3 Documentation

The original network configuration and each subsequent change must be recorded meticulously. This means that there must be documentation for hardware, software, and user accounts. **Hardware documentation** normally involves two sets of documents: maps and specifications. Maps track each piece of hardware and its connection to the network. There can be one general map that shows the logical relationship between each sub-network. There can also be a second general map that shows the physical location of each sub-network. For each sub network, then, there are one or more maps that show all pieces of equipment. The maps use some kind of standardization to be easily read and understood by current and future personnel. Each piece of hardware also needs to be documented. There must be a set of specifications for each piece of hardware connected to the network. These specifications must include information such as hardware type, serial number, vendor (address and phone number), time of purchase, and warranty information. All software must also be documented. Software documentation includes information such as the software type, the version, the time installed, and the license agreement. Most operating systems have a utility that allows the documentation of user accounts and their privileges. The management must make sure that the files with this information are updated and secured. Some operating systems record access privileges in two documents one shows all files and access types for each user; the other shows the list of users that have a particular access to a file.

3.6.4 Fault Management

Complex networks today are made up of hundreds and sometimes thousands of components. Proper operation of the network depends on the proper operation of each component individually and in relation to each other. **Fault management** is the area of network management that handles this issue. An effective fault management system has two subsystems: reactive fault management and proactive fault management.

3.6.5 Reactive Fault Management

A reactive fault management system is responsible for detecting, isolating, correcting, and recording faults. It handles short-term solutions to faults. The first step taken by a reactive fault management system is to detect the exact location of the fault. A fault is defined as an abnormal condition in the system. When a fault occurs, either the system stops working properly or the system creates excessive errors. A good example of a fault is a damaged communication medium. This fault may interrupt communication or produce excessive errors.

The next step taken by a reactive fault management system is to isolate the fault. A fault, if isolated, usually affects only a few users. After isolation, the affected users are immediately notified and given an estimated time of correction. The third step is to correct the fault. This may involve replacing or repairing the faulty component(s).After the fault is corrected, it must be documented. The record should show the exact location of the fault, the possible cause, the action or actions taken to correct the fault, the cost, and time it took for each step. Documentation is extremely important for several reasons:

- The problem may recur. Documentation can help the present or future administrator or technician solve a similar problem.
- The frequency of the same kind of failure is an indication of a major problem in the system. If a fault happens frequently in one component, it should be replaced with a similar one, or the whole system should be changed to avoid the use of that type of component.
- The statistic is helpful to another part of network management, performance management.

3.6.6 Proactive Fault Management

Proactive fault management tries to prevent faults from occurring. Although this is not always possible, some types of failures can be predicted and prevented. For example, if a manufacturer specifies a lifetime for a component or a part of a component, it is a good strategy to replace it before that time. As another example, if a fault happens frequently at one particular point of a network, it is wise to carefully reconfigure the network to prevent the fault from happening again.

3.6.7 Performance Management

Performance management, which is closely related to fault management, tries to monitor and control the network to ensure that it is running as efficiently as possible. Performance management tries to quantify performance by using some measurable quantity such as capacity, traffic, throughput, or response time.

Capacity

One factor that must be monitored by a performance management system is the capacity of the network. Every network has a limited capacity, and the performance management system must ensure that it is not used above this capacity. For example, if a LAN is designed for 100 stations at an average data rate of2 Mbps, it will not operate properly if 200 stations are connected to the network. The data rate will decrease and blocking may occur.

Traffic

Traffic can be measured in two ways: internally and externally. Internal traffic is measured by the number of packets (or bytes) traveling inside the network for a given period. External traffic is measured by the exchange of packets (or bytes) outside the network. During peak hours, when the system is heavily used, blocking may occur if there is excessive traffic.

Throughput

We can measure the throughput of an individual device (such as a router) or a part of the network. Performance management monitors the throughput to make sure that it is not reduced to unacceptable levels.

Response Time

Response time is normally measured from the time a user requests a service to the time the service is granted. Other factors such as capacity and traffic can affect the response time. Performance management monitors the average response time and the peak-hour response time. Any increase in response time is a very serious condition as it is an indication that the network is working above its capacity.

3.6.8 Security Management

Security management is responsible for controlling access to the network based on the predefined policy.

3.6.9 Accounting Management

Accounting management is the control of users' access to network resources through charges. Under accounting management, individual users, departments, divisions, or even projects are charged for the services they receive from the network. Charging does not necessarily mean cash transfer; it may mean debiting the departments or divisions for budgeting purposes. Today, organizations use an accounting management system for the following reasons:

- It prevents users from monopolizing limited network resources.
- It prevents users from using the system inefficiently.
- Network managers can do short- and long-term planning based on the demand for
- Network use.

3.6.10 SNMP Protocol

Since it was developed in 1988, the Simple Network Management Protocol has become the de facto standard for internetwork management. Since it is a simple solution, requiring little code to implement, vendors can easily build SNMP agents to their products. SNMP is extensible, allowing vendors to easily add network management functions to their existing products. SNMP also separates the management architecture from the architecture of the hardware devices, which broadens the base of multivendor support. Perhaps most important, unlike other so-called standards, SNMP is not a mere paper specification, but an implementation that is widely available today. To know more about SNMP you may refer to fifth semester course BCS-052 Network programming and Administration.

SNMP is based on the manager/agent model consisting of a manager, an agent, a database of management information, managed objects and the network protocol. The manager provides the interface between the human network manager and the management system. The agent provides the interface between the manager and the physical device(s) being managed (see the illustration above).

A typical agent usually

- Implements full SNMP protocol.
- Stores and retrieves management data as defined by the Management Information Base.
- Can asynchronously signal an event to the manager
- Can be a proxy for some non-SNMP manageable network node.

A typical manager usually

- Implemented as a Network Management Station (the NMS)
- Implements full SNMP Protocol Able to
- Query agents
- Get responses from agents
- Set's variables in agents
- Acknowledge's asynchronous events from agents.

☞ Check Your Progress 2

1. Select the right choice.

   a) FTP does not use

   (A) Two transfer mode.
   (B) Control connection to remote computer before file can be transferred.
   (C) User Datagram Protocol.
   (D) Authorization of a user through login and password verification.

   ……

   ……

   b) Protocol used to monitor and control network devices operates at:

   (A) Application layer (B) Transport layer
   (C) Network layer (D) Data Link layer

   ……

   ……

2. Discuss TCP connections in FTP.

   ……

   ……

3. What is configuration management in computer network management?

   ……

   ……

3.7 WORD WIDE WEB AND CLIENT SERVER APPLICATIONS

The World Wide Web is an architectural framework for accessing linked documents spread out over millions of machines all over the Internet. In the last decade on so it went from being a way to distribute high-energy physics data to the application that millions of people think of as being "The Internet." Its enormous popularity stems from the fact that it has a colorful graphical interface that is easy for beginners to use, and it provides an enormous wealth of information on almost every conceivable subject.

The Web (also known as **WWW**) began in 1989 at CERN, the European center for nuclear research. CERN has several accelerators at which large teams of scientists from the participating European countries carry out research in particle physics. These teams often have members from half a dozen or more countries. Most experiments are highly complex and require years of advance planning and equipment construction. The Web grew out of the need to have these large teams of internationally dispersed researchers collaborate using a constantly changing collection of reports, blueprints, drawings, photos, and other documents.

The initial proposal for a web of linked documents came from CERN physicist Tim Berners-Lee in March 1989. The first (text-based) prototype was operational 18 months later. In December 1991, a public demonstration was given at the Hypertext '91 conference in San Antonio, Texas.

This demonstration and its attendant publicity caught the attention of other researchers, which led Marc Andreessen at the University of Illinois to start developing the first graphical browser, Mosaic. It was released in February 1993. Mosaic was so popular that a year later, Andreessen left to form a company, Netscape Communications Corp., whose goal was to develop clients, servers, and other Web software. When Netscape went public in 1995, investors, apparently thinking this was the next Microsoft, paid $1.5 billion for the stock. This record was all the more surprising because the company had only one product, was operating deeply in the red, and had announced in its prospectus that it did not expect to make a profit for the foreseeable future. For the next three years, Netscape Navigator and Microsoft's Internet Explorer engaged in a "browser war," each one trying frantically to add more features (and thus more bugs) than the other one. In 1998, America Online bought Netscape Communications Corp. for $4.2 billion, thus ending Netscape's brief life as an independent company.

In 1994, CERN and M.I.T. signed an agreement setting up the **World Wide Web Consortium** (sometimes abbreviated as **W3C**), an organization devoted to further developing the Web, standardizing protocols, and encouraging interoperability between sites. Berners-Lee became the director. Since then, several hundred universities and companies have joined the consortium. Although there are now more books about the Web than you can shake a stick at, the best place to get up-to-date information about the Web is (naturally) on the Web itself.

3.7.1 Architectural Overview (WWW)

From the users' point of view, the Web consists of a vast, worldwide collection of documents or **Web pages**, often just called **pages** for short. Each page may contain links to other pages anywhere in the world. Users can follow a link by clicking on it, which then takes them to the page pointed to. This process can be repeated indefinitely. The idea of having one page point to another, now called **hypertext**, was

invented by a visionary M.I.T. professor of electrical engineering, Vannevar Bush, in 1945, long before the Internet was invented.
Pages are viewed with a program called a **browser**, of which Internet Explorer and Netscape Navigator are two popular ones. The browser fetches the page requested, interprets the text and formatting commands on it, and displays the page, properly formatted, on the screen. Web pages, starts with a title, contain some information, and ends with the e-mail address of the page's maintainer. Strings of text that are links to other pages, called **hyperlinks**, are often highlighted, by underlining, displaying them in a special color, or both. To follow a link, the user places the mouse cursor on the highlighted area, which causes the cursor to change, and clicks on it. Although non graphical browsers, such as Lynx, exist, they are not as popular as graphical browsers. Voice-based browsers are also being developed.

The basic model of how the Web works is shown in Figure 4, Here the browser is displaying a Web page on the client machine. When the user clicks on a line of text that is linked to a page on the *abcd.com* server, the browser follows the hyperlink by sending a message to the *abcd.com* server asking it for the page. When the page arrives, it is displayed. If this page contains a hyperlink to a page on the *xyz.com* server that is clicked on, the browser then sends a request to that machine for the page, and so on indefinitely.

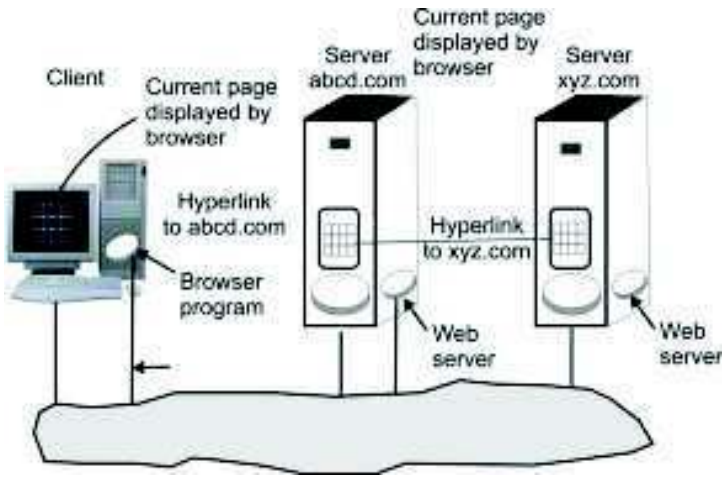


Figure 4: The parts of the Web model.

The Client Side

In essence, a browser is a program that can display a Web page and catch mouse clicks to items on the displayed page. When an item is selected, the browser follows the hyperlink and fetches the page selected. Therefore, the embedded hyperlink needs a way to name any other page on the Web. Pages are named using **URLs** (**Uniform Resource Locators**). A typical URL is http://www.abcd.com/products.html

It is sufficient to know that a URL has three parts: the name of the protocol (*http*), the DNS name of the machine where the page is located (*www.abcd.com*), and (usually) the name of the file containing the page (*products.html*).

When a user clicks on a hyperlink, the browser carries out a series of steps in order to fetch the page pointed to. Suppose that a user is browsing the Web and finds a link on Internet telephony that point to ITU's home page, which is *http://www.itu.org/home/index.html*. Let us trace the steps that occur when this link is selected.

1. The browser determines the URL (by seeing what was selected).

2. The browser asks DNS for the IP address of www.itu.org.
3. DNS replies with 156.106.192.32.
4. The browser makes a TCP connection to port 80 on 156.106.192.32.
5. It then sends over a request asking for file /home/index.html.
6. The www.itu.org server sends the file /home/index.html.
7. The browser displays all the text in /home/index.html.
8. The browser fetches and displays all images in this file.

Many browsers display which step they are currently executing in a status line at the bottom of the screen. In this way, when the performance is poor, the user can see if it is due to DNS not responding, the server not responding, or simply network congestion during page transmission.

Although a browser is basically an HTML interpreter, most browsers have numerous buttons and features to make it easier to navigate the Web. Most have a button for going back to the previous page, a button for going forward to the next page (only operative after the user has gone back from it), and a button for going straight to the user's own start page. Most browsers have a button or menu item to set a bookmark on a given page and another one to display the list of bookmarks, making it possible to revisit any of them with only a few mouse clicks. Pages can also be saved to disk or printed. Numerous options are generally available for controlling the screen layout and setting various user preferences.

In addition to having ordinary text (not underlined) and hypertext (underlined), Web pages can also contain icons, line drawings, maps, and photographs. Each of these can (optionally) be linked to another page. Clicking on one of these elements causes the browser to fetch the linked page and display it on the screen, the same as clicking on text. With images such as photos and maps, which page is fetched next may depend on what part of the image was clicked on.

Not all pages contain HTML. A page may consist of a formatted document in PDF format, an icon in GIF format, a photograph in JPEG format, a song in MP3 format, a video in MPEG format, or any one of hundreds of other file types. Since standard HTML pages may link to any of these, the browser has a problem when it encounters a page it cannot interpret.

Rather than making the browsers larger and larger by building in interpreters for a rapidly growing collection of file types, most browsers have chosen a more general solution. When a server returns a page, it also returns some additional information about the page. This information includes the MIME type of the page. Pages of type *text/html* are just displayed directly, as are pages in a few other built-in types. If the MIME type is not one of the built-in ones, the browser consults its table of MIME types to tell it how to display the page.

There are two possibilities: plug-ins and helper applications. A **plug-in** is a code module that the browser fetches from a special directory on the disk and installs as an extension to itself, as illustrated in Figure 5(a) Because plug-ins run inside the browser, they have access to the current page and can modify its appearance. After the plug-in has done its job (usually after the user has moved to a different Web page), the plug-in is removed from the browser's memory.

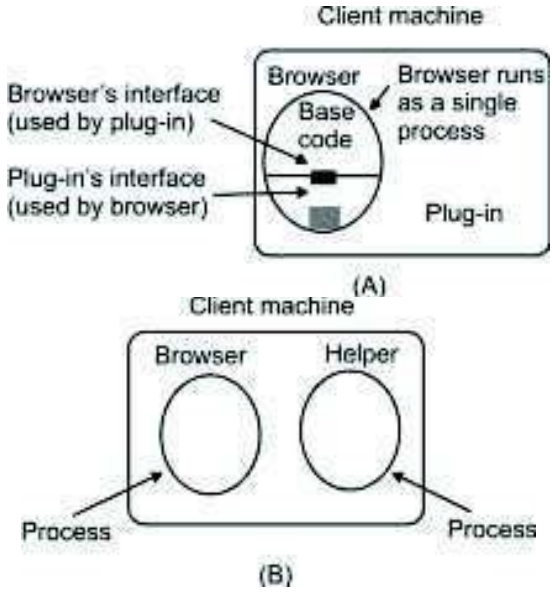


Figure 5: (a) A browser plug-in. (b) A helper application.

Each browser has a set of procedures that all plug-ins must implement so the browser can call the plug-in. For example, there is typically a procedure the browser's base code calls to supply the plug-in with data to display. This set of procedures is the plug-in's interface and is browser specific.

In addition, the browser makes a set of its own procedures available to the plug-in, to provide services to plug-ins. Typical procedures in the browser interface are for allocating and freeing memory, displaying a message on the browser's status line, and querying the browser about parameters.

Before a plug-in can be used, it must be installed. The usual installation procedure is for the user to go to the plug-in's Web site and download an installation file. On Windows, this is typically a self-extracting zip file with extension *.exe*. When the zip file is double clicked, a little program attached to the front of the zip file is executed. This program unzips the plug-in and copies it to the browser's plug-in directory. Then it makes the appropriate calls to register the plug-in's MIME type and to associate the plug-in with it. On UNIX, the installer is often a shell script that handles the copying and registration.

The other way to extend a browser is to use a **helper application**. This is a complete program, running as a separate process. It is illustrated in Figure 5(b). Since the helper is a separate program, it offers no interface to the browser and makes no use of browser services. Instead, it usually just accepts the name of a scratch file where the content file has been stored, opens the file, and displays the contents. Typically, helpers are large programs that exist independently of the browser, such as Adobe's Acrobat Reader for displaying PDF files or Microsoft Word. Some programs (such as Acrobat) have a plug-in that invokes the helper itself.

Many helper applications use the MIME type *application*. A considerable number of subtypes have been defined, for example, *application/pdf* for PDF files and *application/msword* for Word files. In this way, a URL can point directly to a PDF or Word file, and when the user clicks on it, Acrobat or Word is automatically started and handed the name of a scratch file containing the content to be displayed. Consequently, browsers can be configured to handle a virtually unlimited number of document types with no changes to the browser. Modern Web servers are often

configured with hundreds of type/subtype combinations and new ones are often added every time a new program is installed.

Helper applications are not restricted to using the *application* MIME type. Adobe Photoshop uses *image/x-photoshop* and Real One Player is capable of handling *audio/mp3*, for example.

On Windows, when a program is installed on the computer, it registers the MIME types it wants to handle. This mechanism leads to conflict when multiple viewers are available for some subtype, such as *video/mpg*. What happens is that the last program to register overwrites existing (MIME type, helper application) associations, capturing the type for itself. As a consequence, installing a new program may change the way a browser handles existing types.

On UNIX, this registration process is generally not automatic. The user must manually update certain configuration files. This approach leads to more work but fewer surprises.

Browsers can also open local files, rather than fetching them from remote Web servers. Since local files do not have MIME types, the browser needs some way to determine which plug-in or helper to use for types other than its built-in types such as *text/html* and *image/jpeg*. To handle local files, helpers can be associated with a file extension as well as with a MIME type. With the standard configuration, opening *foo.pdf* will open it in Acrobat and opening *bar.doc* will open it in Word. Some browsers use the MIME type, the file extension, and even information taken from the file itself to guess the MIME type. In particular, Internet Explorer relies more heavily on the file extension than on the MIME type when it can.

Here, too, conflicts can arise since many programs are willing, in fact, eager, to handle, say, *.mpg*. During installation, programs intended for professionals often display checkboxes for the MIME types and extensions they are prepared to handle to allow the user to select the appropriate ones and thus not overwrite existing associations by accident. Programs aimed at the consumer market assume that the user does not have a clue what a MIME type is and simply grab everything they can without regard to what previously installed programs have done.

The ability to extend the browser with a large number of new types is convenient but can also lead to trouble. When Internet Explorer fetches a file with extension *exe*, it realizes that this file is an executable program and therefore has no helper. The obvious action is to run the program. However, this could be an enormous security hole. All a malicious Web site has to do is produce a Web page with pictures of, say, movie stars or sports heroes, all of which are linked to a virus. A single click on a picture then causes an unknown and potentially hostile executable program to be fetched and run on the user's machine. To prevent unwanted guests like this, Internet Explorer can be configured to be selective about running unknown programs automatically, but not all users understand how to manage the configuration.

The Server Side

Now let us take a look at the server side. As we saw above, when the user types in a URL or clicks on a line of hypertext, the browser parses the URL and interprets the part between *http://* and the next slash as a DNS name to look up. Armed with the IP address of the server, the browser establishes a TCP connection to port 80 on that server. Then it sends over a command containing the rest of the URL, which is the name of a file on that server. The server then returns the file for the browser to display.

The steps that the server performs in its main loop are as follows:

1. Accept a TCP connection from a client (a browser).
2. Get the name of the file requested.
3. Get the file (from disk).
4. Return the file to the client.
5. Release the TCP connection.

Modern Web servers have more features, but in essence, this is what a Web server does.

A problem with this design is that every request requires making a disk access to get the file. The result is that the Web server cannot serve more requests per second than it can make disk accesses.

One obvious improvement (used by all Web servers) is to maintain a cache in memory of the *n* most recently used files. Before going to disk to get a file, the server checks the cache. If the file is there, it can be served directly from memory, thus eliminating the disk access. Although effective caching requires a large amount of main memory and some extra processing time to check the cache and manage its contents, the savings in time are nearly always worth the overhead and expense.

The next step for building a faster server is to make the server multithreaded. In one design, the server consists of a front-end module that accepts all incoming requests and *k* processing modules, as shown in Figure 6. The $k + 1$ threads all belong to the same process so the processing modules all have access to the cache within the process' address space. When a request comes in, the front end accepts it and builds a short record describing it. It then hands the record to one of the processing modules. In another possible design, the front end is eliminated and each processing module tries to acquire its own requests, but a locking protocol is then required to prevent conflicts.

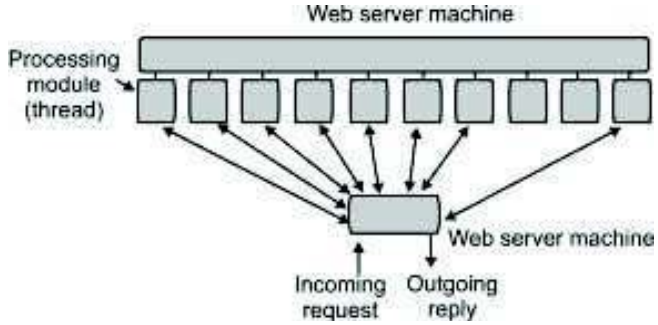


Figure 6: A multithreaded Web server with a front end and processing modules.

The processing module first checks the cache to see if the file needed is there. If so, it updates the record to include a pointer to the file in the record. If it is not there, the processing module starts a disk operation to read it into the cache (possibly discarding some other cached files to make room for it). When the file comes in from the disk, it is put in the cache and also sent back to the client.

The advantage of this scheme is that while one or more processing modules are blocked waiting for a disk operation to complete (and thus consuming no CPU time), other modules can be actively working on other requests. Of course, to get any real

improvement over the single-threaded model, it is necessary to have multiple disks, so more than one disk can be busy at the same time. With *k* processing modules and *k* disks, the throughput can be as much as *k* times higher than with a single-threaded server and one disk.

In theory, a single-threaded server and *k* disks could also gain a factor of *k*, but the code and administration are far more complicated since normal blocking READ system calls cannot be used to access the disk. With a multithreaded server, they can be used since then a READ blocks only the thread that made the call, not the entire process.

Modern Web servers do more than just accept file names and return files. In fact, the actual processing of each request can get quite complicated. For this reason, in many servers each processing module performs a series of steps. The front end passes each incoming request to the first available module, which then carries it out using some subset of the following steps, depending on which ones are needed for that particular request.

1. Resolve the name of the Web page requested.
2. Authenticate the client.
3. Perform access control on the client.
4. Perform access control on the Web page.
5. Check the cache.
6. Fetch the requested page from disk.
7. Determine the MIME type to include in the response.
8. Take care of miscellaneous odds and ends.
9. Return the reply to the client.
10. Make an entry in the server log.

Step 1 is needed because the incoming request may not contain the actual name of the file as a literal string. For example, consider the URL *http://www.cs.vu.nl*, which has an empty file name. It has to be expanded to some default file name. Also, modern browsers can specify the user's default language (e.g., Italian or English), which makes it possible for the server to select a Web page in that language, if available. In general, name expansion is not quite so trivial as it might at first appear, due to a variety of conventions about file naming.

Step 2 consists of verifying the client's identity. This step is needed for pages that are not available to the general public. We will discuss one way of doing this later in this chapter.

Step 3 checks to see if there are restrictions on whether the request may be satisfied given the client's identity and location. Step 4 checks to see if there are any access restrictions associated with the page itself. If a certain file (e.g., *.htaccess*) is present in the directory where the desired page is located, it may restrict access to the file to particular domains, for example, only users from inside the company.
Steps 5 and 6 involve getting the page. Step 6 needs to be able to handle multiple disk reads at the same time.

Step 7 is about determining the MIME type from the file extension, first few words of the file, a configuration file, and possibly other sources.

Step 8 is for a variety of miscellaneous tasks, such as building a user profile or gathering certain statistics.

Step 9 is where the result is sent back and step 10 makes an entry in the system log for administrative purposes. Such logs can later be mined for valuable information about user behavior, for example, the order in which people access the pages.

If too many requests come in each second, the CPU will not be able to handle the processing load, no matter how many disks are used in parallel. The solution is to add more nodes (computers), possibly with replicated disks to avoid having the disks become the next bottleneck. This leads to the **server farm** model of Figure 7. A front end still accepts incoming requests but sprays them over multiple CPUs rather than multiple threads to reduce the load on each computer. The individual machines may themselves be multithreaded and pipelined as before.

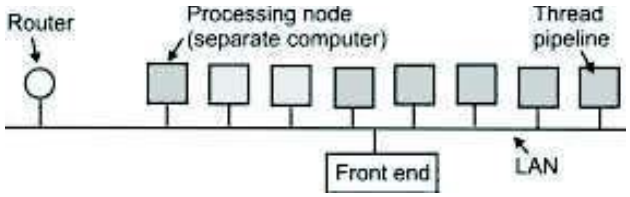


Figure 7: A server farm.

One problem with server farms is that there is no longer a shared cache because each processing node has its own memory unless an expensive shared-memory multiprocessor is used. One way to counter this performance loss is to have a front end keep track of where it sends each request and send subsequent requests for the same page to the same node. Doing this makes each node a specialist in certain pages so that cache space is not wasted by having every file in every cache. Another problem with server farms is that the client's TCP connection terminates at the front end, so the reply must go through the front end. This situation is depicted in Figure 8(a), where the incoming request (1) and outgoing reply (4) both pass through the front end.

Sometimes a trick, called **TCP handoff**, is used to get around this problem. With this trick, the TCP end point is passed to the processing node so it can reply directly to the client, shown in Figure 8(b). This handoff is done in a way that is transparent to the client.

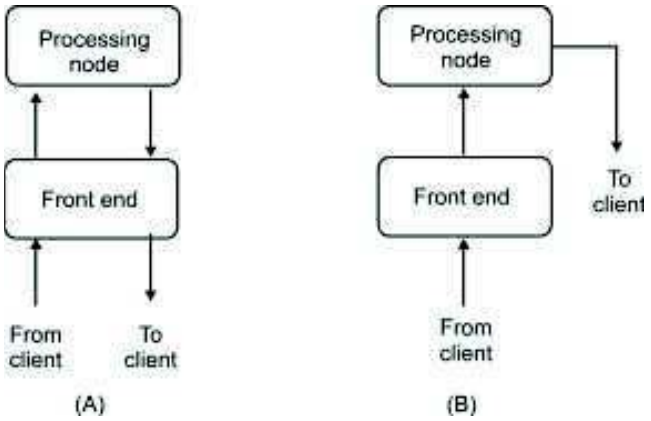


Figure 8: (a) Normal request-reply message sequence. (b) Sequence when TCP handoff is used.

HTTP

Hyper Text Transfer Protocol (HTTP) is used mainly to transfer data on World Wide Web. The commands from the client are embedded in a request message .The contents of the request message are embedded in a response message. HTTP uses the services of TCP at port 80.

HTTP is a stateless protocol since each transaction is independent of the previous transaction. The TCP connection between the client and the server is established for every page. It does not remember anything about the previous request. Keeping HTTP stateless was aimed at making the Web simple.

Sample HTTP request and response transaction is shown below:

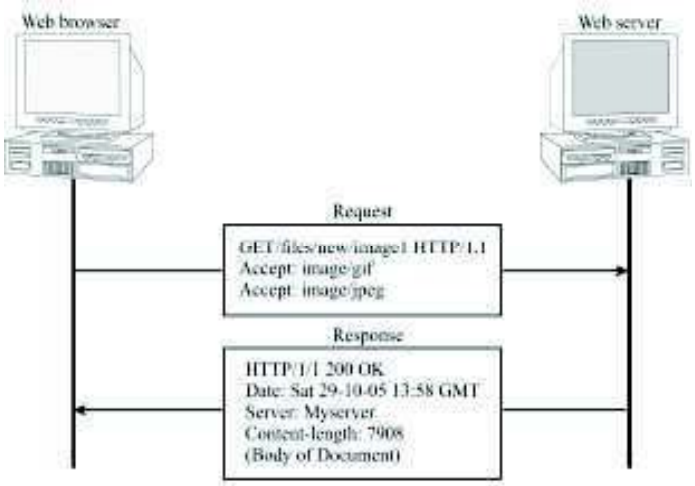


Figure 9: HTTP request and response

As shown in figure 9 above, the GET command requests the web server at www.myserver.com for an image file image1 .The HTTP/1.0 indicates that the browser uses the 1.0 version of the HTTP protocol.

The first line of response indicates that the server is also using HTTP version 1.0 .the return code of 200 indicates that the server processed the request successfully. Request message can be of following types:

Method	**Action**
GET	Requests a document from the server
HEAD	Request information about the document but not the document itself; i.e. head of the HTML page
POST	Sends some information from client to server. It appends the data to the existing document.
PUT	Sends a document from to server. It replaces the existing document.
TRACE	Echoes the incoming request
CONNECT	Reserved
OPTION	Enquire about available options

3.8 ELECTRONIC MAIL

Electronic mail, or **e-mail**, as it is known to its many fans, has been around for over two decades. Before 1990, it was mostly used in academia. During the 1990s, it became known to the public at large and grew exponentially to the point where the number of e-mails sent per day now is vastly more than the number of **snail mail** (i.e., paper) letters.

E-mail, like most other forms of communication, has its own conventions and styles. In particular, it is very informal and has a low threshold of use. People who would never dream of calling up or even writing a letter to a Very Important Person do not hesitate for a second to send a sloppily-written e-mail.

E-mail is full of jargon such as BTW (By The Way), ROTFL (Rolling On The Floor Laughing), and IMHO (In My Humble Opinion). Many people also use little ASCII symbols called **smileys** or **emoticons** in their e-mail.

The first e-mail systems simply consisted of file transfer protocols, with the convention that the first line of each message (i.e., file) contained the recipient's address. As time went on, the limitations of this approach became more obvious. Some of the complaints were as follows:

1. Sending a message to a group of people was inconvenient. Managers often need this facility to send memos to all their subordinates.

2. Messages had no internal structure, making computer processing difficult. For example, if a forwarded message was included in the body of another message, extracting the forwarded part from the received message was difficult.

3. The originator (sender) never knew if a message arrived or not.

4. If someone was planning to be away on business for several weeks and wanted all incoming e-mail to be handled by his secretary, this was not easy to arrange.

5. The user interface was poorly integrated with the transmission system requiring users first to edit a file, then leave the editor and invoke the file transfer program.

6. It was not possible to create and send messages containing a mixture of text, drawings, facsimile, and voice.

3.8.1 Architecture and Services

In this section we will provide an overview of what e-mail systems can do and how they are organized. They normally consist of two subsystems: the **user agents**, which allow people to read and send e-mail, and the **message transfer agents**, which move the messages from the source to the destination. The user agents are local programs that provide a command-based, menu-based, or graphical method for interacting with the e-mail system. The message transfer agents are typically system **daemons**, that is, processes that run in the background. Their job is to move e-mail through the system. Typically, e-mail systems support **five basic functions**.

Composition refers to the process of creating messages and answers. Although any text editor can be used for the body of the message, the system itself can provide assistance with addressing and the numerous header fields attached to each message. For example, when answering a message, the e-mail system can extract the

originator's address from the incoming e-mail and automatically insert it into the proper place in the reply.

Transfer refers to moving messages from the originator to the recipient. In large part, this requires establishing a connection to the destination or some intermediate machine, outputting the message, and releasing the connection. The e-mail system should do this automatically, without bothering the user.

Reporting has to do with telling the originator what happened to the message. Was it delivered? Was it rejected? Was it lost? Numerous applications exist in which confirmation of delivery is important and may even have legal significance ("Well, Your Honor, my e-mail system is not very reliable, so I guess the electronic subpoena just got lost somewhere").

Displaying incoming messages is needed so people can read their e-mail. Sometimes conversion is required or a special viewer must be invoked, for example, if the message is a PostScript file or digitized voice. Simple conversions and formatting are sometimes attempted as well.

Disposition is the final step and concerns what the recipient does with the message after receiving it. Possibilities include throwing it away before reading, throwing it away after reading, saving it, and so on. It should also be possible to retrieve and reread saved messages, forward them, or process them in other ways.

In addition to these basic services, some e-mail systems, especially internal corporate ones, provide a variety of advanced features. Let us just briefly mention a few of these. When people move or when they are away for some period of time, they may want their e-mail forwarded, so the system should be able to do this automatically. Most systems allow users to create **mailboxes** to store incoming e-mail. Commands are needed to create and destroy mailboxes, inspect the contents of mailboxes, insert and delete messages from mailboxes, and so on.

Corporate managers often need to send a message to each of their subordinates, customers, or suppliers. This gives rise to the idea of a **mailing list**, which is a list of e-mail addresses. When a message is sent to the mailing list, identical copies are delivered to everyone on the list.

3.8.2 The User Agent

E-mail systems have two basic parts, as we have seen: **the user agents** and **the message transfer agents**. In this section, we will look at the user agents. A user agent is normally a program (sometimes called a mail reader) that accepts a variety of commands for composing, receiving, and replying to messages, as well as for manipulating mailboxes. Some user agents have a fancy menu- or icon-driven interface that requires a mouse, whereas others expect 1-character commands from the keyboard. Functionally, these are the same. Some systems are menu-or icon-driven but also have keyboard shortcuts.

Sending E-mail

To send an e-mail message, a user must provide the message, the destination address, and possibly some other parameters.

Reading E-mail

Typically, when a user agent is started up, it looks at the user's mailbox for incoming e-mail before displaying anything on the screen. Then it may announce the number of

messages in the mailbox or display a one-line summary of each one and wait for a command.

SMTP server (Simple Mail Transfer Protocol)

For email messaging, every domain maintains an email server. The server runs protocols software that enable email communication. There are two main emails protocols: POP and SMTP. Because both the email protocol software programs run on server computers, the server computers are themselves called POP server and SMTP server. A single server can host both the POP and SMTP server programs.
SMTP is the Internet protocol used to transfer electronic mail between computers. The second generation of SMTP is called ESMTP (for Extended SMTP), but the differences are not important for this introduction.

It actually transfers the email message from the SMTP server of the sender to the SMTP server of the recipient. Its main job is to carry the message between the sender and the receiver. It uses TCP/IP underneath. That is, it runs on top of TCP/IP.
At the sender's site, an SMTP server takes the message sent by a user's computer. The SMTP server at the sender's end then transfers the message to the SMTP server of the recipient.

The SMTP server at the recipient's end takes the message and stores it in the appropriate user's mailbox.

☞ Check Your Progress 3

1. Select the right choice

   a) Internet's initial development was supported by:
   (A) ARPANET
   (B) Bill Rogers
   (C) Bill Gates
   (D) Microsoft

   ……

   ……

   b) What are the uses of the Internet?

   (A) Communication
   (B) Information Retrieval
   (C) Presentation of Information
   (D) All of the above

   ……

   ……

   c) net domain is used for

   (A) Educational institution
   (B) Internet infrastructure and service providers
   (C) International organizations
   (D) None of the above

   ……

   ……

2. Differentiate between http and ftp.

……

……

3.9 SUMMARY

This completes our discussion on the application layer. Include identifying communication partners, determining resource availability, and synchronizing communication. When identifying communication partners, the application layer determines the identity and availability of communication partners for an application with data to transmit. When determining resource availability, the application layer must decide whether sufficient network or the requested communication exists. The unit very well defines the concept of DNS and various internet and communication related issues like www, emailing system, FTP, Telnet etc. To know further about different application layer protocol students may refer to the course material of BCS-052: Network Programming and Administration

3.10 REFERENCES/FURTHER READING

1. Introduction to Data Communication & Networking, 3rd Edition, Behrouz Forouzan, Tata McGraw Hill.
2. Computer Networks, A. S. Tanenbaum 4th Edition, Practice Hall of India, New Delhi. 2003.
3. Douglas E. Comer, Internetworking with TCP/IP Vol.1: Principles, Protocols, and Architecture (4th Edition).
4. James F. Kurose, Computer Networking: A Top-Down Approach Featuring the Internet (3rd Edition).
5. Larry L. Peterson, Computer Networks: A Systems Approach, 3rd Edition (The Morgan Kaufmann Series in Networking).
6. www. wikipedia.org
7. W. Richard Stevens, The Protocols (TCP/IP Illustrated, Volume 1).
8. William Stallings, Data and Computer Communications, Seventh Edition.

3.11 SOLUTIONS / ANSWERS

☞ **Check Your Progress 1**

1. a) D
   b) B
   c) C
2. Managing a large and constantly changing set of names is a nontrivial problem. In the postal system, name management is done by requiring letters to specify (implicitly or explicitly) the country, state or province, city, and street address of the addressee. Conceptually, the Internet is divided into over 200 top-level **domains**, where each domain covers many hosts. Each domain is partitioned into sub domains, and these are further partitioned, and so on. All these domains can be represented by a tree. The top-level domains come in two flavors: **generic** and **countries**.

3. A single name server could contain the entire DNS database and respond to all queries about it. In practice, this server would be so overloaded as to be useless. Furthermore, if it ever went down, the entire Internet would be crippled. To avoid the problems associated with having only a single source of information, the DNS name space is divided into non-overlapping **zones**. A zone will have one primary name server, which gets its information from a file on its disk, and one or more secondary name servers, which get their information from the primary name server. To improve reliability, some servers for a zone can be located outside the zone.

☞ Check Your Progress 2

1. a) (C)

   b) (A)

2. Two TCP connections in FTP are:

   i) **The control connection** is established in the normal client-server fashion. In this case, the server does a *passive open* (is listening) on port 21 for FTP, and waits for the client connection. The client does an *active open* (the2nd handshake in a TCP connection) to establish the control connection. The client uses an ephemeral port number (above 1023) for the control connection. This control connection stays up for the entire time that the client communicates with this server. This connection is used for commands from the client to the server and for the server's replies. The IP type of service for the control connection should be to minimize delay in passing these commands over the TCP connection.

   ii) **A data connection** is created each time a file is transferred between the client and the server. The IP type of service for the data connection should be to maximize throughput since this connection is file transfer, and we want to send this entire file over a high bandwidth line. The specification of FTP includes more than 30 different commands, which are transmitted over the control connection in NVT ASCII format. The commands are not case-sensitive and may have arguments; each command ends with a two character sequence of a carriage return (CR) followed by a line feed (LF). It must be emphasized here that these commands are different from the commands typed by the user at the interface provided by the client. Transferring a single file for instance requires only a single user-level command (e.g., *put* or *get*), but this single command triggers the client to send a set of FTP commands to the server. The FTP server responds to each command with a three-digit reply code (for the FTP client) and an optional text message (for the user).

3. **Configuration management** system must know, at any time, the status of each entity and its relation to other entities. Configuration management can be divided into two subsystems: reconfiguration and documentation.

   Reconfiguration, which means adjusting the network components and features, can be a daily occurrence in a large network. There are three types of reconfiguration: hardware reconfiguration, software reconfiguration, and user-account reconfiguration. Hardware reconfiguration covers all changes to the hardware. For example, a desktop computer may need to be replaced. A router may need to be moved to another part of the network. A sub network may be added or removed from the network.

   The original network configuration and each subsequent change must be recorded meticulously. This means that there must be documentation for hardware, software, and user accounts. **Hardware documentation** normally

involves two sets of documents: maps and specifications. Maps track each piece of hardware and its connection to the network. There can be one general map that shows the logical relationship between each sub-network. There can also be a second general map that shows the physical location of each sub-network. For each sub network, then, there are one or more maps that show all pieces of equipment. The maps use some kind of standardization to be easily read and understood by current and future personnel. Each piece of hardware also needs to be documented. There must be a set of specifications for each piece of hardware connected to the network

☞ Check Your Progress 3

1. a) (A)
   b) (D)
   c) (C)

2. FTP and HTTP were developed to make Internet transmission better. FTP is used to exchange files between computer accounts, to transfer files between an account and a desktop computer (upload), or to access software archives on the Internet. It's also commonly used to download programs and other files to your computer from other servers. It transfers files in two different formats ASCII for text files and Binary format for binary files. This allows a user to perform basic file and directory management operations such as deleting, copying, or renaming. HTTP is used primarily in today's society as a set of rules for exchanging files (text, graphic images, sound, video, and other multimedia files) on the World Wide Web. It also provides access to other protocols like FTP, SMTP, NNTP, WAIS, Gopher, Telnet, and TN3270. Essential concepts that are part of HTTP include (as its name implies) the idea that files can contain references to other files whose selection will elicit additional transfer requests. Any web server machine contains, in addition to the HTML and other files it can serve, an HTTP daemon, a program that is designed to wait for HTTP requests and handle them when they arrive.

UNIT 4 NETWORK APPLICATIONS

Structure **Page Nos.**

4.0 Introduction 76
4.1 Objectives 76
4.2 Internet Applications 76
4.2.1 Email
4.2.2 Chatting
4.3 Social Networking 79
4.3.1 Blogs
4.3.2 Online multiplayer gaming
4.3.3 Facebook
4.3.4 Emerging Trends
4.3.5 Characteristics of social Networking
4.4 Railway Reservation System 85
4.5 Information Sharing 91
4.6 Electronic Governance 92
4.7 Online Processing and Collaborations 95
4.8 Mobile Applications 98
4.9 Summary 100
4.10 References/Further Readings 100
4.11 Solutions/Answers 101

4.0 INTRODUCTION

In this unit we are concentrating on the kind of applications that are used on the **Internet**. It is the part of network protocol (in the sense that they exchange messages with their peers on other machines) and part of traditional application program (in the sense that they interact with the windowing system, the file system, and ultimately, the user). It includes some of the most popular network applications available today Like: The World Wide Web and Email etc. We have also discussed some of the real-time applications like social networking, chatting, Railway Reservation system, Mobile Applications etc.

4.1 OBJECTIVES

After going through this unit you will be able to:

- define the logical structure of the Internet Applications;
- define the structure and working of network applications;
- define the concept of information sharing;
- discuss the basic features of E-Governance; and
- define the various components of mobile Applications;

4.2 INTERNET APPLICATIONS

The Internet is a global system of interconnected computer networks that uses the standard Internet protocol suite (often called TCP/IP, although not all applications use TCP) to serve billions of users worldwide. It is a network of networks that IS MODE of millions of private, public, academic, business, and government networks, of local to global scope, that are linked by a broad array of electronic, wireless and optical networking technologies. The Internet carries an extensive range of information

resources and services, such as the inter-linked hypertext documents of the World Wide Web (WWW) and the infrastructure to support email.

Most traditional communications media including telephone, music, film, and television are reshaped or redefined by the Internet, giving birth to new services such as Voice over Internet Protocol (VoIP) and Internet Protocol Television (IPTV). Newspaper, book and other print publishing are adapting to Web site technology, or are reshaped into blogging and web feeds. The Internet has enabled and accelerated new forms of human interactions through instant messaging, Internet forums, and social networking. Online shopping has boomed both for major retail outlets and small artisans and traders. Business-to-business and financial services on the Internet affect supply chains across entire industries.

The origins of the Internet reach back to research of the 1960s, commissioned by the United States government in collaboration with private commercial interests to build robust, fault-tolerant, and distributed computer networks. The funding of a new U.S. backbone by the National Science Foundation in the 1980s, as well as private funding for other commercial backbones, led to worldwide participation in the development of new networking technologies, and the merger of many networks. The commercialization of what was by the 1990s an international network resulted in its popularization and incorporation into virtually every aspect of modern human life. As of 2011, more than 2.2 billion people – nearly a third of Earth's population — use the services of the Internet.

The Internet has no centralized governance in either technological implementation or policies for access and usage; each constituent network sets its own standards. Only the overreaching definitions of the two principal name spaces in the Internet, the Internet Protocol address space and the Domain Name System, are directed by a maintainer organization, the Internet Corporation for Assigned Names and Numbers (ICANN). The technical underpinning and standardization of the core protocols (IPv4 and IPv6) is an activity of the Internet Engineering Task Force (IETF), a non-profit organization of loosely affiliated international participants that anyone may associate with by contributing technical expertise.

Internet is a short form of the technical term internetwork, the result of interconnecting computer networks with special gateways or routers. The Internet is also often referred to as the Net.

The term the Internet, when referring to the entire global system of IP networks, has been treated as a proper noun and written with an initial capital letter. In the media and popular culture, a trend has also developed to regard it as a generic term or common noun and thus write it as "the internet", without capitalization. Some guides specify that the word should be capitalized as a noun but not capitalized as an adjective.

The terms Internet and World Wide Web are often used in everyday speech without much distinction. However, the Internet and the World Wide Web are not one and the same. The Internet establishes a global data communications system between computers. In contrast, the Web is one of the services communicated via the Internet. It is a collection of interconnected documents (web pages) and other resources, linked by hyperlinks and URLs. In addition to the Web, the Internet also powers a multitude of other services, including (among others) email, file transfer, newsgroups, and online games. Web services can exist apart from the internet, such as on a private intranet.

The Internet allows greater flexibility in working hours and location, especially with the spread of unmetered high-speed connections. The Internet can be accessed almost anywhere by numerous means, including through mobile Internet devices. Mobile phones, data cards, handheld game consoles and cellular routers allow users to connect to the Internet wirelessly. Within the limitations imposed by small screens and other limited facilities of such pocket-sized devices, the services of the Internet, including email and the web, may be available. Service providers may restrict the services offered and mobile data charges may be significantly higher than other access methods.

Educational material at all levels from pre-school to post-doctoral is available from websites. Examples range from CBeebies, through school and high-school revision guides, virtual universities, to access to top-end scholarly literature through the likes of Google Scholar. For distance education, help with homework and other assignments, self-guided learning, whiling away spare time, or just looking up more detail on an interesting fact, it has never been easier for people to access educational information at any level from anywhere. The Internet in general and the World Wide Web in particular are important enablers of both formal and informal education.

The low cost and nearly instantaneous sharing of ideas, knowledge, and skills has made collaborative work dramatically easier, with the help of collaborative software. Not only can a group cheaply communicate and share ideas but the wide reach of the Internet allows such groups more easily to form. An example of this is the free software movement, which has produced, among other things, Linux, Mozilla Firefox, and OpenOffice.org. Internet chat, whether in the form of an IRC chat room or channel, via an instant messaging system, or a social networking website, allows colleagues to stay in touch in a very convenient way when working at their computers during the day. Messages can be exchanged even more quickly and conveniently than via email. These systems may allow files to be exchanged, drawings and images to be shared, or voice and video contact between team members.

4.2.1 Email

Electronic mail is one of the most popular tools made available through the Internet. It is an efficient and effective means of network communication. You can call it as an electronic postal system. One of the most valuable features of communicating via email is that it is asynchronous, meaning the recipient need not be at a computer to receive the message you send. The message will be stored and available to be read when the recipient is ready to read it. In order to send and receive email, you must have access to an Email account.

4.2.2 Chatting

Chatting may refer to any kind of communication over the Internet that offers a real time direct transmission of text-based messages from sender to receiver, hence the delay for visual access to the sent message shall not hamper the flow of communications in any of the directions. Online chat may address point-to-point communications as well as multicast communications from one sender to many receivers and voice and video chat or may be a feature of a Web conferencing service.

Online chat in a lesser stringent definition may be primarily any direct text-based or video-based (webcams), one-on-one chat or one-to-many group chat (formally also known asynchronous conferencing), using tools such as instant messengers, Internet Relay Chat, talkers and possibly MUDs. The expression online chat comes from the word chat which means "informal conversation". Online chat includes web-based applications that allow communication - often directly addressed, but anonymous - between users in a multi-user environment. Web conferencing is a more specific

online service, which is often sold as a service, hosted on a web server controlled by the vendor.

"Real-time communication between two users via computer. Once a chat has been initiated, either user can enter text by typing on the keyboard and the entered text will appear on the other user's monitor. Most networks and online services offer a chat feature."

Issues with Chatting

There has been much criticism about what online chatting has done in today's society. Many people are accusing it of replacing proper English with short hand with an almost completely new hybrid language.

Writing is changing as it takes on some of the functions and features of speech. Internet chat-rooms and rapid real-time conferencing allow users to interact with whoever happens to coexist in cyberspace. These virtual interactions involve us in `talking' more freely and more widely than ever before (Merchant, 2001). With chat-rooms replacing many face-to-face conversations it is necessary to be able to have quick conversation as if the person were present; so many people learn to type as quickly as they would normally speak. Critics are wary that this casual form of speech is being used so much that it will slowly take over common grammar; however, such a change has yet to be seen. With the increasing population of online chat-rooms there has been a massive growth of new words created or slang words, many of them documented on the website Urban Dictionary online chatting can be defined as:

"as new electronic modes of communication provoke similar anxieties amongst critics who express concern that young people are at risk, endangered by a rising tide of information over which the traditional controls of print media and the guardians of knowledge have no control on it".

4.3 SOCIAL NETWORKING

Many people use the World Wide Web to access news, weather and sports reports, to plan and book vacations and to find out more about their interests. People use chat, messaging and email to make and stay in touch with friends worldwide, sometimes in the same way as some previously had pen pals. The Internet has seen a growing number of Web desktops, where users can access their files and settings via the Internet.

Social networking websites such as Facebook, Twitter, and MySpace have created new ways to socialize and interact, some of these social networking groups are depicted in the Figure 1. Users of these sites are able to add a wide variety of information to pages, to pursue common interests, and to connect with others. It is also possible to find existing acquaintances, to allow communication among existing groups of people. Sites like LinkedIn foster commercial and business connections. YouTube and Flickr specialize in users' videos and photographs.



Figure 1: Some of the Social networking groups

The Internet has been a major outlet for leisure activity since its inception, with entertaining social experiments such as MUDs and MOOs being conducted on university servers, and humor-related Usenet groups receiving much traffic. Today, many Internet forums have sections devoted to games and funny videos; short cartoons in the form of Flash movies are also popular. Over 6 million people use blogs or message boards as a means of communication and for the sharing of ideas. The pornography and gambling industries have taken advantage of the World Wide Web, and often provide a significant source of advertising revenue for other websites. Although many governments have attempted to restrict both industries' use of the Internet, in general this has failed to stop their widespread popularity.

4.3.1 Blogs

Blog is a website where entries are written as commentary or news on a particular subject such as food, politics, or local news; some function as more personal online diaries. A typical blog combines text, images, and links to other blogs, web pages, and other media related to its topic. The ability for readers to leave comments in an interactive format is an important part of many blogs. Most blogs are primarily textual, although some focus on art, photographs, videos and music and are part of a wider network of social media. One of the Blogging service named Blogger is depicted in the Figure 2.



Figure 2: One of the Blogging service named Blogger

Blogging is gaining popularity in education, as it removes the technical barriers of writing and publishing online, which encourages students to keep a record of their ideas and thinking over time. Blogging also facilitate readers to give critical feedback on any topic, readers can add comments, where readers can be teachers, other students or a wider viewers. Teachers should investigate the potential of blogs, media-sharing services and other social software, which can be used create new learning opportunities. Students can also use the blogs as blog can provide a personal space online, to ask questions, comment on other questions, publish work, and link to other web sources. However a blog needn't be restricted to a single author, it can merge different kinds of ideas, including fellow students, teachers, and subject specialists. An example: http://edu.blogs.com/.

Blogging platform alternatives are following:

- WordPress.org is not the only blogging platform out there. There are several popular alternatives. First of all there is WordPress.com. Learn here about WordPress.org vs. WordPress.com.
- Google's Blogspot Blogger is another popular option. See here a bit on the difference between WordPress or Blogger.
- Tumblr is another popular choice. If you're wondering which on to pick, check Tumblr vs WordPress article.
- Google offer Blogger, which is easy way to collaborate, discuss, or share your thoughts with others. In the following section we will discuss how to create a blog, and discuss the main features of Blogger.

4.3.2 Online multiplayer gaming

Another area of leisure activity on the Internet is multiplayer gaming. This form of recreation creates communities, where people of all ages and origins enjoy the fast-paced world of multiplayer games. These range from MMORPG to first-person shooters, from role-playing video games to online gambling. While online gaming has been around since the 1970s, modern modes of online gaming began with subscription services such as GameSpy and MPlayer. Non-subscribers were limited to certain types of game play or certain games. Many people use the Internet to access and download music, movies and other works for their enjoyment and relaxation. Free and fee-based services exist for all of these activities, using centralized servers and distributed peer-to-peer technologies. Some of these sources exercise more care with respect to the original artists' copyrights than others.

Internet usage has been correlated to users' loneliness. Lonely people tend to use the Internet as an outlet for their feelings and to share their stories with others, such as in the "I am lonely will anyone speak to me" thread.

Cyber-sectarianism is a new organizational form which involves: "highly dispersed small groups of practitioners that may remain largely anonymous within the larger social context and operate in relative secrecy, while still linked remotely to a larger network of believers who share a set of practices and texts, and often a common devotion to a particular leader. Overseas supporters provide funding and support; domestic practitioners distribute tracts, participate in acts of resistance, and share information on the internal situation with outsiders. Collectively, members and practitioners of such sects construct viable virtual communities of faith, exchanging personal testimonies and engaging in collective study via email, on-line chat rooms and web-based message boards."

A **social networking service** is an online service, platform, or site that focuses on facilitating the building of social networks or social relations among people who, for

example, share interests, activities, backgrounds, or real-life connections. A social network service consists of a representation of each user (often a profile), his/her social links, and a variety of additional services. Most social network services are web-based and provide means for users to interact over the Internet, such as e-mail and instant messaging. Online community services are sometimes considered as a social network service, though in a broader sense, social network service usually means an individual-centered service where as online community services are group-centered. Social networking sites allow users to share ideas, activities, events, and interests within their individual networks.

The main types of social networking services are those that contain category places (such as former school year or classmates), means to connect with friends (usually with self-description pages), and a recommendation system linked to trust. Popular methods now combine many of these, with Facebook, Twitter and Google+ widely used worldwide.

Some social networks have additional features, such as the ability to create groups that share common interests or affiliations, upload or stream live videos, and hold discussions in forums. Geosocial networking co-opts Internet mapping services to organize user participation around geographic features and their attributes.

There is a trend towards more interoperability between social networks led by technologies such as OpenID and OpenSocial. In most mobile communities, mobile phone users can now create their own profiles, make friends, participate in chat rooms, create chat rooms, hold private conversations, share photos and videos, and share blogs by using their mobile phone. Some companies provide wireless services that allow their customers to build their own mobile community and brand it; one of the most popular wireless services for social networking in North America is Facebook Mobile.

4.3.3 Facebook

At present, Facebook is one of the popular Social Networking sites used by millions of people around the world, especially young people including your learners to connect to each other; Figure 3 shows the Login/Signup page for Facebook. This site is a free and effective way of communication on-line with your learners. You can, for example, send the messages, assignments and on-line resources for your subject. They in turn can communicate with you by posting questions etc.



Figure 3: Login/Signup page for Facebook

4.3.4 Emerging Trends

As the increase in popularity of social networking is on a constant rise, new uses for the technology are constantly being observed.

At the forefront of emerging trends in social networking sites is the concept of "real-time web" and "location-based." Real-time allows users to contribute content, which is then broadcast as it is being uploaded - the concept is analogous to live radio and television broadcasts. Twitter set the trend for "real-time" services, wherein users can broadcast to the world what they are doing, or what is on their minds within a 140-character limit. Facebook followed suit with their "Live Feed" where users' activities are streamed as soon as it happens. While Twitter focuses on words, Clixtr, another real-time service, focuses on group photo sharing wherein users can update their photo streams with photos while at an event. Facebook, however, remains the largest photo sharing site - Facebook application and photo aggregator Pixable estimates that Facebook will have more 100 billion photos by mid of 2011. In April, 2012, the image-based social media network *Pinterest* had become the third largest social network in the United States.

Companies have begun to merge business technologies and solutions, such as cloud computing, with social networking concepts. Instead of connecting individuals based on social interest, companies are developing interactive communities that connect individuals based on shared business needs or experiences. Many provide specialized networking tools and applications that can be accessed via their websites, such as LinkedIn. Others companies, such as Monster.com, have been steadily developing a more "socialized" feel to their career center sites to harness some of the power of social networking sites. These more business related sites have their own nomenclature for the most part but the most common naming conventions are "Vocational Networking Sites" or "Vocational Media Networks", with the former more closely tied to individual networking relationships based on social networking principles.

Foursquare gained popularity as it allowed for users to "check-in" to places that they are frequenting at that moment. *Gowalla* is another such service that functions in much the same way that Foursquare does, leveraging the GPS in phones to create a location-based user experience. Clixtr, though in the real-time space, is also a location-based social networking site, since events created by users are automatically geotagged, and users can view events occurring nearby through the Clixtr iPhone app. Recently, Yelp announced its entrance into the location-based social networking space through check-ins with their mobile app; whether or not this becomes detrimental to Foursquare or Gowalla is yet to be seen, as it is still considered a new space in the Internet technology industry.

One popular use for this new technology is social networking between businesses. Companies have found that social networking sites such as Facebook and Twitter are great ways to build their brand image.

There are five major uses for businesses and social media: to create brand awareness, as an online reputation management tool, for recruiting, to learn about new technologies and competitors, and as a lead generation tool to intercept potential prospects. These companies are able to drive traffic to their own online sites while encouraging their consumers and clients to have discussions on how to improve or change products or services.

4.3.5 Characteristics of social Networking

Social networks and science

By sharing information and knowledge with one another, people are able to "increase both their learning and their flexibility in ways that would not be possible within a self-contained hierarchical organization." Social networking is allowing scientific groups to expand their knowledge base and share ideas, and without these new means of communicating their theories might become "isolated and irrelevant".

Social networks and education

Social networks are also being used by teachers and students as a communication tool. Because many students are already using a wide range of social networking sites, teachers have begun to familiarize themselves with this trend in order to leverage student interest in relation to curriculum content. Some of this includes creating chat-room forums and groups to extend classroom discussion to posting assignments, tests and quizzes, through to assisting with homework outside of the classroom setting. Social network services are also being used to foster teacher-parent communication. These services make it possible and more convenient for parents to ask questions and voice concerns without having to meet face-to-face with their children's teachers. The advent of social networking platforms may also be impacting the way(s) in which learners engage with technology in general. The use of online social networks by school libraries is also increasingly prevalent and they are being used to communicate with potential library users, as well as extending the services provided by individual school libraries.

Social networks and their educational uses are of interest to many researchers. “Social networking sites, like much else on the internet, represent a moving target for researchers and policy makers.” Recent trends indicate that 47% of American adults use a social network. A national survey in 2009 found that 73% of online teenagers use SNS, which is an increase from 55% three years earlier. Recent studies have shown that social network services provide opportunities within professional education, curriculum education, and learning. However, there are constraints in this area.

Learning uses within education

Educators and advocates of new digital literacy are confident that social networking encourages the development of transferable, technical, and social skills of value in formal and informal learning. In a formal learning environment, goals or objectives are determined by an outside department or agency. Tweeting, instant messaging, or blogging enhances student involvement. Students who would not normally participate in class are more apt to partake through social network services.

Networking allows participants the opportunity for just-in-time learning and higher levels of engagement. The use of SNSs allow educators to enhance the prescribed curriculum. When learning experiences are infused into a website, students utilize everyday for fun, students realize that learning can and should be a part of everyday life. It does not have to be separate and unattached. Informal learning consists of the learner setting the goals and objectives. It has been claimed that media no longer just influence our culture. They are our culture. With such a high number of users between the ages of 13-18, a number of skills are developed. Participants hone technical skills in choosing to navigate through social networking services. This includes elementary items such as sending an instant message or updating a status. The developments of new media skills are paramount in helping youth navigate the digital world with confidence. Social networking services foster learning through "Participatory Culture." A participatory culture consists of a space that allows engagement, sharing, mentoring, and an opportunity for social interaction. Participants of social network services avail of this opportunity. Informal learning, in the forms of participatory and social learning online, is an excellent tool for teachers to sneak in material and ideas that students will identify with and therefore, in a secondary manner, students will

learn skills that would normally be taught in a formal setting in the more interesting and engaging environment of social learning.

Sites like Twitter provide students with the opportunity to converse and collaborate with others in real time. Social networking services provide a virtual “space” for learners.

Social Interaction

Social networking is a way for one person to meet up with other people on the net. People use social networking sites for meeting new friends, finding old friends, or locating people who have the same problems or interests they have, called niche networking.

More and more relationships and friendships are being formed online and then carried to an offline setting. The relationships which start online are much more likely to succeed.

Being able to meet some-one as a "friend" and see what common interests you share and how you have built up your friend base and "likes" you can truly see a fuller picture of the person you are talking with. Most sites are free instead of being paid based which allows younger people with stricter budgets to enjoy some of the same features as those of adults who are more likely to be able to afford pay based sites. While not the intended or original use for these social sites, a large area of their current function has stemmed from people wanting to meet other people in person and with the extremely busy schedules of most people, it is a fast, reliable and easy way in which to do so that costs you little time and money (if any). Users do not necessarily share with others the content which is of most interest to them, but rather that which projects a good impression of themselves.

4.4 RAILWAY RESERVATION SYSTEM

Journey by rail has its own charm and glitz. And, railway reservation in India is no more a hassle. You can go by online train reservation services or any outlet for that matter. Despite the coming up of cheap fairs in domestic airlines market, a substantial number of passengers and visitors yet journey by train. However, a train travel is both safe & comfortable and cheap. Indian people like to travel by train. A journey by train takes you to unearth the otherwise unexplored sites and mysteries of Mother India. As far as the railway reservation in India is concerned, there are myriad options at ones disposal. You can go by online train reservation system or any railway reservation booking outlet scattered everywhere.

However, Indian railway ticket reservation is no more a tedious job. Just lay your hands on any railway reservation booking outlet around you and make your way to the differing journey. With the onset of online railway reservation system things got much simpler for the passengers to book railway tickets online.

Indian railway is working incessantly to endow simply the best services to the passengers in India. Anyone with a system can have rail reservation instantly with no hassle. However, there are also systems of making railway reservation enquiry from virtually any place with your computer. This is how the whole system of booking railway tickets got easier in terms of accessibility and affordability. You are no longer required to sweat and fret over train ticket reservation in India. For any inquiry or for that matter any info you require regarding railway reservations just log onto the official site of Indian railways and you will have it.

Sitting at your home in front of a computer can give you all the relevant information on booking tickets in Indian railways. Booking any train on Indian Railways computerized passenger reservation system (PRS) network from any originating station or train passing through system station to any destination is that much easy nowadays. Be it about booking tickets, reservation enquiry, internet tickets (i-tickets), electronic tickets (e-tickets) or cancellation of tickets, things are just in place for the convenience of passengers.

Indian Railways-System and Network

Indian Railways is the world's second-largest railway, with 6,853 stations, 63,028 kilometers of track, 37,840 passenger coaches and 222,147 freight cars. Annually it carries some 4.83 billion passengers and 492 million tons of freight. Of the 11 million passengers who climb aboard one of 8,520 trains each day, about 550,000 have reserved accommodations. Their journeys can start in any part of India and end in any other part, with travel times as long as 48 hours and distances up to several thousand kilometers. The challenge is to provide a reservation system that can support such a huge scale of operations — regardless of whether it's measured by kilometers, passenger numbers, routing complexity, or simply the sheer scale of India.

The main challenges in front of the Indian railways are:

- Provide a reservation system that efficiently serves more than half a million people each day
- Ensure maximum uptime so reservation/ticketing/inquiry application is available 24 x 7
- Create a Web site that can accommodate more than one million hits per day
- Traveling on High Technology Indian Railways is one of the most advanced ministries in India, with an innovative and extensive IT environment — and a leading-edge reservation system powered by HP AlphaServer systems running the HP OpenVMS™ operating system and HP Reliable Transaction Router (RTR) middleware. Consider the scope of the operation.
- Good Technology means good service in 1986, the Ministry of Railways established the Centre for Railway Information Systems (CRIS) as an umbrella for all computer activities on Indian Railways. CRIS is responsible for designing, developing and implementing all major computer systems for the Railways. With its own R&D effort, CRIS has become a frontrunner in its field. One of CRIS's key technical achievements is a sophisticated reservation and ticketing application called Country-Wide Network for Enhanced Reservation and Ticketing (CONCERT), which runs on the OpenVMS AlphaServer platform. Centre for Railway Information Systems at Indian Railways, "OpenVMS is an extremely rugged and reliable operating system. Its built-in auditing feature provides us with excellent security." The primary challenge for CRIS is to provide an efficient passenger service by ensuring maximum uptime for its reservation/ticketing and inquiry application. The Railway must prepare charts that map passengers with their seats, and must post these charts outside each coach. CONCERT software enables the preparation of skeleton charts in advance for each train for the next three journey days. Indian Railway's current CONCERT application represents a steady progression of using the latest technologies available. In the mid-1980s, Indian Railways first computerized its reserved ticketing operation on VAX systems running VMS. This was done from five regional passenger reservation centres, each of which was a stand-alone site with its own local database. During the mid- to late 1990s, CRIS introduced CONCERT, which linked the five passenger reservation centres so

that reserved tickets from any station of Indian Railways could be issued to any other station from a single window. "CONCERT from CRIS has been able to improve the services to the passenger by offering single-window service to the passengers. RTR gives the user location transparency for the distributed database system. Thus, the reservation from one station to any other station can be given from a single window covering the round trip, which means passengers only have to stand in one queue. Indian railways perform various services to the passengers by using information technology, these are as follows:

i) Passenger Reservation System Solution

ii) Unreserved Ticketing System for Railways

iii) Mobile Ticketing

iv) Web Ticketing

v) Kiosk-based Ticketing

vi) Centralized (Hybrid) Ticketing System

vii) Time Table and Scheduling System

viii) Traffic Management Systems

ix) Passenger Information Display System

As more and more people turned to the Web to find information about various services, Indian Railways decided to provide information related to passenger reservations to the public over the Internet. In 2000, CRIS designed and implemented Indian Railways' own Web site, which receives a staggering 1.2 million hits per day. The site is hosted by CRIS and runs on the OpenVMS AlphaServer platform.

In 1985, CMC Software company (a subsidiary company of TATA Consultancy Services) developed IMPRESS, the railway reservation system based online transaction processing (OLTP), for the Indian Railways, which has been successfully operating it since 1987. Since then, however, the system has undergone a major change for networking all nodes in the railway network. The current software is CONCERT, implemented by the Centre for Railway Information Systems (CRIS).

The impact of IMPRESS / CONCERT on the system's users as well as on the Indian Railways has been tremendous. The benefits include substantial savings in transportation costs and in reservation time, telescopic fare benefits for cluster journeys, reduced malpractice and, above all, a modern, efficient and convenient system.

For the Railways, there is substantial reduction in cost per ticket issued, manpower savings (a 40 per cent increase in transactions handled per day), savings in space required, less strenuous work, higher productivity and fewer errors in fare computation, concession calculations, etc.

IMPRESS is being enhanced proactively, using state-of-the-art relational database management systems on open systems. The enhanced IMPRESS is built around an RDBMS core and supports full client-server architecture. It can also work on character-based terminals (used in the reservation and charting modules) in a host-based environment.

The application has been designed as an open distribution system, so that the data and transaction volume can be segregated between multiple host sites. Networking is an inherent feature of the application.

- The IMPRESS software can support both graphic user interface (GUI) and character-based terminals, which act as front-ends installed at the booking counters to cater to passenger requests.
- This software conforms to open standards. Hence, it can be interfaced to other applications like airline reservation systems, hotel reservation systems, etc., which are also based on open standards.
- The IMPRESS software is 'parametric' in terms of data and business rules, for fare computation, refund rules, cancellation, break journey rules, etc. Here, business rules are also kept as data items in the back-end repository instead of being part of the application logic. Therefore, the system can absorb changes in business rules immediately, without having to regenerate the object code.
- The application is secured against intrusion by two-level user authentications as the topmost guard. Below it, the data is secured from external access through multiple-level privileges. A data encryption facility is available across the WAN to prevent hacking.
- Ticketing for Indian Railways

Almost 14 million of the 15 million people whom the Railways transports every day travel on unreserved tickets. Handling them has been a huge problem. The Indian Railways plans to cover 943 more stations in 2006-07, and ensure that a total of 6,000 stations have UTS as of March 31, 2009. Unreserved tickets were earlier offered only two hours before the scheduled departure of trains. This not only caused inconvenience to passengers (as they wait in long queues to purchase their tickets) but it also affected the IR adversely in terms of loss in revenues, cumbersome reporting and in poor demand analysis. While some of the trains ran overcrowded, the others went partially vacant.

The implementation of UTS eliminated the earlier lapses in ticketing and helped the IR to substantially control overcrowding. The system comprised a network of terminals wherefrom the passengers could buy unreserved tickets for any journey 30 days in advance.

The Unreserved Ticketing System allowed advance planning and rational analysis of passenger demands for unreserved coaches. It also helped the IR to effectively monitor sale of tickets on various trains and regulate the train capacities to the fluctuating demands of passengers. With an aggressive use of leading hardware, data management and network technology, IR could successfully address the needs of the passengers of unreserved trains.

A network covering 63, 140 route kms (as on March 31, 2002), the Indian Railways traverse the length and breadth of the country. Even though the railways have been divided into zones for better management and functionality, the railway reservations or the process for booking train tickets is centrally computerised. Operating 14,444 trains daily, the IRCTC or Indian Railway Catering and Tourism Corporation Limited, ensures that train schedule and train timings remain prompt. In effect, Indian Railways is the largest railway system in the world to be functioning under a single management. .

The most cost efficient mode of transport, Indian Railways enjoy preference over other public transport systems. Used extensively for passenger and freight transfer, Indian Railways proves itself the forerunner in the transport sector as the most affordable, convenient and well connected network. With thousands of railway stations across the country, superior safety standards, lower environmental hazard and relatively low train fare, Indian Railways is the first choice for transfer goods and commodities .

While the Indian Railway booking system has always been well organised, one immensely crucial step has been the launch of the IRCTC website (IRCTC online rail ticket reservation system is depicted in the following Figure 4), making all processes related to Indian Railways a breeze. Besides allowing online booking of rail ticket (s), it offers everything from railway map to railway ticket fare to timetable of train (s). A complete railway enquiry system is in place, with a click of the mouse allowing one to access / check the Indian Railways timetable, railway ticket availability and booking for any sector besides complete online train ticket reservations. Visitors to the website can use it not only for railway ticket reservation / booking but also to find out ticket availability and ticket confirmation or PNR status for any ticket on any train within the Indian Railways network



Figure 4: IRCTC online rail ticket reservation system

With various seating arrangements like AC 1 tier, AC 2 tier, AC 3 tier, AC Chair Car, Sleeper and General in most long distance trains and another range of seat availability in the metro trains within urban areas, the Indian Railways offers a plethora of choices. Travellers can take their pick and enjoy the convenience and affordability of this mode of transport.

The list of **major Indian Railway Zones** is as follows:

Central Indian Railways: This is the oldest of Indian railway zones and one of the largest of the 16 zones formed by Indian railways.

Eastern Indian Railways: The Eastern Railway (ER) zone is one of the important Indian Railway zones. With its headquarters in Kolkata, the Eastern Zone is divided further into four divisions namely Mald, Howrah, Asansol and Sealdah for better working.

Northern Indian Railways: It is one of the nine older zones of Indian Railways. New Delhi, the national capital of India serves as headquarters of this Indian Railway Zone

Southern Indian Railways: This is the first zone formed after India got liberated

from British Rule. Southern Indian Railway Zone was established on April 14, 1951 by combination of always of three states.

Western Indian Railways: This Indian railways zone is amongst the most hustling and lively rail networks of the country. The headquarters of Western Indian Railway is situated in Mumbai city.

☞ Check Your Progress 1

1. Select the right choice.
   a) Which of the following identifies a specific web page and its computer on the Web?
      A) Web site
      B) Web site address
      C) URL
      D) Domain Name

   ……

   ……

   b) What type of telecommunications hardware allows you to access the web?
      A) Browser
      B) Modem
      C) FTP protocol
      D) IRC

   ……

   ……

   c) The mail server as defined in the text uses the _____ protocol.
      A) HTTP
      B) FTP
      C) POP
      D) SMTP

   ……

   ……

2. Discuss standards of email protocol?

   ……

   ……

   ……

   ……

3. Write a note on social networking service?

   ……

   ……

   ……

   ……

4.5 INFORMATION SHARING

Information sharing describes “the exchange of data between various organizations, people and technologies”.

There are several types of information sharing:

- Information shared by individuals (such as a video shared on Facebook or YouTube).
- Information shared by organizations (such as the RSS feed of an online weather report).
- Information shared between firmware/software (Such as the IP addresses of available network nodes or the availability of disk space).

The advent of wide distributed networks, intranets, cross-platform compatibility, application porting, and standardization of IP protocols have all facilitated the huge growth in global information sharing.

When it comes to personal information however, no matter how easy it is to port the actual data, there are laws in most countries prohibiting the sharing of personal data without explicit permission being granted. In the US and Europe it is a criminal offense to share any personal data about anyone without such explicit permission.

.

There is plenty of other information sharing that does not fall under the law and information sharing is increasing as more networks and organizations connect and information becomes easier to share across the internet.

Data was formerly frequently kept in silos and often not shared among other entities due to its proprietary, non-portable format or the inability to import/export data. Even simple items such as dates were stored in a whole range of different formats making the sharing of such a simple field a potential nightmare. The same applied to a whole range of data, and even if it was compatible it was often not possible to physically transfer the data from one platform to another. .

Today these problems have all been coded out and information sharing is common between computer networks; information sharing has become especially prevalent due to social networking. These 21st century network models actively encourage the sharing of information across social networks.

Facebook has 750 million accounts, YouTube has over 400 million and the other social networking sites and applications have established between them a sharing network of over a billion people. In terms of information sharing this is a global proportion with almost 10% of the world's population sharing information across common networks regularly. .

After the terrorist attacks of September 11th, information sharing became one of the United States government's goals in developing their resources to try to avert such atrocities. It was mandated among government agencies and departments that personnel create a methodology for regularly sharing relevant information. The US needed information sharing improvements to respond to various threats more effectively. The lesson was learned that when information is hoarded instead of shared, those needing it may not be able to react in a timely manner Using information sharing intelligently has been shown to be a more effective way to manage any organization; a government or a business. .

Information sharing is crucial to many businesses, helping to promptly meet customer and client needs through customer relationship systems which share information about products and services and improve access to their customers. .

Information sharing has also allowed easy availability of credit history details which helps consumers access more services. Consumers can have access to banking, financial and credit products from across the nation and even internationally where appropriate. .

Hospitals sharing medical records (under stringent conditions) about people so that their medical personnel can make better decisions, is a good example of how organizations can share information for productive purposes rather than for social entertainment as with Facebook. .

Overall, when used intelligently, information sharing is a useful way of lowering costs, improving overall accuracy of public data and allowing organizations and individuals alike to have access to information that they might need and entertainment that they want to experience.

4.6 ELECTRONIC GOVERNANCE

It is the use of a range of modern Information and Communication Technologies such as Internet, Local Area Networks, mobiles etc. by Government to improve the effectiveness, efficiency, service delivery and to promote democracy. E-Government can transform citizen service, provide access to information to empower citizens, enable their participation in government and enhance citizen economic and social opportunities, so that they can make better lives, for themselves and for the next generation. Governments are specialized institutions that contribute to governance. Representative governments seek and receive citizen support, but they also need the active cooperation of their public servants. Governance is the outcome of politics, policies, and programs.

The primary delivery models of e-Government can be divided into:

- Government-to-Citizen or Government-to-Consumer (G2C)
- Government-to-Business (G2B)
- Government-to-Government (G2G)
- Government-to-Employees (G2E)

Within each of these interaction domains, five kinds of activities take place:

1. Informing the citizen
2. Representing the citizen
3. Encouraging the citizen to vote
4. Consulting the citizen
5. Involving the citizen

The ultimate goal of the E-Government is to be able to offer an increased portfolio of public services to citizens in an efficient and cost effective manner. E-government allows for government transparency. Government transparency is important because it allows the public to be informed about what the government is working on as well as the policies they are trying to implement. Simple tasks may be easier to perform through electronic government access. Many changes, such as marital status or address changes can be a long process and take a lot of paper work for citizens. E-government allows these tasks to be performed efficiently with more convenience to

individuals. E-government is an easy way for the public to be more involved in political campaigns. It could increase voter awareness, which could lead to an increase in citizen participation in elections. It is convenient and cost-effective for businesses, and the public benefits by getting easy access to the most current information available without having to spend time, energy and money to get it.

E-government helps simplify processes and makes access to government information more easily accessible for public sector agencies and citizens. In addition to its simplicity, e-democracy services can reduce costs. The anticipated benefits of e-government include efficiency, improved services, better accessibility of public services, and more transparency and accountability. One goal of e-government will be greater citizen participation. Through the internet, people from all over the country can interact with politicians or public servants and make their voices heard. Blogging and interactive surveys will allow politicians or public servants to see the views of the people they represent on any given issue. Chat rooms can place citizens in real-time contact with elected officials, their offices or provide them with the means to replace them by interacting directly with public servants, allowing voters to have a direct impact and influence in their government. These technologies can create a more transparent government, allowing voters to immediately see how and why their representation in the capital is voting the way they are. This helps voters better decide who to vote for in the future or how to help the public servants become more productive. A government could theoretically move more towards a true democracy with the proper application of e-government. Government transparency will give insight to the public on how decisions are made and hold elected officials or public servants accountable for their actions. The public could become a direct and prominent influence in government legislature to some degree.

Characteristics of E-Governance

E-government allows citizens to interact with computers to achieve objectives at any time and any location, and eliminates the necessity for physical travel to government agents sitting behind desks and windows. Improved accounting and record keeping can be noted through computerization, and information and forms can be easily accessed, equaling quicker processing time. A network architecture of E-governance is shown in the Figure 5 given blow. It shows the different parties and agencies of governance can be connected using different networks devices and Internet.



Figure 5: Network Architecture of E-Governance. Source: http://www.mapit.gov.in

On the administrative side, access to help find or retrieve files and linked information can now be stored in databases versus hardcopies stored in various locations. Individuals with disabilities or conditions no longer have to be mobile to be active in government and can be in the comfort of their own homes.

The primary delivery models of e-Government are classified depending on who benefits. In the development of public sector or private sector portals and platforms, a system is created that benefits all constituents. Citizens needing to renew their vehicle registration have a convenient way to accomplish it while already engaged in meeting the regulatory inspection requirement. On behalf of a government partner, business provides what has traditionally, and solely, managed by government and can use this service to generate profit or attract new customers. Government agencies are relieved of the cost and complexity of having to process the transactions.
To develop these public sector portals or platforms, governments have the choice to internally develop and manage, outsource, or sign a self-funding contract. The self-funding model creates portals that pay for themselves through convenience fees for certain e-government transactions, known as self-funding portals.

Social networking is an emerging area for e-democracy. The social networking entry point is within the citizens' environment and the engagement is on the citizens' terms. Proponents of e-government perceive government use of social networks as a medium to help government act more like the public it serves. Examples can be found at almost every state government portal through Facebook, Twitter,
and YouTube widgets. Government and its agents also have the opportunity to follow citizens to monitor satisfaction with services they receive. Through ListServs, RSS feeds, mobile messaging, micro-blogging services and blogs, government and its agencies can share information to citizens who share common interests and concerns. Government is also beginning to Twitter.

E-Government Forum

Since electronic government is new to everyone throughout the world, why couldn't an international matching system be developed so that anyone around the world can search by government function (e.g., Chief Information Officer); role (e.g., state or federal Department of Transportation); city, state, or country; type of technology used in development of the e-government system; sharing of best practices; licensing or buying existing software or platforms from the agency concerned; collaborating in the integrated development of a vertical (statewide) or horizontal (national or international) systems; or exchanging ideas. Theoretically, ALL government employees in the United States or throughout the world could be included. This could be easily achieved using modern technology and foster greater peace and understanding. Almost like an online Peace Corps.

☞ Check Your Progress 2

1. Select the right choice.
   a) Why do we store information? To refer to it:
      (A) When a decision is to be made or to confirm a fact.
      (B) If the original data is lost or corrupted.
      (C) Analyze to make some interpretation, explanation or resolution with it.
      (D) Distribute for further use
      (E) All of the above.

   ……

   ……

   b) Most of the electronic payment systems on internet use _____ to ensure confidentiality and security of the payment information.
      (A) Quantum Computing
      (B) Cryptography
      (C) Both (A) & (B)
      (D) None of the above

   ……

   ……

   c) An electronic check book device is a combination of
      (A) Hardware and a digital signature
      (B) Software and information about user
      (C) Secure hardware and software
      (D) None of the above

   ……

   ……

2. Discuss the Goal's of e-government?

   ……

   ……

   ……

   ……

3. Write a note on social networking service?

   ……

4.7 ONLINE PROCESSING AND COLLABORATIONS

The terms "**online**" and "**offline**" (also stylized as "**on-line**" and "**off-line**") have specific meanings in regard to computer technology and telecommunications. In general, "online" indicates a state of connectivity, while "offline" indicates a disconnected state. In common usage, "online" often refers to the Internet or the World Wide Web.

The concepts have however been extended from their computing and telecommunication meanings into the area of human interaction and conversation, such that even offline can be used in contrast to the common usage of online. For example, discussions taking place during a business meeting are "online", while issues that do not concern all participants of the meeting should be "taken offline" — continued outside of the meeting.

To be considered online, one of the following must apply to a device:

- Under the direct control of another device
- Under the direct control of the `system` with which it is associated
- Available for immediate use on demand by the system without human intervention
- Connected to a system, and is in operation
- Functional and ready for service

In contrast, a device that is offline meets none of these criteria (e.g., its main power source is disconnected or turned off, or it is off-power).

An online system is differs from an offline system in that templates are employed whenever the user inserts content into a web page for publication. These templates area said to be ‘on demand’ in that they are applied as per the user’s requirements.

Content is stored in a database which is a mandatory requirement, especially if the user wishes to view, amend or upgrade that content.
Template processing takes place as and when requested compared to offline systems which pre-process all of their content and apply their templates beforehand.

One advantage of an online processing is its extensibility which means greater functionality in the form of add-on’s such as image galleries, forum, wikis and blogs. This scalability is an important feature of these types of CMS.
Other advantages include:

- The user views content in real-time (published)
- The user is only shown relevant content, e.g. content which is out of date is withdrawn once it has reached it’s ‘sell by date’.
- The content can be accessed by multiple authors/editors in a variety of locations.
- Content can be maintained from any location which has internet access.

The downsides to this type of system include a slowing down of performance due to the fact that every time a user views a page, that page is retrieved from the database which slows down the processing speed exponentially.

This system requires a fast connection and an up to date browser especially in regard to content creation. Both of these are improving all the time but performance speeds are quicker on a local computer as used for an offline system.

Collaboration is working together to achieve a goal. It is a recursive process where two or more people or organizations work together to realize shared goals, (this is more than the intersection of common goals seen in co-operative ventures, but a deep, collective, determination to reach an identical objective. Most collaboration requires leadership, although the form of leadership can be social within a decentralized and egalitarian group. In particular, teams that work collaboratively can obtain greater resources, recognition and reward when facing competition for finite resources. Collaboration is also present in opposing goals exhibiting the notion of adversarial collaboration, though this is not a common case for using the word. Structured methods of collaboration encourage introspection of behavior and communication. These methods specifically aim to increase the success of teams as they engage in collaborative problem solving. Forms, rubrics, charts and graphs are useful in these situations to objectively document personal traits with the goal of improving performance in current and future projects.

Types of Collaborations:

Project management

Project Management developed from different fields of application including construction, engineering, and defense.

Learning community

A learning community is a group of people who share common emotions, values or beliefs, are actively engaged in learning together from each other, and by habituation. Such communities have become the template for a cohort-based, interdisciplinary approach to higher education. This may be based on an advanced kind of educational or 'pedagogical' design. the participants of learning community must feel some sense of loyalty and belonging to the group (membership) that drive their desire to keep working and helping others, also the things that the participant do in must affect what happened in the community, that means, an active and not just a reactive performance (influence). Besides a learning community must give the chance to the participants to meet particular needs (fulfillment) by expressing personal opinions, asking for help or specific information and share stories of events with particular issue included (emotional connections) emotional experiences.

Business

Collaboration in business can be found both inter- and intra-organization and ranges from the simplicity of a partnership and crowd funding to the complexity of a multinational corporation. Collaboration between team members allows for better communication within the organization and throughout the supply chains. It is a way of coordinating different ideas from numerous people to generate a wide variety of knowledge. Collaboration with a selected few firms as opposed to collaboration with a large number of different firms have been shown to positively impact firm performance and innovation outcomes. The recent improvement in technology has provided the world with high speed internet, wireless connection, and web-based collaboration tools like blogs, and wikis, and has as such created a "mass collaboration." People from all over the world are efficiently able to communicate and

share ideas through the internet, or even conferences, without any geographical barriers. The power of social networks it beginning to permeate into business culture where many collaborative use are being found including file sharing and knowledge transfer.

Education

Generally defined, an Educational Collaborative Partnership is ongoing involvement between schools and business/industry, unions, governments and community organiza tions. Educational Collaborative Partnerships are established by mutual agreement between two or more parties to work together on projects and activities that will enhance the quality of education for students while improving skills critical to success in the workplace.

Collaboration in Education- two or more co-equal individual voluntarily brings their knowledge and experience together by interacting toward a common goal in the best interest of students for the betterment of their education success. Students achieve team building and communication skills meeting many curricular standards. Students have the ability to practice real-world communication experiences. Students gain leadership through collaboration and empower peer to peer learning.

Technology

Due to the complexity of today's business environment, collaboration in technology encompasses a broad range of tools that enable groups of people to work together including social networking, instant messaging, team spaces, web sharing, audio conferencing, video, and telephony. Broadly defined, any technology that facilitates linking of two or more humans to work together can be considered a collaborative tool. Wikipedia, Blogs, even Twitter are collaborative tools. Many large companies are developing enterprise collaboration strategies and standardizing on a collaboration platform to allow their employees, customers and partners to intelligently connect and interact.

4.8 MOBILE APPLICATIONS

A **mobile application** (or **mobile app**) is a software application designed to run on smart phones, tablet computers and other mobile devices. They are available through application distribution platforms, which are typically operated by the owner of the mobile operating system, such as the Apple App Store, Google Play, Windows Phone Marketplace and BlackBerry App World. Some apps are free, while others have a price. Usually, they are downloaded from the platform to a target device, such as an iPhone, BlackBerry, Android phone or Windows Phone 7, but sometimes they can be downloaded to less mobile computers, such as laptops or desktops. For apps with a price, generally a percentage, 20-30%, goes to the distribution provider (such as iTunes), and the rest goes to the producer of the app.

The popularity of mobile applications has continued to rise, as their usage has become increasingly prevalent across mobile phone users.

For mobile applications, the fixed telephone system is not suitable. Mobile phones are currently in widespread use for voice and will soon be in widespread use for data. The first generation was analog, dominated by AMPS. The second generation was digital, with D-AMPS, GSM, and CDMA the major options. The third generation will be digital and based on broadband CDMA.

An alternative system for network access is the cable television system, which has gradually evolved from a community antenna to hybrid fiber coax. Potentially, it

offers very high bandwidth, but the actual bandwidth available in practice depends heavily on the number of other users currently active and what they are doing.

Mobile apps were originally offered for general productivity and information retrieval, including email, calendar, contacts, and stock market and weather information. However, public demand and the availability of developer tools drove rapid expansion into other categories, such as mobile games, factory automation, GPS and location-based services, banking, order-tracking, and ticket purchases. The explosion in number and variety of apps made discovery a challenge, which in turn led to the creation of a wide range of review, recommendation, and curation sources, including blogs, magazines, and dedicated online app-discovery services.

Mobile Apps are apps or services that can be pushed to a mobile device or downloaded and installed locally.

Classification of Mobil Apps:

- **Browser-based**: apps/services developed in a markup language
- **Native**: compiled applications (device has a runtime environment). Interactive apps such as downloadable games.
- **Hybrid**: the best of both worlds (a browser is needed for discovery)

Mobile application development is the process by which application software is developed for low-power handheld devices, such as personal digital assistants, enterprise digital assistants or mobile phones. These applications can be pre-installed on phones during manufacture, downloaded by customers from various mobile software distribution platforms, or delivered as web applications using server-side or client-side processing (e.g. JavaScript) to provide an "application-like" experience within a Web browser.

☞ Check Your Progress 3

1. Select the right choice.
   a) What do the letters GSM currently mean?
      (A) Global Special Mobile
      (B) Greater System's Mobile
      (C) Global Systems for Mobile Communications
      (D) None of the above!

   ……

   ……

   b) How is this problem solved in analogue cellular network?
      (I) Base Station continuously transmits the Mobile Identification Number (MIN) and received by the mobile phone.
      (II) Mobile phone continuously transmits the Signal Audio Tone (SAT) and received by the base station.
      (III) Overall size of the cluster is increased
      (IV) Increasing the bandwidth allocated to each user.

   ……

   ……

   c) Why does GSM use TDMA, as opposed to CDMA?

(I) When GSM was planned, CDMA was not approved as a multiple access system.
(II) TDMA is better than CDMA.
(III) CDMA is not really needed in Europe.
(IV) CDMA is too expensive to implement.
(A) I and III
(B) I only
(C) II and III
(D) I and IV

……

……

2. What is collaboration? Discuss in detail.

……

……

……

……

……

3. Discuss the advantages of Mobile applications?

……

……

……

……

……

4.9 SUMMARY

This completes our discussion on the introductory concepts of Network Applications. The internet architecture discussed in the unit, we have also discussed WWW and email formats in detail. The information given on various topics such as Information Sharing, Railway Reservation System, E-Governance, Social Networking, Online Processing etc. We have also includes the discussions based on latest trends and technology used for the internet applications (Mobile Applications) which helps the reads to keep your knowledge up to date. In addition to further readings and test their skills question answer sessions are included at the end of each sections. In the next block of this course, you will learn fundamental details for setting up a small local area network including wired and wireless setup. Next block will also cover the foundational details of network security protocols and wireless networking.

4.10 REFERENCES/FURTHER READING

1. Introduction to Data Communication & Networking, 3rd Edition, Behrouz Forouzan, Tata McGraw Hill.
2. Computer Networks, A. S. Tanenbaum 4th Edition, Practice Hall of India, New Delhi. 2003.

3. Douglas E. Comer, Internetworking with TCP/IP Vol.1: Principles, Protocols, and Architecture (4th Edition).

4. James F. Kurose, Computer Networking: A Top-Down Approach Featuring the Internet (3rd Edition).

5. Larry L. Peterson, Computer Networks: A Systems Approach, 3rd Edition (The Morgan Kaufmann Series in Networking).

6. www.wikipedia.org

7. W. Richard Stevens, The Protocols (TCP/IP Illustrated, Volume 1).

8. William Stallings, Data and Computer Communications, Seventh Edition.

4.11 SOLUTIONS/ANSWERS

☞ Check Your Progress 1

1. a) (D)
   b) (B)
   c) (C)

2. The main standards that relate to the protocols of email transmission and reception are:

   i) **S**imple Mail Transfer Protocol (SMTP) - which is used with the TCP/IP protocol suite? It has traditionally been limited to the text based electronic messages.

   ii) **M**ultipurpose Internet Mail Extension (MIME) - Which allows the transmission and reception of mail that contains various types of data, such as speech, images, and motion video? It is a newer standard than STMP and uses much of its basic protocol.

   iii) **S**/MIME (Secure MIME). RSA Data security created S/MIME which supports encrypted e-mail transfer and digitally signed electronic mail.

3. A **social networking service** is an online service, platform, or site that focuses on facilitating the building of social networks or social relations among people who, for example, share interests, activities, backgrounds, or real-life connections. A social network service consists of a representation of each user (often a profile), his/her social links, and a variety of additional services. Most social network services are web-based and provide means for users to interact over the Internet, such as e-mail and instant messaging. Online community services are sometimes considered as a social network service, though in a broader sense, social network service usually means an individual-centered service where as online community services are group-centered. Social networking sites allow users to share ideas, activities, events, and interests within their individual networks.

☞ Check Your Progress 2

1. a) (D)
   b) (B)
   c) (B)

2. **Goal's of e-government**

One goal of e-government will be greater citizen participation. Through the internet, people from all over the country can interact with politicians or public servants and make their voices heard. Blogging and interactive surveys will allow politicians or public servants to see the views of the people they represent on any given issue. Chat rooms can place citizens in real-time contact with elected officials, their offices or provide them with the means to replace them by interacting directly with public servants, allowing voters to have a direct impact and influence in their government. These technologies can create a more transparent government, allowing voters to immediately see how and why their representation in the capital is voting the way they are. This helps voters better decide who to vote for in the future or how to help the public servants become more productive. A government could theoretically move more towards a true democracy with the proper application of e-government.

3. A **social networking service** is an online service, platform, or site that focuses on facilitating the building of social networks or social relations among people who, for example, share interests, activities, backgrounds, or real-life connections. A social network service consists of a representation of each user (often a profile), his/her social links, and a variety of additional services. Most social network services are web-based and provide means for users to interact over the Internet, such as e-mail and instant messaging. Online community services are sometimes considered as a social network service, though in a broader sense, social network service usually means an individual-centered service where as online community services are group-centered. Social networking sites allow users to share ideas, activities, events, and interests within their individual networks.

☞ Check Your Progress 3

1. a) (C)
   b) (D)
   c) (A)

2. **Collaboration** is working together to achieve a goal. It is a recursive process where two or more people or organizations work together to realize shared goals, (this is more than the intersection of common goals seen in co-operative ventures, but a deep, collective, determination to reach an identical objective. Most collaboration requires leadership, although the form of leadership can be social within a decentralized and egalitarian group. In particular, teams that work collaboratively can obtain greater resources, recognition and reward when facing competition for finite resources. Collaboration is also present in opposing goals exhibiting the notion of adversarial collaboration, though this is not a common case for using the word. Structured methods of collaboration encourage introspection of behavior and communication. These methods specifically aim to increase the success of teams as they engage in collaborative problem solving. Forms, rubrics, charts and graphs are useful in these situations to objectively document personal traits with the goal of improving performance in current and future projects.

3. Mobile Apps are internet applications designed to run on smartphones and other mobile devices. mobile applications help users by connecting them to Internet services more commonly accessed on desktop or notebook computers. While opportunities abound, we have identified three advantages of using mobile apps for your business: speed, volume of information, and advertising.

UNIT 1 BUILDING A SIMPLE NETWORK

Structure **Page No.**

1.0 Introduction 5
1.1 Objectives 5
1.2 Structure Cabling 5
1.2.1 Assembling patch cable
1.3 Integrating Home Computers 14
1.3.1 How to connect two computers by using cross-over cable?
1.3.2 How to share data between two computers?
1.4 Creating a small Network 16
1.4.1 How to connect computers using hub / switch ?
1.4.2 How to create cluster of switches/hubs ?
1.4.3 How to configure a wireless network?
1.5 Case: Designing & Development of Small Networks 19
1.6 Summary 24
1.7 Solutions /Answers 24
1.8 References 25

1.0 INTRODUCTION

Computer Networks forms the basis of the present day's communication. It comprises of the technology that makes the world to work. While structuring the network, one need to have sound knowledge of both software and hardware components associated with computer networking. Hardware settings involves structuring of cables, electrical connectivity, fixing access points etc, whereas the software setting helps the network administrator to make the hardware component work properly.

In this unit you will learn about the ways and means, required to build a simple network i.e. a network that can be used for your day to day working viz. sharing of files, configuring a network device, configuring a network in wired/wireless mode and so on. *We will sum up this unit with a simple case discussed in section 1.5 of this unit, the case relates to "Designing & Development of small networks", through this case you will be able to understand the practical utility and benefit of this unit.*
In a wired computer network the structured cabling forms the backbone of the network. This unit starts with the discussion over structured cabling, which is later extended to the depth of computer networks, suitable for your level.

1.1 OBJECTIVES

After going through this unit you will be able to:

- Identify the prominent problems associated with networking;
- Propose basic network solution for the identified network problems;
- Perform basic hardware structuring, required for network layout and ;
- Perform software settings, required to make a workable network.

1.2 STRUCTURE CABLING

The term structured cabling is related the cabling and connectivity products used to integrate voice, data, video etc. over LAN(Local Area Network).The cables and connectivity products are desired to be used in a systematic way, such that the organized cabling system can be easily understood by installers, network administrators, and any other technician that deals with cabling. To maintain the

world wide code of conduct for structured cabling, standards are laid by industry viz. The EIA/TIA (Electronic Industries Association / Telecommunication Industry Association) and ISO/IEC (International Standards Organization/ International Electrotechnical Commission) have created industry standards for cabling. These standards results the standardized cabling architectures, which allows a single delivery method to be designed for support and services in the workspace.

However, to ensure the efficient and effective structured cabling design, three rules are advised to be followed :

1. **Look for a complete connectivity**: Connectivity includes all the systems that are designed to connect, route, manage, and identify cables in structured cabling systems.

2. **Plan for future growth :** The number of cables installed should also meet future requirements. Category 5e, Category 6, and fiber-optic solutions should be considered to ensure that the future needs will be met.

3. **Freedom of choice in vendors**. Even though a closed and proprietary system may be less expensive initially, this could end up being much more costly over the long term. A non-standard system from a single vendor may make it more difficult to make moves, adds, or changes at a later time.

Before applying the rules to ensure reasonably good cabling mechanism, we need to do some home work, related to the length of cables required (Number of Bundles), secondly type of cable required viz. shielded or unshielded (UTP-Unshielded Twisted Pair cable). Further, we need to choose the cable as per the distance i.e. for ~ 100m length of network coverage cat5e option of cable is fine but for ~150m length of network coverage cat6 is to be opted, after that length we need to use repeaters. Now, we need to understand where to use sheathed cable and where to use unsheathed cable. The Shielded cable is thick and more protected to physical damages, thus it is generally used in the situations where physical endurance is more required viz. dragging the cable through some pipe or so. Further, you need to understand the components involved in entire cabling process, viz. The connectors, patch cords, cable and its types. So, we start with the understanding of related components viz connectors, patch cords etc. in a sequential manner.

The Structured cabling of an Ethernet systems, leads to increase the flexibility and cost-effectiveness of transmitting voice, data, and multimedia over integrated networks. Ethernet patch cords are fast, and they are becoming a familiar part of our everyday experience. These ubiquitous cables have played a central role in the development of generic and structured cabling systems, and today are used for connecting virtually all networking components, without regard to a particular application or industry. In all of these ways, patch cords are the Ether of the Ethernet. These Ethernet patch cords are clubbed with RJ45 (RJ-Registered Jack) connectors, these are the connectors which holds 8P8C ("8 position, 8 conductor") configuration. Refer to figure-1 to map RJ45 with the 8P8C configuration



Figure 1: 8P8C connector plug commonly referred as RJ 45

In Ethernet networks, these RJ-45 plugs and jacks form a modular, gendered connector system that helps in making dynamic alterations in network components in a fast and easy way. The male plugs and female jacks are held together by a spring-loaded tab—called a hook—that keeps them securely in place while in use, but allows them to be easily unplugged when changes are made to a network system or work area.

The patch cords used in most Ethernet systems are constructed using UTP(Unshielded Twisted-Pair) cable. UTP cable consists of eight insulated copper-core conductors grouped into four pairs, with each pair twisted together along the cable's length. The conductor pairs and individual conductors in UTP cables are represented by a color code that assigns a primary color—blue, orange, green, or brown—to each of the 4 twisted pairs. The insulation of a conductor within a pair is either a solid primary color, or white striped with that primary color. In this way, all conductors are identified as members of a specific twisted pair, and as individual members within that pair. The conductor pairs are numbered 1 to 4 as shown in Figure -2 below, where Pair 1 corresponding to the blue pair, Pair 2 to the orange pair, Pair 3 to the green pair, and Pair 4 to the brown pair. The individual conductors in UTP cables can be solid copper-core wires with a well-defined thickness, or bundles of fine copper wire strands. Even though the solid-conductor cables are less expensive and easier to terminate, patch cords are almost always made from stranded cables. This is because the stranding of the conductors increases the cable's flexibility and durability.



Figure 2: UTP Cable Cross Section



Figure 3: CAT-5E UTP Cable

You might be thinking , what's the use of twisting the cable, why not we use the straight strands of the cable. To answer your question you need to understand a lot of Physics associated with it, but in short, The twisted conductor pairs in UTP cables form a balanced circuit. This is because the voltages of each member in a given pair has the same amplitude (the same voltage magnitude), but their voltages are opposite in phase (one voltage is positive, and the other is negative). The uniform twisting of each of these balanced pairs reduces electromagnetic interference (EMI) and radio frequency interference (RFI) originating from other conducting pairs inside the cable, or from equipment in the cable's environment. The conductor pairs inside a twisted-pair cable influence one another through a type of EMI called crosstalk. Crosstalk occurs when the electromagnetic field generated by one pair is large enough (the pair's signal is strong enough) to cross over to the location of a neighboring pair.

You are required to go through the following key points given in the form of notes, below. These key points will let you to understand various aspects related to the various questions which might be boggling in your mind like "How the number of turns in the UTP, relates to its performance ? " Or " What is the relevance of shielding the Cable?" Or " What are the various IEEE and EIA/TIA cabling Standards, how they differ ?" Or " When to use which type of cable?" Or " What is the difference between CAT 5/CAT 5e/Cat6 cables, when to use which cable?". in the discussion below we try to answer all these questions.

NOTE:

1. **How the number of turns in the UTP, relates to its performance ?**

   The greater the number of conductor twists, the better a cable's immunity to EMI and RFI. This immunity gets even better when the number of twists per unit length (the twist rate) is varied among the four pairs. For example, manufacturers of higher-grade cables employ variations in the twist rates of individual conductor pairs, using a different twist rate for each of the four pairs in order to minimize the crosstalk between them.

2. **What is the relevance of shielding the Cable?**

   Wrapping each conductor pair with a foil shielding further reduces the crosstalk among pairs, and wrapping all four of the twisted-pairs in a foil or braided metallic shield reduces a cables susceptibility to EMI and RFI in noisy cable environments. Thus, STP (Shielded Twisted Pair) cables employ both types of shielding, giving them the highest immunity to all interference types. FTP (Foil Twisted Pair) and ScTP (Screened Twisted Pair) cables employ only the outer foil or braided-conductor shielding, giving them enhanced immunity against external EMI and RFI, but no more protection against crosstalk than an equally-constructed UTP cable.

3. **What are the various IEEE cabling Standards, how they differ?**

   10Base-T and 100Base-T are the IEEE (Institute of Electrical and Electronics Engineers) standards defining the electrical and physical characteristics of twisted-pair cabling for use in 10 Mbps (Megabits per second) and 100Mbps Ethernet connections. The "T" stands for Twisted pair, and these two Ethernet connections use wire pairs 2 and 3 to transmit and receive information, corresponding to the orange and green twisted pair conductors shown in Figure 2. Nowadays we use the Gigabit Ethernet (or 1000Base-T), where all four conductor pairs shown in Figure 2 above, are used to transmit and receive information simultaneously.

4. **What are the various EIA/TIA cabling Standards, how they differ?**

   568A and 568B are EIA/TIA(Electronics Industry Association/Telecommunications Industry Association) wiring standards specifying two different RJ-45 pin assignments for the orange and green conductor pairs in Category-type twisted-pair cables. The wiring for two different conductor/pin configurations is shown in Figure 4, and the same are tabulated in Table-1below. You should observe that, the Ethernet patch cords with connectors wired using the same standard on both ends, are referred as *"Straight Through Cable"* and those with different standards are referred as *"Crossover cable".* In Brief, to create a straight-through cable, you'll have to use either T-568A or T-568B on both ends of the cable. To create a cross-over cable, you'll wire T-568A on one end and T-568B on the other end of the cable. The general structuring of Straight Through and Cross Over Cable is shown in Figure 5 below.

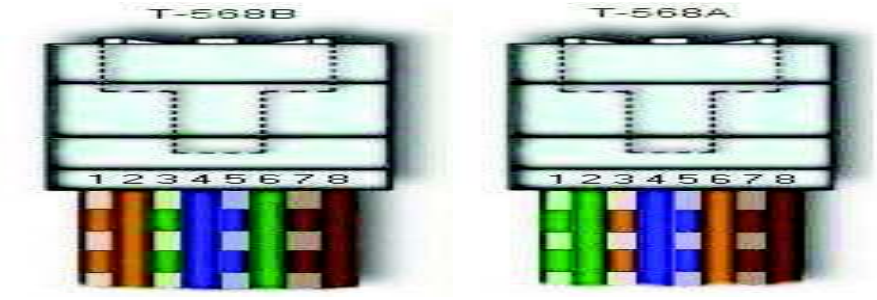


Figure 4: 568A and 568B are EIA/TIA wiring standards -specifying different RJ-45 pin assignments

Table 1: Wiring Diagram for EIA/TIA Standards 568a and 568b

Pair #	Wire	Pin #
1-White/Blue	White/Blue	5
	Blue/White	4
2-White/Green	White/Green	1
	Green/White	2
3-White/Orange	White/Orange	3
	Orange/White	6
4-White/Brown	White/Brown	7
	Brown/White	8
568-A Wiring Diagram		

Pair #	Wire	Pin #
1-White/Blue	White/Blue	5
	Blue/White	4
2-Wht./Orange	White/Orange	1
	Orange White	2
3-White/Green	White/Green	3
	Green/White	6
4-White/Brown	White/Brown	7
	Brown/White	8
568-B Wiring Diagram		

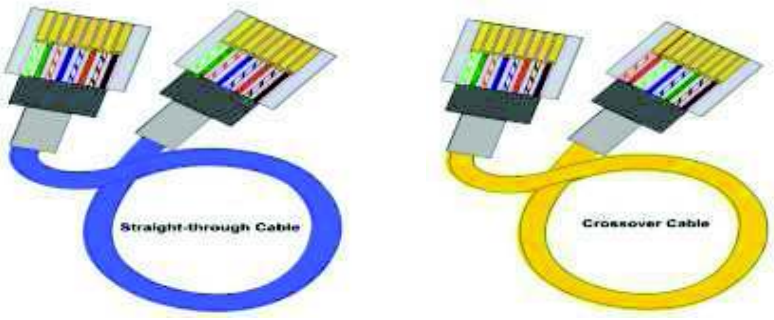


Figure 5: Straight-Through Cable and Cross Over Cable

5. **When to use which type of cable?**

The straight-through cables are used when connecting Data Terminating Equipment (DTE) to Data Communications Equipment (DCE), such as computers and routers to modems (gateways) or hubs (Ethernet Switches). The cross-over cables are used when connecting DTE to DTE, or DCE to DCE equipment; such as computer to computer, computer to router; or gateway to hub connections. The DTE equipment terminates the signal, while DCE equipment do not. To simplify, we tabulated the generalized situations, where you might be expected to use the respective cables. i.e. Crossover cable or Straight Through Cable:

Computer to Computer	–	Crossover
Switch to Switch	–	Crossover
Computer to Modem	–	Straight Through
Computer to Switch	–	Straight Through
Switch To Router	–	Both (if problems, go with Crossover)

6. **What is the difference between CAT 5/ 5e / 6 cables, when to use which cable?**

Making the choice between types of Ethernet cables available for networking and connecting their computers to the Internet viz. Cat 5, Cat 5e, and Cat 6 cables can be confusing. To distinguish between the various types of cables, you have to understand the nomenclature, the term *Cat* being short for "Category", whereas the numbers and letters to follow are all used to indicate performance. These performance designations make it easier to choose the right type for various purposes such as networking computers together or using peripherals including hubs and routers. All three types of cables, Cat 5, Cat 5e,

and Cat 6, are comprised of four pairs of UTP (unshielded twisted pair), but the amount of transmissions the cable will be able to support is up to its category rating.

The Original Cat 5 Cable : An old standard in the industry, Cat 5 cable is able to perform up to 100MHz and is still widely used for a variety of applications, although most new installations will use Cat 5e or higher. Able to support 10/100 Ethernet and fast Ethernet, Cat 5 cable is upwardly compatible with the Cat 5e version.

The Improved Cat 5e Cable : With improved durability over Cat 5, the protective outer covering of Cat 5e cable is thicker and therefore more suitable and reliable for more situations than its earlier counterpart. There are several other differences between this version and its predecessor including its backwards compatibility, as it will work along with either 10BaseT or 100Base T networking hubs and cards. There is also less cross talk or electronic interference with Cat 5e as opposed to Cat 5 cable thanks to improved signal capabilities. In terms of bandwidth, Cat 5e supports gigabit Ethernet connections of up to 350MHz, more than trebling the 100MHz of a Cat 5 cable.

Remember that Cat 5e cable is not rated for outdoor use, although many people do without incident. If you must use this cable outside, add a conduit such as one made from PVC to keep moisture away. The safe operating temperature for Cat 5e cable is anywhere from 10 degrees Celsius to 60 degrees Celsius.

Also, with this particular category cable, 100 meters is the maximum length you will be able to use the cable without the benefit of either a network bridge, hub, or amplification to strengthen the signal.

*The Cat 6 Cable :*Certified and designed specifically for gigabit use, Cat 6 cable reduces cross talk even more than its predecessors by improving upon the original Cat 5 version with wires featuring extra twists. The use of Cat 6 cable does not guarantee that the network will be a full gigabit network, for this to be achieved each and every one of the components must be gigabit certified. Unless your network meets this criteria, opt for Cat 5e which will provide high quality speeds while saving money in the process.

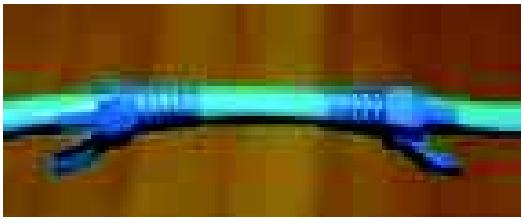
For quick reference, here are the ratings of the various category cables: Cat 5 up to 100MHz ; Cat 5e up to 350MHz; Cat 6 up to 550MHz

1.2.1 Assembling Patch Cable

By learning the theoretical aspects of structured cabling, you might be exhausted. So, let's apply our learnt skills in a practical manner, just follow the instructions given below and you will be able to produce your own patch cable assembly.

Steps to assemble Patch Cable:

1. Cut the cable to the length that you will need.



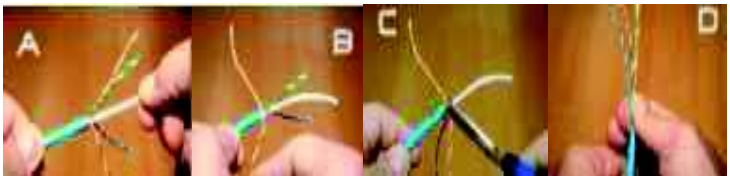
2. Skin the cable about 1.5” down.



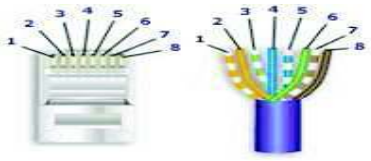
3. Remove all of the twists in the cables pairs. Un-twist each pair, and straighten each wire between the fingers.



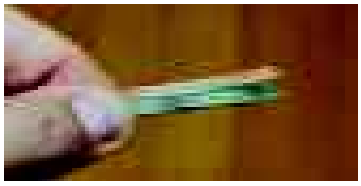
4. Cat 6 cable has a center spine that needs to be removed. Pull on the spine and fold the pairs back. Then cut the spine as close to the cables end as possible. The process is shown in steps A,B,C,D to be executed sequentially



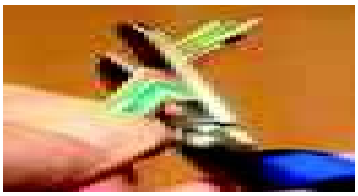
5. Place the wires in the order of one of the two diagrams shown in Figure 4 above, i.e. for EIA/TIA - 568B or 568A. Here we have chosen the 568B diagram which is by far the most popular. If you are unsure, go with the 568B wiring.



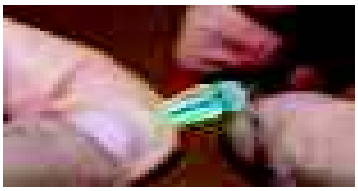
6. Bring all of the wires together, until they touch. Hold the grouped (and sorted) wires together tightly, between the thumb, and the forefinger. At this point, recheck the wiring sequence with the diagram.



7. Cut the wires on a very sharp angle to make it easier to install the load-bar(in the next step).



8. Insert the loadbar on the wires one wire at a time.
This is why we recommended cutting the wires on an angle.

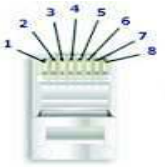


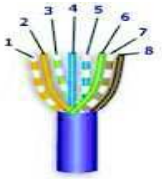
9. Check the wiring sequence one more time. Than slide the load bar down all the way and make a straight cut about 0.25 past the loadbar. A perfectly straight cut is essential here.



10. Insert the connector onto the loadbar assembly.
Hold the plug with the copper connectors up and the locking clip facing down.
In this configuration, the Brown Pair of wires should be to the right side







11. For Crimping, push the connector all of the way in and then squeeze down all the way on the crimper. Remove the connector from the crimper body.



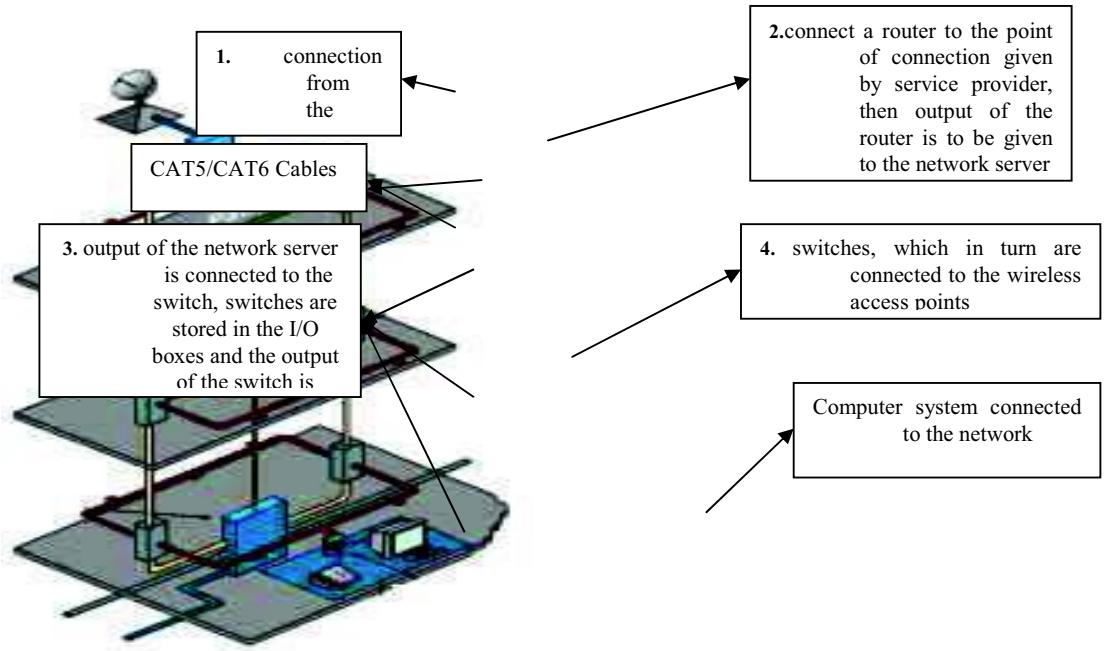
12. Repeat the procedure on the other end of the cable using the same wiring diagram. NOTE: If you wish to make a crossover cable, than use the other diagram (in this case 568-A)

13. Test the cable using a high quality four pair LAN cable tester.



Now you are perfectly ready to do structured cabling and design the network of your own. To perform the structured cabling in a building or so you need to refer to Figure 6, given below. It will clear your understanding, related to structured cabling in big layouts, thus it relates to the role of structured cabling in network design. The Figure 6 explains that, Once the connection from the

service provider is installed in the institutional premises, they left it with connecting point, after which you are suppose to start your network. You are required to connect a router to the point of connection given by service provider, then output of the router is to be given to the network server which is responsible for content management, bandwidth regulation, malware protection etc. The output of the network server is connected to the switch with desired number of ports. The switches are stored in the I/O boxes and the output of the switch is cascaded with the other switches, which in turn are connected to the wireless access points. Now in between the devices the paired cable, dully connected with the connectors at both the ends is running. Nowadays patch cords/cables, as per the standards are also available, but they are bit expensive. In general, networking engineers purchases the cable bundle and crimp the RJ45 connectors at the ends of the cables through the crimping tool.



☞ Check Your Progress 1

1. Differentiate between Straight Through Cable and Crossover Cable?

……

……

……

2. Identify the suitable cable i.e. straight through/ crossover cable, required to connect the following

a) Computer to Computer
b) Switch to Switch
c) Computer to Modem
d) Computer to Switch
e) Switch To Router

……

……

……

3. Differentiate between 10 Base T and 100 Base T - IEEE standards of twisted pair cabling

…….

…….

…….

1.3 INTEGRATING HOME COMPUTERS

Electronic environment of any house involves wired or wireless connectivity among various devices viz. computers themselves, between computers and printers, etc. Since, you had already studied the concept of structured cabling, you are expected to firstly understand "How the patch cables can be used to directly connect the computers?" , then we will extend our discussion in the subsequent section, to let you understand "How switches or Hubs can be used to connect the computer systems?" and in the similar manner we will proceed for the development of wireless networks too

So, let's start our discussion with "**How to connect two computers simply by cross-over cable (without router or switch)?".**The steps are listed below, just follow them and you will get it done

1.3.1 How to Connect Two Computers by Using Cross-Over Cable?

This section involves the connectivity of the computers through a cross over cable, without using the network devices like switch or hub. However we will discuss the establishment of computer network by using the network devices in our subsequent section 1.4.1. In this unit we are assuming that the user are having Windows operating system .

STEPS

1. Switch ON the computers
2. Connect both computers with a cross over cable(Cat 5/6) having RJ 45 connector crimped at its both ends.
3. Go to control panel
4. Click on network connections
5. Right click on cable connections.
6. Click properties
7. Pick internet protocol (TCP/IP) & press properties.
8. Click on choose following IP address

   IP address: 1 choose 192.168.1.1
9. Network is 255.255.255.0
10. Press ok and close
11. Now repeat the steps (1 to 10) on the other computer but choose different IP address say it is 192.168.1.2

12. Now test the connection by using cmd command

- Go to start
- Click Run
- Type cmd
- Type ping IP address if you are on system with IP 192.168.1.1 (i.e. ping 192.168.1.2)
- If it says time-out, that means that you don't have a connection with other computer

Interconnectivity facilitates the data sharing among the computers. So, you are required to understand, "How to share the data among the computers, connected to each other in either mode i.e. wired or wireless " Just follow the steps listed below and you will get the data shared among the computers, listed steps work for both wired and wireless connections.

1.3.2 How to Share Data Between Two Computers?

In this section we are going to exploit one of the basic need of computer network, i.e. the sharing of data between the computers. The section gives you the stepwise guidance, to perform the task of data sharing. In this unit we are assuming that the user are having Windows operating system .

STEPS

1. Assign IP address to both computers (in the same manner as discussed above)
2. Go to control panel.
3. Choose network and internet option.
4. How choose network and sharing center option.
5. Choose manage network connections option.
6. Right click area connection and select properties option.
7. Select TCP/IPV 4 and click properties
8. Select "Use the following IP address" say it be 10.1.1.1, say subnet mask is 255.0.0.0. now click ok.
9. Now click close.
10. Close the network connections window.
11. Close the network and sharing center window.
12. Close the network and internet option window.
13. Connect both computers by cable
    a) for different devices: straight cable e.g Switch → Computer
    b) for same devices: Cross over cable
14. Now you have to share folder or file that you want to access from other computer
    i) Create a folder say on desktop

ii) Right click the folder and choose properties

iii) Select the sharing tabs.

iv) Select the advanced sharing button.

v) Check the share this folder option.

vi) Press the permission button.

vii) Check the permission say full control/ Change/ Read to be allowed or deny.

viii) Click the apply button for all opening you made in sharing section and then finally close the sharing properties tab.

15. Turn file sharing ON.

a) Go to control panel.

b) Select network and sharing center.

c) Turn ON the file sharing option and click apply.

d) Choose the option “No make the network that I am connected to a private network” if you don’t want data to be shared by All. Otherwise choose “Yes, turn on file sharing for all public networks”.

e) For security you may activate the “Password protected sharing” by turning ON the password protected sharing and click apply.

f) Now close all the windows/ tabs opened till the above task is done.

16. To access folder that you have shared

a) Go to other computer by typing //10.1.1.2

b) Select share folder icon

c) Create new folder and close the opened windows.

17. To get other computer on to your screen

a) Type or Run mstsc command in the search option.

b) Remote desktop option get activated.

c) Type the IP of the computers you want to connect to say 10.1.1.2 in the computer section and click connect

18. Now you can share the data of 10.1.1.2

1.4 `

Till the moment you understood that a network could be as simple as two users sharing information through a diskette or as complex as the Internet that we have today. The Internet is made up of thousands of networks interconnected through devices called hubs, bridges, routers and switches. These connecting devices are the building blocks of a network and each of them performs a specific task to deliver the information that is flowing in the network. So, it's time to learn how to connect the computers by using these connecting devices. We will limit our discussion to hubs and switches only, as they are widely used in developing LAN. So, let's understand the devices and their utilities in brief.

HUB : A hub is a connecting device that all end workstations are physically connected to, so that they are grouped within a common domain called a network

segment. A hub functions at the physical layer of the OSI model; it merely regenerates the electrical signal that is produced by a sending workstation, It is a shared device, which means if all users are connected to a 10Mbps Ethernet hub, then all the users share the same bandwidth of 10 Mbps. As more users are plugged into the same hub, the effective average bandwidth that each user has decreases.

SWITCH: Switch is another important device when we talk about computer network on broader spectrum. It is used at the same place as hub is but the only difference between the two is that switch possess switching table with in it. Switching tables store the MAC addresses of every computer it is connected to and send the data to only requested address unlike hub which broadcasts the data too all the ports.

NOTE:

1. A switch functions at the same OSI layer as the bridge, the data link layer. In fact, a switch can be considered a multi-port bridge. While a bridge forwards traffic between two network segments, the switch has many ports, and forwards traffic between those ports. One great difference between a bridge and a switch is that a bridge does its job through software functions, while a switch does its job through hardware implementation. Thus, a switch is more efficient than a bridge, and usually costs more.

2. Switches are introduced to partition a network segment into smaller segments, so that broadcast traffic can be reduced and more hosts can communicate at the same time. This is called micro segmentation, and it increases the overall network bandwidth without doing major upgrade to the infrastructure.

3. Hub is Unmanaged device where as switch can be a managed or unmanaged. Both support full duplex communication i.e. any computer can send data to any other computer connected through the connecting device. The devices can have 4/8/16/32 ports and you may connect two or more than two switches or hubs, to form the cluster of networks. To a N port hub/switch, one port may be used to connect to the server and other N-1 ports may be used to connect the client devices.

4. you are not desired to configure the HUB/SWITCH they got automatically adapted to the networks, unlike the case of Routers and Access points where you need to explicitly configure the network.

1.4.1 How to Connect Computers USING HUB / SWITCH ?

In section 1.3.1 we discussed How to connect computers using cross over cable?, in this section we are extending the concept of the computers connectivity through network devices like hub/switches. The steps desired to be performed are given below:

1. Connect the hub to the power source through its adapter and switch it ON

2. Take Straight cables with RJ 45 connector connected to its both ends, use it to connect the Network Interface Card(NIC) of all computer system to the different ports of Hub/switch as shown in the figure below.



3. Switch ON the computers
4. Go to control panel
5. Click on network connections
6. Right click on cable connections.
7. Click properties
8. Pick internet protocol (TCP/IP) & press properties.
9. Click on choose following IP address
10. IP address: 1 choose 192.168.1.1
    Network is 255.255.255.0
11. Press ok and close
12. Now repeat the steps (1 to 10) on the other computer but choose different IP address say 192.168.1.2 for second computer and so on.
13. Now test the connection by using cmd command
    a) Go to start
    b) Click Run
    c) Type cmd
    d) Type ping IP address if you are on system with IP 192.168.1.1 (i.e. ping 192.168.1.2)
    e) If it says time-out, that means that you don't have a connection with other computer

1.4.2 How to Create Cluster OF Switches/Hubs ?

Let say you have a network at home, the Hub/Switch you bought only got 4 Ethernet LAN ports. 2 ports are connected to computers and 1 port is connected to notebook. You then found out you still have 1 computer and 1 notebook to connect to network, but you only left 1 Ethernet LAN port on Hub/Switch, so how to connect both devices to the network and solve this problem?

The solution is easy. You can create a network cluster by connecting one more hub/switch to one of the ports of the existing hub/switch by using cross-over cable. After that, you can connect computer and notebook to the switch's normal port by using straight cable, finally they are all connected to network and able to access Internet. The LED on the switch will show you which ports are connected.

1.4.3 How to Configure a Wireless Network?

After going through the sections given above , you might have understood the efforts involved in the development of any wired network. So, to simplify the complexities of wired networks, the technology has explored the option of wireless network, which involves one more network device i.e. Access Point. In this section we will let you understand the configuration of wireless network. The steps to configure a wireless network are given below :

1. Switch on the computer and Access point
2. Activate the wireless network mode of computers

3. Connect Access point to the computer through straight cable
4. Open the web browser
5. Type the IP address of Access point given with the access point at the place where URL is typed, and press enter
6. Access point window will be opened
7. Generate SSID and Password from the opened Access point window
8. Ping the access point through you computer by typing "ping ip address of access point" from the command prompt, successful response assures the connectivity
9. Now you may disconnect the wired connection between the computer and the access point.
10. Activate the wireless network mode of other computers in the near vicinity of the access point, they will automatically detect the network.
11. Once the network is detected, to get connected to that wireless network, just select the respective network and it will ask for the SSID code and Password.
12. Provide the assigned SSID code and password, press enter and you are connected to that network.
13. Now you may assign the respective ip addresses to the computer systems connected to the access point and use the same process as discussed above to verify the connectivity among the computers

☞ **Check Your Progress 2**

True / False

a) Switch is used to partition the network

b) Hub is an unmanaged device

c) Hubs/Switches got automatically adapted to the networks,

d) Routers and Access points are required to be configured explicitly.

e) A switch can be considered a multi-port bridge.

…….

…….

1.5 CASE: DESIGNING & DEVELOPMENT of SMALL NETWORKS

A reputed educational group has three institutes in the same campus. All institutes have separate independent facilities to administer and manage. However the institute was lagging in information resource management as a whole. All facilities are available but people are unaware to use them optimally and systematically. The Campus was catered with fantastic internet facility of 2Mbps speed Lease line and broadband too. The broadband connections were available in their hostels, which are off campus for boys and in campus for girls.

The off campus students were catered with internet facility through the separate

broadband connection, which were generally tempered by the students residing there and hence connectivity problem was a regular feature, apart from this the occupancy of the hostel was approx 27 students, nine students at each floor and a connection is available at ground and second floor, the third floor students are sharing the connection with the broadband connection available at second floor. As 18 students got attached to one connection the speed of internet got considerably decreased, so the students contacted the IT department for a solution. The problem was expected to be solved without going for a separate broadband for third floor.

However the in campus, girls hostel was enjoying the uninterrupted internet access because the 2Mbps speed was directly at their disposal. The students here were using it generally for chatting, movie download, torrents etc. and the purpose was not at all academic. As many movies, software and other downloadable contents were put in process of download, the actual working of entire campus was hampered. The internet speed was drastically reduced and sometimes it got choked too, thus entire campus was suffering.

The IT situation was pathetic in the sense that other departments were not at all utilizing the existing resources in the sense, the Computer Lab as a whole was on WIFI and the systems deployed there were desktops; the accounts section was taking data backup on CDs. The faculties were using pen drives to carry their presentation to the classes which actually let VIRUSES to enter into the network and hence the systems need frequent maintenance. The students who participates in the lab sessions frequently complaints that on which ever system they worked in the last class they are unable to get that computer system in the subsequent class and as a result the tasks executed in the previous class are non traceable. Apart from this the students were also equipped with laptops and entire campus is WIFI, the students have a regular practice to change their IP addresses assigned to them by IT department as a result of which IP conflict occurs and it leads to create problems in accessing the internet for other students too.

The internet service provider is a reputed organization , providing sufficiently nice services but generally because of the excavation process and other tasks performed by other companies in real estate or so, the cables required to provide internet services in the institute campus are damaged hence working/communication of entire institute is disturbed. IT manager was expected to resolve this issue too. Apart from this problem, the institute is in expansion mode and frequently new EPABX numbers are desired with a traceability that how many lines are making calls outside the campus and their billing was also expected to be maintained and informed to the accounts section for necessary actions. Sometimes the faculties are on leave and their lecture suffers, it was desired by management to make some arrangements that , faculty should be able to deliver the lecture even when he/she is out station. Institute was planning to develop this facility of video lectures to use for conferences or so, to be organized in the campus. After all, institute need to have internet website/intranet website and extranet facility to facilitate the employees working even from outside.

After going through the case given above, you might have realized the presence and importance of networking in our day to day life. Apart from this you might have identified various network components required to establish the computer network. Since you had already gone through various networking concepts in the previous units of this course BCS 041, you are required to make yourself comfortable, to do the tasks given below

ASSIGNED TASKS:

TASK-1 Being an IT manager of the Institute, Identify the problem areas and problems specific to the identified area in the institute. Present your identified

problems in a tabular format.

TASK-2 Identify solutions (both hardware and software), which may be used to resolve the identified problems. Identify the cost effective solution you wish to implement, justify your choice with suitable arguments.

TASK-3 Prepare a summarized requirement report, targeting the identified problem with the proposed solution, in the form of a table, so as to simplify decision making at management end.

SOLUTIONS TO THE ASSIGNED TASKS

Lack of knowledge related to networking and related techniques is the prime cause of problems in the entire campus of the educational institution. Technical Workforce of the institute is unaware of network devices and their usage viz.

1. where to go for wired networking and where for wireless networking
2. which servers are to be designed viz. DNS, Backup server etc
3. non awareness of firewalls or content filtering software
4. how to administer the network connection directly coming from the service provider.

Apart from the mentioned lacunas in the existing network of the campus, there are many more deficiencies; we will discuss them as the discussion proceeds.

From the given Case, we identified following ***problems*** persists, in the respective areas and sub areas :

1. **Entire Campus :**
   a) The connectivity from service provider is wired connectivity, as a result of which as and when the connection cable got damaged due to excavation the entire campus got disconnected from the internet services.
   b) The bandwidth distribution is open in the sense, the lease line from the service providers router is directly catering the institutes access points, thus the bandwidth can be used by the persons outside the institutional premises, and this is quite unsafe because someone it's a security abuse for the institutional network.
   c) Further, the usage of bandwidth by outsiders leads to network choking i.e. network hangout.
   d) Students equipped with laptops were changing the ip addresses, thus ip conflict is a frequent issue of the respective network.
   e) No Intranet or extranet web facility, total reliance on Internet. Thus, in the absence of Internet connectivity, entire communication is on standby mode.
   f) Comparatively low bandwidth, as desired for video conferencing or online lectures.
2. **Hostel :**
   a) On Campus Hostel :

      Since there is no control over the bandwidth regulation, the users are consuming the available bandwidth for non – academic jobs, viz. online

gaming, movies download etc.

b) Off Campus Hostel :

i) The Internet connectivity is given through broadband connection, the positioning of the broadband device is not safe, thus the users were able to hack the device password by using hacking software by directly connecting their systems to the broadband device through the network cable.

ii) The broadband connection was overloaded, thus the Internet speed is slowed down.

3. **Computer Lab. :**

a) Just to save the cost of wiring the entire laboratory systems were on wireless connectivity through the access points, I agree its cost effective, but at the time of lab maintenance and up gradation, the situation become quite challenging, because the network speed is quite slow in wireless mode. Apart from this if connection is lost in between the software installation, then entire file gets corrupt.

b) Since it is very much impossible to find the seat on the same computer system in the subsequent class, the students are to redo the task performed in previous session. Thus the students are unable to recollect the data or the task executed in their past classes.

c) Usage of flash memories/pendrives for porting the data, leads to virus prone network.

4. **Accounts Section :**

a) The backup is taken on the CD/DVDs. if prior to take the backup, system crashes out, and then nothing can be done.

We know that if the problems are identified then half of the job is done.

So, its time to talk about solutions and related alternatives, further we are suppose to identify the optimal and feasible solution.

Network solutions are proposed in the sequence, the problems are identified above As per the identified problem following are the requirements for troubleshooting.

Hardware: Gigabit Ethernet cards, Cat 6 cables, servers (Domain name Server, Backup server),Online Ups for servers, switches, I/O boxes, Connectors, LAN meter, Crimping tools etc.

Software: Proprietary or Open source server software, content filtering software, Bandwidth regulator , Anti- malware software, or Software Firewall.

Let's see how above mentioned resources should be utilized in network designing and development, such that the identified problems of respective identified areas are solved. First, let's start from the Campus Related issues.

Identified Network Solution for identified areas and sub areas:

1. **Entire Campus :**

a) Solution to connectivity problem: The campus should have wireless lease line and not the wired one or may have both types, because wired connectivity has its own advantage related to the speed of operation.

b) Solution to bandwidth distribution: Instead of directly connecting the access points to the service provider connection available through their router. The Network Input/output mechanism should be laid by using the Gigabit Ethernet cards, where the cards are installed on the motherboard slots of the computer system, such that connection from the service provider router goes to the input cards and the output card is connected to the access points through the switches. In between the input and the output card, respective software works viz. open source content management software like squid, software firewalls, anti viruses, anti spyware, anti spam software etc, thus the connection is secure and well regulated. Actually, this computer system acts as a server, which is responsible for bandwidth regulation, content management etc.

c) Solution to usage of bandwidth by outsiders : The router should be protected by necessary SSID(Service Set Identifier) number and password protection mechanism, which prevents outsiders from accessing the network connectivity.

d) Solution to ip conflict: To get rid of the ip conflict problem, we may bind the allotted ip address with the Mac id of the laptop OR we may keep the administrative rights with the system administrator and let the student to act as a user, so no permissions are available to change the ip address.

e) Solution to total reliance on Internet: The institute must design at least a mail server, such that the in campus communication goes on without hindrance. Further, they should design an intranet website, because everything cannot be for public domain.

f) Low Bandwidth: Bandwidth requirement is to be reworked as per the usage, and the connection capacity is to be revised from 2 mbps to the required one.

2. **Hostel :**

a) On Campus Hostel :

Since the on campus hostel is very much in premises so the solution to the bandwidth distribution discussed above, solves this problem. The software firewall have the feature to allot the bandwidth to the particular series of ip addresses which may be allotted to the students or teachers, they are having many other options too.

b) Off Campus Hostel :

i) Broadband connection safety: The broadband device should be protected by installing an I/O Box, mounted close to ceiling, thus the students cannot access the port connections of the broadband device.

ii) Overloaded broadband connection: Two solutions are possible; either we should go for one more broadband connection, which incurs a recurring cost to the institution. The other solution is that we install a switch for the top floor of the hostel where wired connectivity is provided, this involves one time cost of switch and cabling, further the advantage of mobility is curtailed.

3. **Computer Lab. :**

a) It is advised that the computer labs should have wired connectivity

because, wired network has better speed then wireless network. Apart from the speed the connection is dedicated, thus the possibility of losing network connection, in between the process of installation or so, is rare. We may be using the labs for exam purpose or so, in that situation loss of connectivity or so leads to tremendous problems. I agree that the wireless connectivity is quite manageable and cost effective, but at the time of lab maintenance and up gradation, the situation become quite challenging, because the network speed is quite slow in wireless mode. Apart from this if connection is lost in between the software installation, then entire file got corrupt. So, Labs should have wired connectivity

b) Recollecting the data: Here is the requirement to design a DNS (Domain Name Server), where some memory space is allotted to each student, which may be according to their roll numbers or so. Thus through DNS, students are always able to work in their allotted space, and can recollect the job done in previous class. But, there is a requirement of On-Line Ups with the DNS server, because if the power goes Off, then restart of DNS is time consuming

c) Data Portability: Again DNS will be the solution for accessibility of data in class rooms or labs or anywhere else, thus we can block the USB ports and let the entire network be managed through DNS and Intranet.

4. **Accounts Section :**

a) Backup on CD/DVDs : A Backup server is desired to be designed for solving this problem.

1.6 SUMMARY

After going through this unit you are now equipped with the skills desired to structure a wired or wireless computer network. Now you are required to make practice of the learned concepts and realize the facts and figures of networking. Here you learned the concepts related to the Structure Cabling which is further extended to the skill based assembling of Patch Cables, which are widely required to connect computers and network devices, in a wired network. The concepts of wired network are covered under the heading integrating home computers, which enables us to understand the concepts related to How to connect two computers by using crossover cable? and How to share data between two computers? The unit also explored the creation of a small network in both wired and wireless mode, by using hubs, switches and access points. The understanding of the concepts learned in this unit, enabled your application skills through a case given in the end. Hope, you are in the position to apply the learned concepts.

1.7 SOLUTIONS / ANSWERS

☞ Check Your Progress 1

1. The straight-through cables are used when connecting Data Terminating Equipment (DTE) to Data Communications Equipment (DCE), such as computers and routers to modems (gateways) or hubs (Ethernet Switches). The cross-over cables are used when connecting DTE to DTE, or DCE to DCE equipment; such as computer to computer, computer to router; or gateway to hub connections. The DTE equipment terminates the signal, while DCE equipment does not.

2. a) Computer to Computer – Crossover

   b) Switch to Switch – Crossover

c)	Computer to Modem	–	Straight Through
d)	Computer to Switch	–	Straight Through
e)	Switch To Router	–	Both (if problems, go with Crossover)

3. 10Base-T and 100Base-T are the IEEE (Institute of Electrical and Electronics Engineers) standards defining the electrical and physical characteristics of twisted-pair cabling for use in 10 Mbps (Megabits per second) and 100Mbps Ethernet connections. The "T" stands for Twisted pair, and these two Ethernet connections use wire pairs 2 and 3 to transmit and receive information, corresponding to the orange and green twisted pair Nowadays we use the Gigabit Ethernet (or 1000Base-T), where all four conductor pairs, are used to transmit and receive information simultaneously.

☞ **Check Your Progress 2**

True / False

a) True

b) True

c) True

d) True

1.8 REFERENCES

WEBLINKS

- http://www.andcable.com/files/UnderstandingEthernetPatchCords.pdf
- http://en.wikipedia.org/wiki/Structured_cabling
- http://www.lanshack.com/make-cat5E.aspx
- http://www.iplocation.net/tools/rj45-wiring.php

EBOOKS

- Cisco Networking Academy Program CCNA 1: Networking Basics v3.1
- *IP Network Design Guide from IBM by Martin W. Murhammer, Kok-Keong Lee, Payam Motallebi,Paolo Borghi, Karl Wozabal*

UNIT 2 INTRODUCTION TO NETWORK ARCHITECTURES

Structure **Page No.**

2.0 Introduction 26
2.1 Objectives 26
2.2 X.25 Architecture 26
2.3 Atm Network 28
2.4 IPv4 and IPv6 Overview 41
2.4.1 Classes of IP Address
2.5 Summary 45
2.6 Solutions/Answers 45

2.0 INTRODUCTION

In the simplest form, data transfer can take place between two devices which are directly connected by some form of communication medium. But it is not practical for two devices to be directly point to point connected, because of two reasons (i) the devices are very far apart and (ii) there is a set of devices, each of whom may require to connect to others at various times. Solution to this problem is to connect each device to a communication network. As you know computer networks means interconnected set of autonomous computers, in order to meet the needs of various applications, networks are available with different interconnection layouts and plans, methods of access, protocols and data carrying capacities. Network architecture is a complete design of a communications network. Primarily we can say that it is a framework for the specification of a network's physical components, their functional organization and configuration. Network architecture also includes the operational principles and procedures. This unit is an introduction to network architecture, in which we will discuss about different network architectures like X.25, Frame Relay, ATM. Further, it covers IPv4 and IPv6 protocol details; we will also discuss the mechanisms for implementing/deploying IPv6.

2.1 OBJECTIVES

After going though this unit, you should be able to:

- Understanding the working of various Network architectures
- differentiate between X.25, Frame Relay and ATM Architecture
- Know the functions of X.25, Frame Relay and ATM layers
- describe how X.25, Frame Relay and ATM protocols works;
- Know the need of IPv6 protocol
- Compare between the IPv4 and IPv6

2.2 X.25 ARCHITECTURE

Before discussing about X.25, we will refresh our knowledge about switching techniques. As you may know following are the basic switching techniques:
Circuit Switching: Circuit switching is used in the telephone networks to transmit voice and data signals. To enable synchronised transmission, circuit switching establishes a dedicated connection between the sender and receiver involved in the data transfer over the network. As a result, the connection consumes network

capacity whether or not there is an active transmission taking place; for example, the network capacity is used even when a caller is put on hold.

Packet Switching: In contrast to circuit switching, packet switching ensures that the network is utilised at all times. Data to be sent is broken down into chunk of bits or packets. Each packet contains data and header information for control. At each node the packet is received, stored briefly and passed on. At each node the packets may be put on a queue for further movement into the network. It does this by sending signals even in the small unused segments of the transmission — for example, between the words of a conversation or when a caller is put on hold. There are two approaches to the above kind of transport:

1. ***Datagram***, where each packet can take any path through the network as long as they all reach the destination.

2. ***Virtual Circuit***, where all the packets are routed through the same path without having the path dedicated. The path segments may carry many virtual circuits. Datagram allows for dynamic handling of congestion and no call setup is necessary. Virtual circuits allow for sequencing, error and flow control.

X.25 is an old standard protocol suite for packet based wide area network. The old networks mainly telecommunications companies and ATM's (automated teller machines) were following X.25 protocols for packet switching based network. These WAN's are having packet-switching exchanges and leased communication channels. At present X.25 protocols has been replaced by other better and less complex protocols of TCP/IP suit however, the service is still in use and functioning in some places and applications. Some student are interested to know that why it is called with such name X.25? The reason is International Telecommunication Union (ITU) publishes some series of technical books, among these technical books; there is a larger set of X-Series specifications on public data networks. The X.25 specification is only a part of that X-Series specification on public data networks.

The common perception for development of X.25 was to develop a universal standard for packet switching network. X.25 does not specify how the network operates internally; it specifies only the interface between public switched networks and the users. As shown below in the figure DTE (data terminal equipment) is a user/subscriber, DCE (data communications equipment) is a device between a network and user, in general it is MODEM device, DSE are nothing but data switching exchanges in a packet switching based WAN.

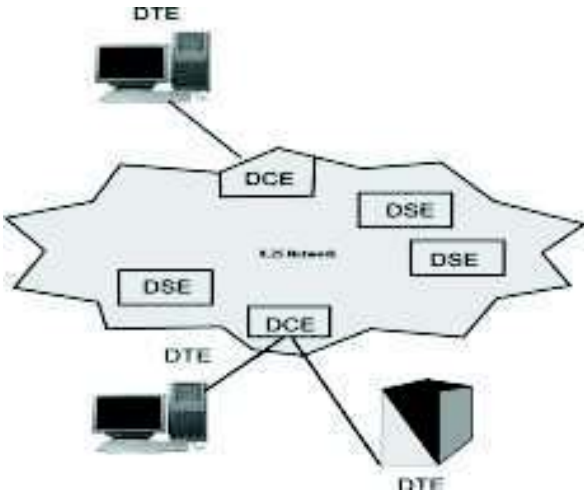


Figure 1: X.25 Network and its components

X.25 is specified on 3 layers:

1. Physical layer
2. Data link layer
3. Network layer

X.25 Network provides the means for these users (DTE) to communicate with each other. In the context of X.25 Data link and Network Layers, an X.25 DCE is the local network node to which the DTE is connected. The X.25 protocol defines the rules for the communication between the DTE and the DCE. You may again note that communication within the WAN may be entirely by some other mechanism. Following are details of each layer of X.25:

- Physical layer: Specify the physical, electrical and interface between host and network. It also specifies functional and procedural characteristics to control the physical link between a DTE and a DCE. Common implementation is X.21protocol.
- Data link layer: Deal with data transmission over an between user equipment and routers. Error control and flow control are its main responsibilities. This layer consists of the link access procedure for data interchange on the link between a DTE and a DCE.
- Network layer: this layer specify a packet-layer protocol for exchanging control and user data packets to form a packet-switching network based on virtual calls. It has main functions like Addressing, Flow control, Delivery confirmation, etc. Also, it allow to established Virtual Circuit and send packet reliably.

X.25 is connection oriented architecture and support switched virtual circuits (SVC) and permanent virtual circuits (PVC). Switched virtual circuits are established on the need basis. SVC is established when a call is made and broken down after the call is completed. On the other hand, permanent virtual circuits are almost leased kind of connections, which provide a dedicated connection between DTE's. X.25 sessions are established when one DTE device contacts another to request a communication session. The DTE device that receives the request can either accept or refuse the connection. If the request is accepted, the two systems begin full-duplex information transfer. Either DTE device can terminate the connection. After the session is terminated, any further communication requires the establishment of a new session.

2.3 FRAME RELAY

Frame Relay is a virtual-circuit based WAN that was designed to provide more efficient transmission scheme than X.25. It provides connection oriented services at reasonable speed and low cost. Interestingly in Frame relay, the packets are now of variable length (called as frame, which is a reason such architecture is named FRAME RELAY) with less overheads. Some of the main drawbacks of X.25 are as follow:

1. X.25 has a low 64 kbps data rate, [In 1990 It was very less]
2. X.25 has extensive flow Central and error central at both the data link and network layer (Because in 1970-80 available media was more prone of these errors and an objective of X.25 was to develop a global system which may have more possibility of errors). It creates large overhead and slow transmission.
3. X.25 was designed for private use, not for Internet (public use). It has its awn network layer and Internet has its awn hence packet is encapsulated in X.25 and than Internet, which increase overheads.

Frame relay overcome from the above drawbacks. It is a wide Area Network (WAN) with following features:

1. It operates a higher speed (1.5 mbps, and 44.376 mpbs)
2. Frame relay operates in only physical and data link layer. (so it can easily be used as backbone network to other protocols have network layer with less overheads)
3. Frame Relay allows bursty data. It means if at some point large amount of data is sent by someone than network should able to handle it properly.
4. Frame relay allow a Frame size of 9000 bytes, which can accommodates all LAN Frame sizes
5. It is less expensive than previous WANs, particularly with X.25.
6. It has error detection at data link layer only.
    - No Flow control, No error control, No re-transmission policy.
    - If frame is damaged, it is dropped.

Now can you answer why Frame Relay is faster than X.25? The answer is given above because it has fewer overheads of error control and flow control.

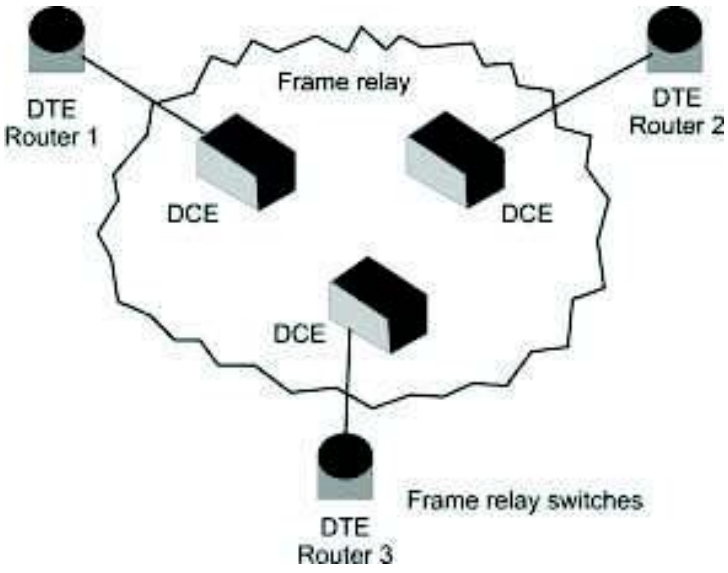


Figure 2: Network Architecture of Frame Relay

In Frame Relay each user/subscriber gets a leased line to a Frame Relay node, however the transmission paths are changed frequently and this is totally transparent to the users. Frame Relay is used for both voice and data transmission. Here, the data is packed in variable-size units called "frames" and necessary error-correction left for the end units. In Frame relay most of the services are based on permanent virtual circuit (PVC), which gives a feel good factor that they have a leased line connection at very low cost. As we discussed earlier, Frame relay operate in only physical and data link layer, so that it can easily be used as backbone-network to other protocols have network layer. Frame Relay layers are:

1. Physical Layer: The role of physical layer is similar with other architectures. However in frame relay no specific protocol is defined for physical layer to give

flexibility and better connectivity with other architectures. It supports any of the protocol recognized by ANSI. (American National Standard Institute)

2. Data Link Layer: Frame Relay uses simple protocol that does not support Flow Control, error Control, only it has error detection mechanism. However, the error correction is left for the end-user machines.

Format of Frame

Each Frame Relay Protocol data unit (PDU) consists of the following fields:

8	16	variable	16	8
Flag	**Address**	**Information ….**	**FCS**	**Flag**

Start and End Flag: Flag Field is 8 bit size, used to perform "synchronization" which indicates the beginning and end of the frame. Please refer to the unit 1 of block 1, where we have given similar example of start and end bits used for asynchronous communication. But what will happen if the flag bit pattern which we are using for end or start a communication occurs in between the flags. To avoid it we use bit stuffing and de-stuffing procedures at the source and destination respectively.

Frame Check Sequence (FCS): This is a 16 bits Field, which carries 16 bits of cyclic redundancy check (CRC) used at each switching node in the network for error detection.

Information: This field is a variable size field because user can send any data bits in this field. This is the actual data which network will pass on to receiver.

Address: This is a 16 bit or 2 bytes field having following fields inside of address:

DLCI	C/R	EA	DLCI	FECN	BECN	DE	EA
6	1	1	4	1	1	1	1

DLCI: Data link connection identifier used to identify virtual circuit in the Frame Relay.

DLCI field is of 10 bit size placed at two positions in the address field as given below:

i) The 1st DLCI is the 6 bits of first Bytes of address field

ii) The 2nd DLCI is the first 4 bits of second Bytes of address field

Command/Response (C/R): This is a 1 bit field. It is provided for upper layers to identify whether "a frame" is a command or a response. (This is not for Frame Relay)
Extended Address: This is 1 bit field, which inform the protocols about the address, such as:

- If, EA = 0 : Another address byte is to follow.(extended address can be 24 bit or 32 bit)
- If, EA = 1 : Current byte is the final address

FECN (Forward Explicit Congestion Notification): FECN bit can be set ("1") by any switch of the network to indicate that traffic is congested in the frames travelling towards the destination machine. This bit informs the destination that congestion has occurred, so destination should is ready for delay or packet loss.

BECN (Backward Explicit Congestion Notification): BECN bit also indicate congestion in a Network. BECN bit can be set ("1") by any switch of the network to indicate that traffic is congested in the frames travelling towards the source machine. This bit informs the sender machine that congestion had occurred in the network, hence slow-down the processing to prevent further delay or packet loss.

Discard Eligibility (DE): This is a 1 bit field, which indicates the priority of a frame. Sometime, switches have to discard frame (like congestion). If DE is set to "1", switch may discard the frame in problematic situation else it is very important frame and should not be discarded.

Frame Relay switching:

Here is an example of switching being done in the frame relay switch:

Table 1: Frame Relay Switching Data

Incoming		Outgoing	
Interface	DLCI	Interface	DLCI
2	78	10	37
2	121	12	147
2			

Interface 2 has received 2 pkts with DLCI values 78 and 121., Table maintained by switch show that a pkt arriving at interface 2 with DLCI = 78 should be souted to interface 10 with DLCI = 37. (Table tells the Frame Relay how to forward Frames from incoming interface to outgoing path)

Congestion Control in Frame Relay

The Frame Relay network is designed to handle busty data, whenever due to the high load and data bursts in some services, frame-relay networks provides some effective mechanisms to control the congestion. Remember, flow control is not performed in data link layer of Frame Relay so congestion avoidance mechanism as given below is used in Frame Relay:

Congestion avoidance is done through sharing information between sender/receiver nodes about backward/forward congestion notification in the network:

- Receiver can send BECN bit as a part of one of the ACK (acknowledgement). Any Frame Relay switch, send a special packet having BECN bit to the sender, so that sender may act accordingly.
- Through FECN bit, we can warn the destination that congestion has occurred, Destination can send ACK with BECN bit Set. Also, delay in sending ACK, may force the sender for deliberate delay in sending further data and consequently reduce congestion.

☞ Check Your Progress 1

1. Differentiate between virtual circuit and datagram.

……

……

……

……

2. Compare between SVC and PVC of X.25?

……

……

……

……

3. Write any four differences between X.25 and Frame Relay.

……

……

……

……

4. Explain the used of FECN and BECN in Frame Relay.

……

……

……

……

2.4 ATM NETWORK

Asynchronous Transfer Mode (ATM) is a form of data transmission that allows voice, video and data to be sent along the same network. In contrast to ATM, in the past, voice, video and data were transferred using separate networks. For example, voice was transmitted over the phone, video over cable networks and data over an internet work. ATM has its similarities with the frame relay, particularly in the term of data unit size, frame replay used a variable lengh data unit called frame. On the contrary, ATM used fixed data unit named as "cell", we can say ATM as Cell-Relay in analogy to frame relay.

ATM was emerged as a viable technology for effective transmission of voice, video and data. Some of its features are:

- ATM is a packet network like X.25, frame relay.
- ATM integrates all different types of traffic into one network.
- ATM supports multiplexing of multiple logical connections over single physical channel.
- ATM does not provide flow Control and error control at data link layer.
- ATM can serve as a LAN or WAN backbone without requiring any network replacement.
- ATM can be used in existing physical channels and networks. Because ATM was developed to have such a wide range of compatibility with existing networks, its implementation does not require replacement or over-building of telephone, data or cable networks. It is also compatible with wireless and satellite communications.

ATM Cell

As we had already discussed that ATM used a fixed size data unit called cell. As packet size is one of the key issues for protocol design, we would like to discuss the reasons for deciding the cell size. First let's assume a situation of using large packet size.

Large packets are better in a sense that they use less number of headers for data transfer. So, large packets may cause less overhead in a network. Another, important point is if a we are using a large size packet, than sometime the system has to wait till the packet is completely filled before sending any data. Remember the data sending requirement are not same at all time. Just to solve this problem, we can use variable size packet for different type of data. For example, Voice traffic can be sent in small packet and data traffic into large packet. But the variable size packet may increase additional Complexity such that variable packet size can leads to starvation problem for small packets.

The team of ATM designer had discarded the idea of both large packet and variable packets, and agreed for a fixed size data unit of 53 bytes (a 5-byte cell header and 48 bytes of data), which can achieve both higher data rate and less transmission delay. What was so special about '48 bytes'? Some people say that US telecommunication organizations wants 64 bytes Cell but the Europeans and Japanese telecommunication organizations want 32 bytes Cell. So as a compromise, 48 byte was decided.

5 Byte Header	48 bytes Data Unit

Figure 3: ATM Cell

We have various advantages of using fixed size small Cell, like it reduced queuing delay for a high priority cell. This concept simplifies the implementation of switching mechanism in hardware. The fixed cell size ensures that time-critical information such as voice or video is not adversely affected by long data frames or packets. Also, the cell header is organized for efficient switching, virtual-circuit identifiers and header error checks.

ATM cell has two formats for user to network interface and network to network interface as shown in the Figure 4:

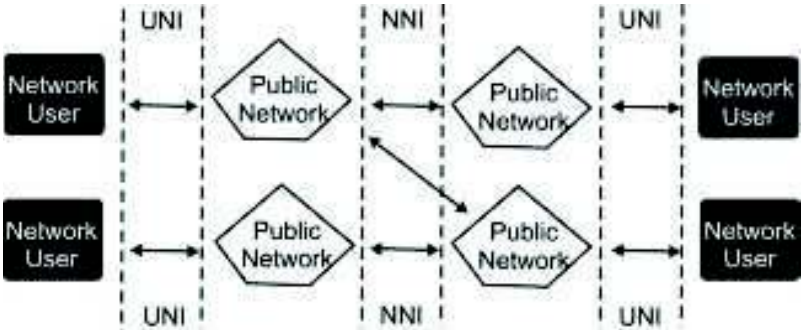


Figure 4: UNI and NNI of ATM

The Header Format

The structure of the header is different in UNI and NNI. In the network-network interface, the virtual path identifier field is expanded from 8 to 12 bits.

8	7	6	5	4	3	2	1
Generic Flow Control*				Virtual Path Identifier			
Virtual Path Identifier				Virtual Channel Identifier			
Virtual Channel Identifier							
Virtual Channel Identifier				Payload Type ID			CLP
Header Error Control							

Figure 3: User-network Interface

8	7	6	5	4	3	2	1
Virtual Path Identifier							
Virtual Path Identifier				Virtual Channel Identifier			
Virtual Channel Identifier							
Virtual Channel Identifier				Payload Type ID CLP			
Header Error Control							
INFORMATION PAYLOAD (48 Bytes)							

Figure 4: Network-network interface

Let's now look at the characteristics of each of the fields of the header format of an ATM cell.

Generic Flow Control (GFC)

The GFC field of the header is only defined across the UNI and does not appear in the NNI.

Function

- It controls the traffic flow across the UNI.

Virtual Path Identifier (VPI)

The VPI is an 8-bit field for the UNI and a 12-bit field for the NNI.

Function

- It constitutes a routing field for the network and is used to identify virtual paths. In an idle cell, the VPI is set to all 0's.
- Together with the Virtual Channel Identifier, the VPI provides a unique local identification for the transmission.

Virtual Channel Identifier (VCI)

It is a 16-bit field used to identify a virtual channel. For idle cells, the VCI is set to all 0's.

Function

- It functions as a service access point and it is used for routing to and from the end user.
- Together with the Virtual Path Identifier, the VCI provides a unique local identification for the transmission.

Payload Type Identifier (PTI)

The PTI field indicates the type of information in the information field. The value in each of the three bits of PTI indicate different conditions.

- Bit 1 is set to 1 to identify operation, administration, or maintenance cells (i.e. anything other than data cells).
- Bit 2 is set to 1 to indicate that congestion was experienced by a data cell in transmission and is only valid when bit 4 is set to 0.
- Bit 3 is used to convey information between end-users.

Cell Loss Priority (CLP)

The 1-bit CLP field is used for indication of the priority of the cell. It is used to provide guidance to the network in the event of congestion. When set to value 1, it indicates that the cell may be discarded within the network when congestion occurs. When the CLP value is set to 0, it indicates that the cell is of relatively high priority and should be discarded only in situations when no alternative is available.

Header Error Control (HEC)

Each ATM cell includes an 8-bit HEC that is calculated based on the remaining 32 bits of the header.

Function

- It detects all single-bit errors and some multiple-bit errors. As an ATM cell is received at a switch, the HEC of the cell is compared and all cells with HEC discrepancies (errors) are discarded. Cells with single-bit errors may be subject to error correction if supported or discarded. When a cell is passed through the switch and the VPI/VCI values are altered, the HEC is recalculated for the cell prior to being passed out to the port.

ATM Layers

ATM is a connection-invented protocol. ATM has a layered structure that is similar to the 7-layerd OSI model. However, ATM only addresses the functionality of the two lowest layers of the OSI model, i.e.

- The physical layer and
- The data link layer.

Apart from these two layers, all other layers of the OSI model are irrelevant in ATM, as these layers are only part of the encapsulated information portion of the cell which is not used by the ATM network. In ATM, three layers handle the functionality of the two lower OSI layers.

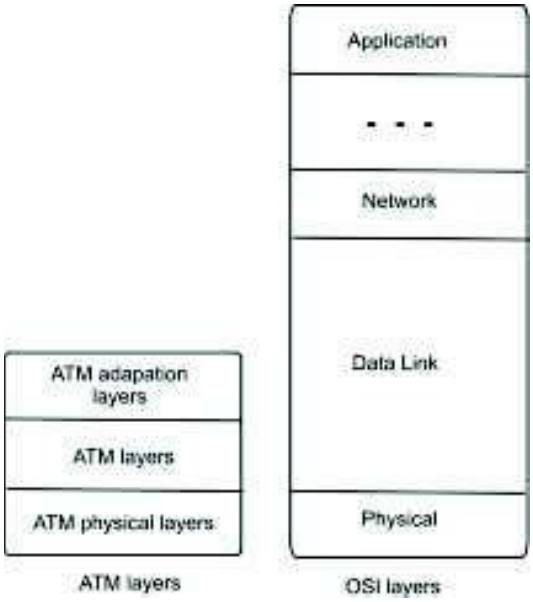


Figure 5: ATM and OSI Model

i) **Physical Layer:** The physical layer defines the specification of a transmission medium (copper, fiber optic, coaxial, HFC, wireless) and a signal encoding scheme and electrical to optical transformation. It provides convergence with physical transport protocols such as SONET as well as the mechanism for transforming the flow of cells into a flow of bits. The ATM form has left most of the specification for this level to the implementer.

ii) **The ATM Layer:** The ATM layer deals with cells and cell transport. It defines the layout of a cell and tells what the header fields mean. The size of a cell is 53 bytes (5 bytes of header and 48 bytes of payload). Because each cell is the same size and all are relatively small, delay and other problems with multiplexing different sized packets are avoided.

It also deals with establishment and release of virtual circuits. Congestion control is also located here. It resembles the network layer of the OSI model as it has got the characteristics of the network layer protocol of OSI model like Routing, Switching, End to end virtual circuit set up and Traffic management.

Switches in ATM provide both switching and multiplexing. A Cell format of ATM Layer are distinguished as,UNI (User Network Interface) and NNI (Network-Network Interface)

In both cases the cell consists of a 5 byte header followed by a 48 bytes pay-load but the two headers are slightly different.

iii) **ATM Adaptation Layer:** The ATM Adaptation Layer (AAL) maps the higher-level data into ATM cells to be transported over the ATM network, i.e. this layer segments the data and adds appropriate error control information as necessary. It is dependent on the type of services (voice, data etc.) being transported by the higher layer.

ATM is connection oriented and allows the user to specify the resources required on a per-connection basis (per SVC) dynamically. There are the five classes of service defined for ATM (as per ATM Forum UNI 4.0 specification).

ATM Service Classes

Service Class	Quality of Service Parameter
Constant bit rate (CBR)	CBR class is used for emulating circuit switching. The cell rate is constant with time. CBR applications are sensitive to cell-delay variation. Examples of applications that can use CBR are telephone traffic (i.e. nx64 kbps), video conferencing and television.
Variable bit rate–non-real time (VBR–NRT)	VBR-NRT class allows users to send traffic at a rate that varies with time depending on the availability of user information. Statistical multiplexing is provided to make optimum use of network resources. Multimedia e-mail is an example of VBR–NRT.
Variable bit rate–real time (VBR–RT)	This class is similar to VBR–NRT but is designed for applications that are sensitive to cell-delay variation. Examples for real-time VBR are voice with speech activity detection (SAD) and interactive compressed video.
Available bit rate (ABR)	ABR class provides rate-based flow control and is aimed at data traffic such as file transfer and e-mail. Although the standard does not require the cell transfer delay and cell-loss ratio to be guaranteed or minimized, it is desirable for switches to minimize delay and loss as much as possible. Depending upon the state of congestion in the network, the source is required to control its rate. The users are allowed to declare a minimum cell rate, which is guaranteed to the connection by the network.
Unspecified bit rate (UBR)	UBR class is widely used today for TCP/IP.

The ATM Forum has identified certain technical parameters to be associated with a connection.

Depending on the type of data, several types of AAL layers have been defined. However, no AAL is restricted to a specific data class or type; all types of data could conceivably be handled by any of the AALs. The various AAL protocols defined are:

1. AAL 1
2. AAL 2
3. AAL 3/4 (layer 3 and 4 were merged to avoid function overlapping)
4. AAL 5

Each layer of ATM is further divided into two sublayers

- SAR (Segmentation and Reassembly)
- CS (Convergence Sublayer).

Segmentation & Reassembly: This is the lower part of the AAL. The SAR sublayer breaks packets up into cells on the transmission side and puts them back together again at the destination. It can add headers and trailers to the data units given to it by the CS to form payloads. It is basically concerned with cells.

Convergence Sublayer: The CS sublayer makes it possible to have ATM systems offer different kind of services to different applications. The CS is responsible for accepting bit streams or arbitrary length messages from the application and breaking them into units of 44 or 48 bytes for transmission.

Working of ATM

When a user sends data over the ATM network, the higher-level data unit is passed down to the Convergence Sublayer of the AAL Layer, which prepares the data for the ATM Layer according to the designated AAL protocol. The data is then passed down to the Segmentation and Reassembly Sublayer of the AAL Layer, which divides the data unit into appropriately sized segments.

These segments are then passed down to the ATM Layer, which defines an appropriate cell header for each segment and encapsulates the header and payload segment into a 53-byte ATM cell. The cells are then passed down to the Physical Layer, which streams the cells at an appropriate pace for the transmission medium being used, adding empty cells as needed.
ATM circuit connections are of two types:

1. Virtual Paths and

2. Virtual Channels.

A virtual channel is a unidirectional pipe made up from the concatenation of a sequence of connection elements. A virtual path **consists of a set of these virtual channels.** Each virtual channel and virtual path has an identifier associated with it. Virtual path is identified by Virtual Path Identifiers (VPI) and a virtual channel is identified by a Virtual Channel Identifier (VCI). All channels within a single path must have distinct channel identifiers but may have the same channel identifier as channels in different virtual paths.

An individual channel can therefore be uniquely identified by its virtual channel and virtual path number. Cell sequence is maintained through a virtual channel connection.
ATM connections can be categorised into two types:

i) ***Point-to-point connections:*** These are the connections which connect two ATM end-systems. Such connections can be unidirectional or bidirectional.

ii) ***Point-to-multipoint connections:*** These are the connections which connects a single source end-system known as the root node) to multiple destination end-systems (known as leaves).

The basic operation of an ATM switch is very simple to understand.

1. The ATM switch receives a cell across a link on a known VCI or VPI value.

2. The ATM switch looks up the connection value in a local translation table to determine the outgoing port (or ports) of the connection and the new VPI/VCI value of the connection on that link.

3. The ATM switch then retransmits the cell on that outgoing link with the appropriate connection identifiers.

The manner in which the local translation tables are set up determine the two fundamental types of ATM connections:

- ***Permanent Virtual Connections (PVC):*** A PVC is a connection set up by some external mechanism, typically network management, in which a set of switches between an ATM source and destination ATM system are programmed with the appropriate VPI/VCI values.
- ***Switched Virtual Connections (SVC):*** An SVC is a connection that is set up automatically through a signalling protocol. SVCs do not require the manual interaction needed to set up PVCs and as such, are likely to be much more widely used.

Traffic Control

An ATM network needs efficient Traffic Control mechanisms to allocate network resources in such a way as to separate traffic flows according to the various service classes and to cope with potential errors within the network at any time.
Network Resource Management: Network Resource management deals with allocation of network resources in such a way that traffic is separated on the basis of the service characteristics. A tool of network resource management which can be used for Traffic Control is the **virtual path technique**. A virtual path connection (VPC) groups several virtual channel connections (VCC)s together such that only the collective traffic of an entire virtual path has to be handled. In this type of setup, priority control can be supported by reaggregating traffic types requiring different qualities of service through virtual paths. Messages for the operation of traffic control can be more easily distributed, a single message referring to all the virtual channels within a virtual path will do.

Connection Admission Control: Connection Admission Control is the set of actions taken by the network in protecting itself from excessive input loads. When a user requests a new virtual path connection or virtual channel connection, the user needs to specify the traffic characteristics in both directions for that connection. The network establishes such a connection only if sufficient network resources are available to establish the end-to-end connection with the required quality of service. The agreed quality of service for any of the existing channels must not be affected by the new connection.

Usage Parameter Control and Network Parameter Control: After a connection is accepted by the Connection Admission Control function, the UPC function of network monitors the connection to check whether the traffic conforms to the traffic contract.
The main purpose of UPC/NPC is to protect the network resources from an overload on one connection that would affect the quality of service of other already established connections. Usage Parameter Control (UPC) and Network Parameter Control (NPC) do the same job at different interfaces. The UPC function is performed at the UNI, while the NPC function is performed at the NNI.

Functions performed by the Usage parameter control include:

- Checking the validity of VPI/VCI values.

- Monitoring the traffic volume entering the network from all active VP and VC connections to ensure that the agreed parameters are not violated.
- Monitoring the total volume of the accepted traffic on the access link.
- Detecting violations of contracted (agreed) parameter values and taking appropriate actions.

Priority Control: Priority control is an important function as its main objective is to discard lower priority cells in order to protect the performance of higher-priority cells.

Congestion Control: Congestion is a state of network wherein the network resources are overloaded. This situation indicates that the network is not able to guarantee the negotiated quality of service to the established connections and to the new connection requests. ATM Congestion control refers to the measures taken by the network to minimize the intensity, spread and duration of network congestion.

1. As a high-bandwidth medium with low delay and the capability to be switched or routed to a specific destination, ATM provides a uniformity that meets the needs of the telephone, cable television, video and data industries. This universal compatibility makes it possible to interconnect the networks — something that is not currently possible because of the various transmission standards used by each industry.

2. One of the key advantages of ATM is its ability to transmit video without creating a jittery picture or losing the synchronisation of the sound and picture. This is possible due to proper resource allocation and admission control.

3. ATM also provides dynamic bandwidth for bursty traffic.

4. Telephone networks connect each telephone to every other telephone using a dedicated path, but carry narrow bandwidth signals. Cable networks carry broadband signals, but only connect subscribers to centralised locations. To build a network that would provide a dedicated connection between sender and receiver for broadband communications would be prohibitively expensive. For this reason, ATM seems to be the best hope since it can use existing networks to deliver simple voice and data as well as complex and time-sensitive television signals. ATM can also handle bi-directional communications easily.

5. Unlike packet switching, ATM is designed for high-performance multimedia networking.

☞ Check Your Progress 2

1. What are VPI and VCI in ATM network? Write the importance of each.

   ……

   ……

   ……

   ……

2. Explain how ATM layers are divided into sub-layers.

   ……

   ……

   ……

2.5 IPv4 AND IPv6 OVERVIEW

The primary goal of the Internet is to provide an abstract view of the complexities involved in it. Internet must appear as single network of computers. At the same time network administrators or users must be free to choose hardware or various internetworking technologies like Ethernet, Token ring etc. Different networking technologies have different physical addressing mechanisms. Therefore, identifying a computer on Internet is a challenge. To have uniform addressing for computers over the Internet, IP software defines an IP address which is a logical address. Now, when a computer wants to communicate to another computer on the Internet, it can use logical address and is not bothered with the physical address of the destination and hence the format and size of data packet.

2.5.1 Classes of IP Address

Internet addresses are 32 bits long, written as four bytes separated by periods (full stops). They can range from 0.0. 0.0 to 223. 255. 255. 255. It's worth noting that IP addresses are stored in big-endian format, with the most significant byte first, read left to right. This contrasts with the little-endian format used on Intel- based systems for storing 32- bit numbers. This minor point can cause a lot of trouble for PC programmers and others working with raw IP data if they forget.
IP addresses comprise two parts, the network ID and the host ID. An IP address can identify a network (if the host part is all zero) or an individual host. The dividing line between the network ID and the host ID is not constant. Instead, IP addresses are split into five classes, which allow for a small number of very large networks, a medium number of medium- sized networks and a large number of small networks. The classes of IP address are briefly explained below, the structure of these classes are also shown in.

IP Address Class	High Order Bit(s)	Format	Range	No. of Network Bits	No. of Host Bits	Max. Hosts	Purpose
A	0	N.H.H.H	1.0.0.0 to 126.0.0.0	7	24	2^{24}-2	Few large organisations
B	1,0	N.N.H.H	128.1.0.0 to 191.254.0.0	14	16	2^{16}-2	Medium-size organisations
C	1,1,0	N.N.N.H	192.0.1.0 to 223.255.254.0	21	8	2^{8}-2	Relatively small organisations
D	1,1,1,0	N/A	224.0.0.0 to 239.255.255.255	N/A	N/A	N/A	Multicast groups (RFC 1112)
E	1,1,1,1	N/A	240.0.0.0 to 254.255.255.255	N/A	N/A	N/A	Future Use (Experimental)

Figure 4: Classes of IPv4 address

IP follows these rules to determine the address class:

- **Class A:** If the first bit of an IP address is 0, it is the address of a class A network. The first bit of a class A address identifies the address class. The next 7 bits identify the network, and the last 24 bits identify the host. There are fewer than 128 classes a network numbers, but each class A network can be composed of millions of hosts.

- **Class B:** If the first 2 bits of the address are 1 0, it is a class B network address. The first 2 bits identify class; the next 14 bits identify the network, and the last 16 bits identify the host. There are thousands of class B network numbers and each class B network can contain thousands of hosts.
- **Class C:** If the first 3 bits of the address are 1 1 0, it is a class C network address. In a class C address, the first 3 bits are class identifiers; the next 21 bits are the network address, and the last 8 bits identify the host. There are millions of class C network numbers, but each class C network is composed of fewer than 254 hosts.
- **Class D:** If the first 4 bits of the address are 1 1 1 0, it is a multicast address. These addresses are sometimes called class D addresses, but they don't really refer to specific networks. Multicast addresses are used to address groups of computers all at one time. Multicast addresses identify a group of computers that share a common application, such as a video conference, as opposed to a group of computers that share a common network.
- **Class E:** If the first four bits of the address are 1 1 1 1, it is a special reserved address. These addresses are called class E addresses, but they don't really refer to specific networks. No numbers are currently assigned in this range.

IP addresses are usually written as four decimal numbers separated by dots (periods). Each of the four numbers is in the range 0-255 (the decimal values possible for a single byte). Because the bits that identify class are contiguous with the network bits of the address, we can lump them together and look at the address as composed of full bytes of network address and full bytes of host address. If the value of the first byte is:

- Less than 128, the address is class A; the first byte is the network number, and the next three bytes are the host address.
- From 128 to 191, the address is class B; the first two bytes identify the network, and the last two bytes identify the host.
- From 192 to 223, the address is class C; the first three bytes are the network address, and the last byte is the host number.
- From 224 to 239, the address is multicast. There is no network part. The entire address identifies a specific multicast group.
- Greater than 239, the address is reserved.

To learn further about IP address and CIDR you can see the course material of BCS-061: TCP/IP programme which you will study in your next semester.

IPv6 Overview

With the advancement in the technologies, mobile-handheld devices and emerging applications, it is quite evident that soon the IP addresses provided by IPv4 are not sufficient. In the recent future we can operate and use various smart deices (like TV, Fridges, cameras, ACs, phone, mobiles, etc). Each of such devices will require a unique IP address, which will increase the demand of IP addresses exponentially.

The number of IPv4 unique addresses is not that large in relation to the current rate of expansion of the Internet. Consequently, a new addressing system has been devised which is a part of Internet Protocol version 6 (IPv6), which uses 128-bit addresses, means total addresses will be 2^{128} . IPv4 uses 32-bit addresses, means total addresses will be 2^{32} around 4,294,967,296 unique addresses. IPv6 has almost 7.9×10^{28} times more addresses than IPv4.

It is possible that IPv6 would not be used or implemented completely in the coming couple of years. This IPv6 (Internet Protocol version 6) is a revision of the earlier

Internet Protocol (IP) version 4. As you know IPv4 address is 32 bit and divided into four octets separated by dot for example 192.186.12.10, on the other hand IPv6 addresses are consist of eight groups of four hexadecimal digits separated by colons, for example 2001:0db8:85a3:0042:0000:8a2e:0370:7334. IPv6 is designed to swap the existing IPv4, which is the main communications protocol for most Internet traffic as of today. IPv6 was developed to deal with the long-anticipated problem of IPv4 running out of addresses, some of the reasons and need for implementing IPv6 are following:

- The short term solutions like sub-netting, classless addressing cannot fulfill the massive future demand of address space.
- The internet must accommodate the real-time audio and video transmission with best quality of services.
- Internet protocol must provide the necessary security implementation for some applications.
- There is a need of multicasting in current IPv4, where the transmission of a packet to multiple destinations can be sent in a single send operation.
- IPv4 need a major revision in various issues like privacy, mobility, routing, QoS (quality of services), extendibility and addressing.

Address format

As we discussed before, IPv6 addresses are consist of eight groups of four hexadecimal digits separated by colons (:), for example 2001:0db8:85a3:0042:0000:8a2e:0370:7334. Lets see the bit composition of an IPv6 address, as we know each hexadecimal should be of 4 bits each, in a group we have four hexadecimal bits hence a group has 16 bits. Now we have 8 groups so 16 multiple with 8 is 128 bits. Any IPv6 address may be reduced and interpreted using the following rules:

- First thing is leading zeroes from the groups of hexadecimal digits can be removed, similar to the currency where leading zeros are nothing. For example, convert the group 0036 to 36.
- Always remember that hexadecimal digits in the groups are not case-sensitive just like the c programming; e.g., the groups 08DB and 08db are same.
- Next you may merge successive groups of one or more zeroes, using a double colon (::) to indicate the omitted groups. But, double colon may only be used once in any given address

The initial process and few implementation of IPv6 have been done, but still the transition process of replacing from IPv4 with IPv6 will continue for couple of years. We must consider that at present IPv4 is backbone of Internet, replacing it, is not an easy process. Definitely it will be done slow transition from one stage to another. Following are the approaches being used for replacing from IPv4 with IPv6:

1. Protocol Translation
2. Dual IP Stack
3. Tunneling

Protocol Translation

Like any other protocol both IPv4 and IPv6 are using their own headers. There are different kinds of IPv4 to IPv6 translators possible

- IP header translator: At the IP layer, we replaces IPv4 header by IPv6 header through trnslation. IP header translator is similar to NAT, Network Address Translator.
- TCP relay: At the TCP layer, we can transmit IPv4 TCP connection to IPv6 TCP connection, and vice versa, regardless of the application protocol used over TCP.
- Application gateway: In this technology we works in application protocol layer (such as FTP, HTTP), and uses application protocol-specific mechanism for protocol translation.

Protocol translation may interfere with an objective of end-to-end transparency in network communications. Also, the use of protocol translators cause problems with NAT and limit the use of addressing.

Dual IP stack

In dual IP stack implementation, we will use both IPv4 and IPv6 protocol stacks together at the Internet layer. Dual IP stack seems to be a fair solution for IPv6 implementation, as it avoids many complexities and overheads. Dual-stacked hosts running on a dual-stack network allow applications to migrate one at a time from IPv4 to IPv6. Applications and devices (that are not upgraded according to IPv6 stack) can coexist with upgraded IPv6 applications on the same network system.
In dual stack we will need the devices having capability of handling both IPv4 and IPv6 can use any IPv4 or IPv6, depending on the requirement. Dual stack approach will be costly and in some cases network devices may not support this mplementation.

Tunneling

In tunneling, we mean to encapsulate the packets of one protocol into the packets of another protocol. Something likes keeping one letter envelope into another envelope. Assume a situation as shown in the figure when two isolated IPv6 networks need to communicate over an IPv4 network, dual-stack routers at the network edges can be used to set up a tunnel which encapsulates the IPv6 packets within IPv4 and communicate (between two IPv6 networks) without updating the inter-mediatory IPv4 network infrastructure.

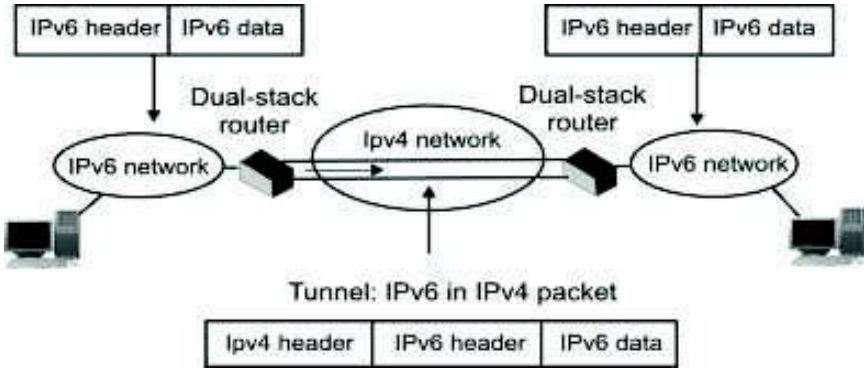


Figure 5: Tunneling Mechanism for IPv6

☞ Check Your Progress 3

1. Discuss the need of IPv6.

……

……

……

2. Explain the dual stack approach for IPv6 implementation.

……

……

……

……

1.5 SUMMARY

Network architecture is a complete design of a communications network. Primarily we can say that it is a framework for the specification of a network's physical components, their functional organization and configuration. In this unit you have learnt about X.25, Frame Relay and ATM Architectures. X.25 is an old standard protocol suite for packet based wide area network. The old networks mainly tele-communications companies and ATM's (automated teller machines) were following X.25 protocols for packet switching based network. Frame Relay is a virtual-circuit based WAN that was designed to provide more efficient transmission scheme than X.25. It provides connection oriented services at reasonable speed and low cost. Asynchronous Transfer Mode (ATM) is a form of data transmission that allows voice, video and data to be sent along the same network. In contrast to ATM, in the past, voice, video and data were transferred using separate networks. In this unit you have also studied about ISP and different address schemes of TCP/IP protocols suits. Now you know that the number of IPv4 unique addresses is not that large in relation to the current rate of expansion of the Internet. Consequently, a new addressing system has been devised which is a part of Internet Protocol version 6 (IPv6), which uses 128-bit addresses, means total addresses will be 2^{128} . In this unit you have also learnt about different approaches, which can be used for replacing from IPv4 with IPv6.

1.6 SOLUTIONS/ANSWERS

☞ **Check Your Progress 1**

1. Datagram allows for dynamic handling of congestion and no call setup is necessary. Virtual circuits allow for sequencing, error and flow control.

2. X.25 is connection oriented architecture and support switched virtual circuits (SVC) and permanent virtual circuits (PVC). Switched virtual circuits are established on the need basis. SVC is established when a call is made and broken down after the call is completed. On the other hand, permanent virtual circuits are almost leased kind of connections, which provide a dedicated connection between DTE's.

3. Following are the differences between X.25 and Frame Relay:

   - Frame Relay operates a higher speed
   - Frame relay operate in only physical and data link layer. (so it can easily be used as backbone network to other protocols have network layer with less overheads)
   - Frame Relay allows bursty data. It means if at some point large amount of data is sent by someone than network should able to handle it properly.
   - Frame relay allow a Frame size of 9000 bytes, which can accommodates all LAN Frame sizes.
   - It is less expensive than X.25.

- It has error detection at data link layer only.

4. FECN and BECN are used in Frame Relay mainly for congestion control

FECN (Forward Explicit Congestion Notification): FECN bit can be set ("1") by any switch of the network to indicate that traffic is congested in the frames travelling towards the destination machine. This bit informs the destination that congestion has occurred, so destination should is ready for delay or packet loss.

BECN (Backward Explicit Congestion Notification): BECN bit also indicate congestion in a Network. BECN bit can be set ("1") by any switch of the network to indicate that traffic is congested in the frames travelling towards the source machine. This bit informs the sender machine that congestion had occurred in the network, hence slow-down the processing to prevent further delay or packet loss.

☞ Check Your Progress 2

1. Virtual Path Identifier (VPI) is an 8-bit field for the UNI and a 12-bit field for the NNI. It constitutes a routing field for the network and is used to identify virtual paths. In an idle cell, the VPI is set to all 0's. Together with the Virtual Channel Identifier, the VPI provides a unique local identification for the transmission.

   Virtual Channel Identifier (VCI) is a 16-bit field used to identify a virtual channel. For idle cells, the VCI is set to all 0's. It functions as a service access point and it is used for routing to and from the end user. **Together with the Virtual Path Identifier, the VCI provides a unique local identification for the transmission.**

2. Each layer of ATM is further divided into two sublayers SAR (Segmentation and Reassembly) and CS (Convergence Sublayer).

   Segmentation & Reassembly: This is the lower part of the AAL. The SAR sublayer breaks packets up into cells on the transmission side and puts them back together again at the destination. It can add headers and trailers to the data units given to it by the CS to form payloads. It is basically concerned with cells.

   Convergence Sublayer: The CS sublayer makes it possible to have ATM systems offer different kind of services to different applications. The CS is responsible for accepting bit streams or arbitrary length messages from the application and breaking them into units of 44 or 48 bytes for transmission.

☞ Check Your Progress 3

1. With the advancement in the technologies, mobile-handheld devices and emerging applications, it is quite evident that soon the IP addresses provided by IPv4 are not sufficient. In the recent future we can operate and use various smart deices (like TV, Fridges, cameras, ACs, phone, mobiles, etc). Each of such devices will require a unique IP address, which will increase the demand of IP addresses exponentially.

   The number of IPv4 unique addresses is not that large in relation to the current rate of expansion of the Internet. Consequently, a new addressing system has been devised which is a part of Internet Protocol version 6 (IPv6), which uses 128-bit addresses, means total addresses will be 2^{128} . IPv4 uses 32-bit

addresses, means total addresses will be 2^{32} around 4,294,967,296 unique addresses. IPv6 has almost $7.9x10^{28}$ times more addresses than IPv4.

2. In dual IP stack implementation, we will use both IPv4 and IPv6 protocol stacks together at the Internet layer. Dual IP stack seems to be a fair solution for IPv6 implementation, as it avoids many complexities and overheads. Dual-stacked hosts running on a dual-stack network allow applications to migrate one at a time from IPv4 to IPv6. Applications and devices (that are not upgraded according to IPv6 stack) can coexist with upgraded IPv6 applications on the same network system.

   In dual stack we will need the devices having capability of handling both IPv4 and IPv6 can use any IPv4 or IPv6, depending on the requirement. Dual stack approach will be costly and in some cases network devices may not support this implementation.

UNIT 3 INTRODUCTION TO WIRELESS AND MOBILE NETWORKS

Structure **Page Nos.**

3.0 Introduction Systems 48
3.0.1 Wired Communication System
3.0.2 Wireless Communication System
3.1 Objectives 50
3.2 Wireless Communication Systems 51
3.2.1 Paging System
3.2.2 Cordless Telephone System
3.2.3 Cellular Mobile System
3.2.4 Bluetooth
3.2.5 Wireless Local Area Network (WLAN)
3.3 Wireless Generations 54
3.3.1 First Generation (1G) –
3.3.2 Second Generation (2G) –
3.3.3 Evolution To Mid of Second Generation (2.5G) –
3.3.4 Third Generation (3G) –
3.4 Introduction to Cellular Mobile Systems – GSM 56
3.5 Code Division Multiple Access (CDMA) 59
3.6 Cellular System Design Fundamental 60
3.6.1 Frequency Reuse
3.6.2 Hand-Off and Signal Strength
3.6.3 Interference
3.6.4 Coverage and Capacity Improvements
3.7 Summary 64
3.8 Suggested Reading 64
3.9 Solutions / Answers 65

3.0 INTRODUCTION SYSTEMS

Communication Systems enables two or more person to communicate each other irrespective of their geographic location distance. This is only because of communication technology which provide such seamless service to their customers. Therefore, this communication service has tremendously grown among people in the past few years as it has eliminated the obstacle caused by geographic distance between people. Now there are modes of this communication system through which this service can be provided. The first one is the wired communication system and the other one is wireless communication system. Please note that in this chapter, the terms "wireless" and "cellular" are used interchangeably. Both are used in view of mobility. Moreover, in all diagrams, a dotted line represents a wireless link whereas a solid line represents a wired communication link.

3.1.1 Wired Communication System

The **Wired Communication System** depends solely on wires as e all the users (or communicators) are connected to each other through wires. The typical example (Figure 1) of such wired communication system can be a Local Area Network (LAN) where people are communicating and linked with each other through wires. Wire can be a cable wire, optical wire or any other type of wire.

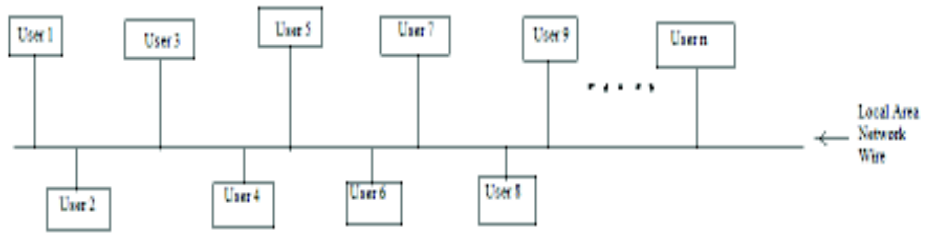


Figure 1: Wired Communication Systems

Following are the **pros of Wired Communication System**:-

- Wired technology has good data transfer speed.
- It is more secure as compared to wireless technology as the data transfer takes place in wired medium.

The **cons of Wired Communication System** are:-

- The position of a user becomes fixed due to attachment with a wire.
- As some wires are underground, it is difficult to perform their maintenance as one have to dig the ground to repair it.

3.1.2 Wireless Communication System

In **Wireless Communication System**, communication medium is the space and not the wires. Each user communicates with each other through a wireless link. This link can be a radio link which typically works on Radio Frequency (RF) concept. The typical example of a wireless communication system can be a mobile system (Figure 2) like Global System for Mobile Communication (GSM) OR Code Division Multiple Access (CDMA) in which the communication signals travel through air medium.

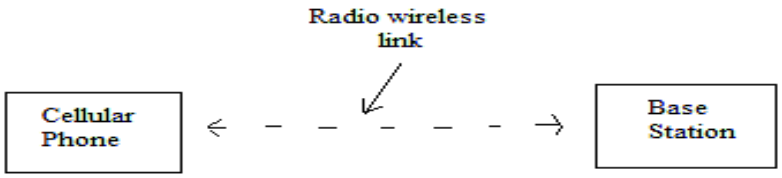


Figure 2: Wireless Communication System

Following are the **pros of Wireless Communication System**:-

- Wireless network provides mobility to its users.
- Users are free from wires and this reduces their effort to manage and maintain them.

The **cons of Wireless Communication System** are:-

- The technology is less secure as the whole communication takes place in an open wireless medium
- The data transfer speed is low as compared to wired technology.
- The quality of network and signals depend on the weather condition as rainy season deviate the signals in air which degrades the performance of a wireless network.

Modes of Wireless Communication System in small distance

Despite of cons of wireless network, this technology usage is increasing day by day and reaching to every people due to its advantage of getting people free form wires and providing them mobility. Therefore, our focus we will be on Wireless Technology. Now, we will discuss the two modes in which a wireless technology works.

- **Access Point (AP) Wireless Communication System :-**

  The first mode is called **Access Point (AP) Wireless Communication System.** In this Access Point (AP) wireless communication system (Figure 3), the data transfer takes place from source (User 1) to destination (User 2) through an Access Point. This access point can be a modem or a switch etc. This access point decides the path for data transfer to follow, data transfer speed, calculates the shortest path etc from source (User 1) to destination (User 2). No two users or more users can communicate directly or without an access point in this mode. The example of Access Point (AP) wireless technology is a wireless LAN network where r a modem decides the path and transfers the data form User 1 to User 2.

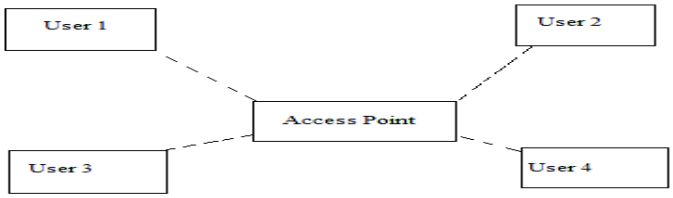


Figure 3: Access Point (AP) Wreless Communication System

- **Ad-hoc wireless Communication System :-**

  The second mode is Ad-hoc wireless communication system mode. In this mode (Figure 4), the data transfer takes place directly from source (User 1) to destination (User 2). There is no need of access point in this type of mode. Every user in this network mode communicates directly with each other. Such types of network are temporary and its establishment is very quick as compared to access point mode. These networks are successful where there is a requirement of temporary network only for few days.

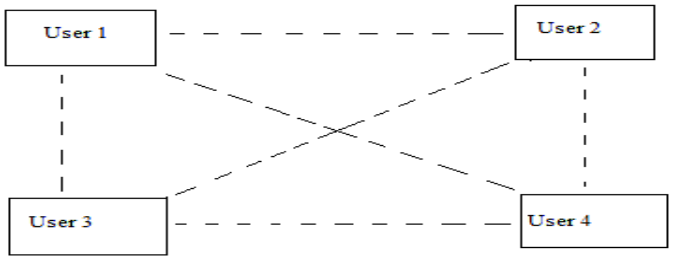


Figure 4: Ad-hoc Wireless Communication System

3.1 OBJECTIVES

After going through this unit you will be able to:

- define the wired and wireless communication systems;
- discuss the various wireless communication systems;

- define the wireless generations;
- define the Global System Mobile (GSM);
- define Code Division Multiple Access (CDMA); and
- define cellular system design fundamental.

3.2 WIRELESS COMMUNICATION SYSTEMS

The technology has grown tremendously and the consequence of which is modern wireless communication that has helped in eradicating the disadvantages of the typical wireless communication systems such as paging system and cordless telephone system. The examples of modern wireless communication systems are Cellular Mobile System, Bluetooth, and Wireless Local Area Network (WLAN). Below given are the few examples of Wireless Communication Systems which we will discuss in brief:-

3.2.1 Paging System

Paging systems are the systems which broadcasts the messages to its user for performing any action. Such message can be a service message in which a user can subscribe to a missed call alert service, caller tune service, internet service or any other such service. This message is broadcast to the users in a service area using same base stations. Coverage of a paging system can be of a range of 2 to 5 km or it can cover a wide area using wide area paging systems.

3.2.2 Cordless Telephone System

Cordless Telephone system consists of a landline telephone which is a fixed port (Figure 5). This landline is connected to the telephone exchange called Public Switched Telephone Network (PSTN). The landline telephone has a wireless (or cordless) handset which is connected to the landline telephone through a radio link. Therefore, through this cordless system, the user has the freedom to move while on a call. But this has a range or distance limitation which is around few tens of meters only.

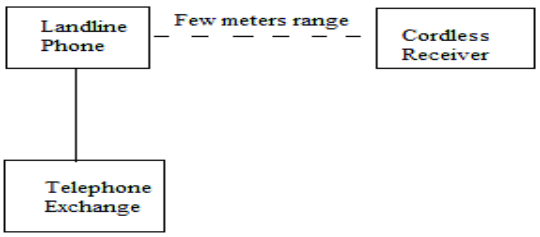


Figure 5: Cordless Telephone System

3.2.3 Cellular Mobile System

Cellular Mobile Systems eradicates the distance limitation imposed by a cordless telephone system. In this mobile system (Figure 6), a user can easily move from one place to another while on a call without getting disconnected from call. The user is constantly connected to the called user through radio links. As the user passes from one area (or cell) to another area (or another cell), the Base Station Controller (BSC) of one cell informs Base Station Controller (BSC) of other cell about the call transfer. Every cell has at least one BSC. Further all these BSC are connected to Mobile Switching Center (MSC). And finally all MSC are connected to PSTN.

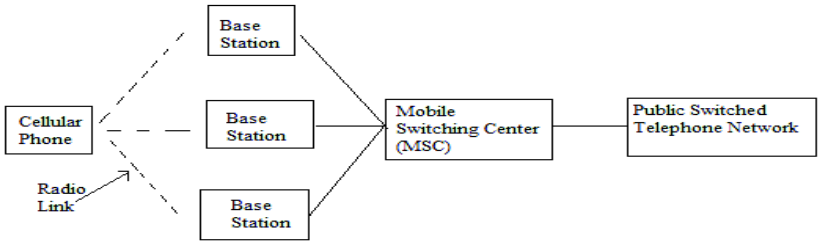


Figure 6: Cellular Mobile System

3.2.4 Bluetooth

As discussed above, Bluetooth works on ad-hoc mode in which the network is formed quickly and is of temporary basis. Bluetooth technology is created by a telecom company called Ericsson in 1994. It was developed in order to connect two devices without wires. The range in which Bluetooth technology works is of 10 meters (or 30 feet approx) only. The name "Bluetooth" is after tenth-century king Harald I of Denmark and parts of Norway who united Danish tribes into a single kingdom. The implication is that Bluetooth does the same with communications protocols, uniting them into one universal standard.

Bluetooth works on 2.4 GHz ISM band (Industrial, Scientific and Medical band) which divides the data into parts and sends it on up to 79 bands. It uses Frequency Hopping Spread Spectrum (FHSS) with Time Division Duplexing (TDD) technique at the rate of 1600 hops/sec. Moreover, the modulation technique employed is Guassian Frequency Shift Keying (GFSK) which was the only available modulation technique at the time of Bluetooth. Data rate is around 128 Mbps (Mega Bits per Second) and can support up to 8 devices simultaneously in Master-Slave mode. Bluetooth has versions started from version 1.0 to version 4.0.

3.2.5 Wireless Local Area Network (WLAN)

WLAN is a local area network but the end point at which the user gets the service is a wireless end. As you can see in below Figure 7, the user is connected to Access Point through a wireless link. The Access point is further connected to a LAN line which is wired. Therefore, in WLAN, only the last end is wireless and rest is wired network.

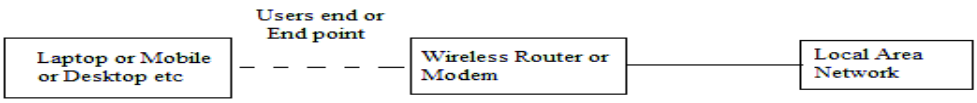


Figure 7: Wireless Local Area Network

WALN technology works on IEEE 802.11 standard. Components of 802.11 are Basic Service Set (BSS), Extended Service Set (ESS), Access Point (AP) and Distribution Systems (DS). We will now discuss these components in brief.
Basic Service Set (BSS) – BSS contains one or more mobile user (Figure 8 & 9). BSS can work in two modes. One is independent mode in which all users are connected to each other directly. The other mode is infrastructure mode in which all users communicate through an Access Point (AP).

Access Point – An AP can be a modem, router or a switch through which users communicate with each other. When a network employs this component, that network is called infrastructure mode. All the data passes through this AP.

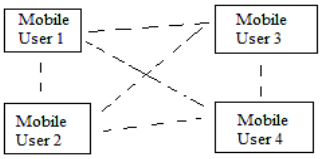


Figure 8: BSS Without Acess Point (AP)

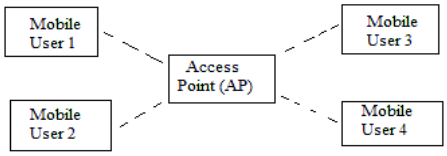


Figure 9: BSS With Acess Point(AP)

Extended Service Set (ESS) - All separate BSS (either in independent mode or in infrastructure mode) when connected to each other is called an Extended Service Set (Figure 10).

Distribution System (DS) – Distribution system connects AP of different ESS. This increases network coverage as all the users of different BSS will be connected with each other through DS (Figure 10). All the links connecting APs to DS can be wireless or wired.

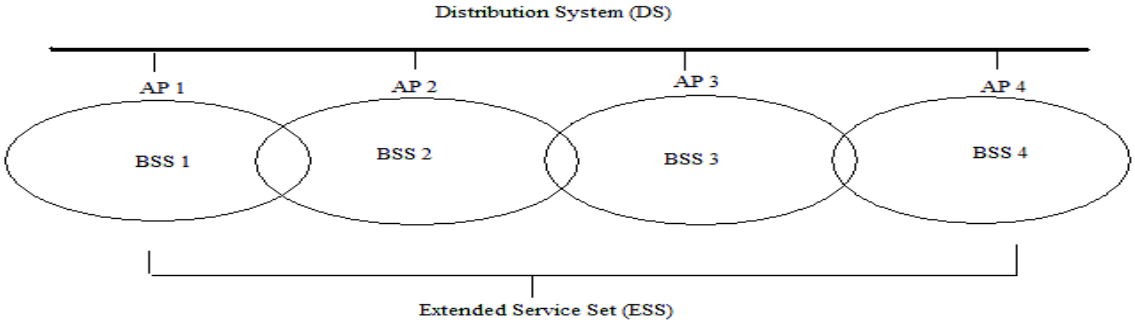


Figure 10: Extended Service Set (ESS) and Distribution Systems(DS)

☞ Check Your Progress 1

1. **State True or False**

   a) Wired communication systems are less secure than wireless communication system. ☐

   b) Wireless communication systems are easy to set up. ☐

   c) Data transfer speed is less in wireless communication system. ☐

   d) Ad-hoc wireless communication system uses an access point for users to get connected to each other. ☐

   e) In Wireless Local Area Network (WLAN), each point is a wireless point. ☐

   f) Basic Service Set (BSS) provides the ability for all Access Point (APs) to get connected to each other. ☐

2. Discuss about Bluetooth Technology?

   ……

   ……

   ……

   ……

3. Explain how Cordless Telephone System works?

……

……

……

3.3 WIRELESS GENERATIONS

Cellular Mobile System has come a long way as at present scenario, every person carries a cellular phone in his/her hand. This tremendous growth has established the growth of cellular technology. From a cordless phone which gave mobility to users but only of short distance in meters to a basic phone which has overcome the disadvantage of short range cordless phone. And now today, a basic cellular phone is converted to a smart multimedia cellular phone which is used not only for making calls but is used to click pictures, listen songs, record voice, checking mails etc. In this section, we will draw your attention towards the emerging generations of a wireless cellular phone.

Before starting with generations, we will discuss the two channel access technologies - Time Division Multiple Access (TDMA) and Frequency Division Multiple Access (FDMA) which are used in these generations.

Frequency Division Multiple Access (FDMA) allocates individual frequency channel to an individual user at a time. Each user gets a frequency channel whenever the user demands for it. Any other user cannot use the same frequency channel until the assigned channel is given back by the user to the pool of freely available frequency channel.
Time Division Multiple Access (TDMA) divides an individual frequency channel into number of time slots. These time slots are then allocated to users on demand. Unlike FDMA, one or more user can share the same frequency channel. Each slot is used either for transmitting or receiving signals. Therefore, the data transmission is non-continuous in nature which makes the hand-off simpler. Consecutive slots are used to transmit the data.

3.3.1 First Generation (1G) –

First Generation also called 1G is based on analog Frequency Modulation (FM) and Frequency Division Multiple Access (FDMA). 1G uses circuit switched technology and came in 1980. According to this generation, each user has allocated with a dedicated frequency channel. Moreover, this generation made solely to provide voice services to its users. It was not intended for any data services. These lacked features in 1G were the biggest reason behind the rise of Second Generation (2G).

3.3.2 Second Generation (2G) –

Second Generation provides the voice as well as data services to its users. Unlike 1G, no user has allocated dedicated frequency channel. 2G uses digital modulation technique and multiple access techniques like Time Division Multiple Access (TDMA) and Code Division Multiple Access (CDMA). 2G came in 1990 and uses circuit switched technology. Every user in 2G uses a time-sharing frequency channel. In this generation, there are 3 TDMA standards which are – Global System Mobile (GSM), Interim Standard 136 (IS 136) and Pacific Digital Cellular (PDC) and one CDMA standard called 2G CDMA or Interim Standard 95 (IS 95) CDMA.

- *Global System Mobile (GSM)* – For every 200 KHz radio channel, there are 8 times slotted users. 2G TDMA standard GSM is used in countries like Europe, Australia, Asia and South America. It is also used in India

- *Interim Standard 136 (IS 136)* - For every 30 KHz radio channel, there are 3 times slotted users. 2G TDMA standard IS 136 is used in countries like Australia, North America and South America. IS 136 is also known as US Digital Cellular (USDC) or North American Digital Cellular (NADC).
- *Pacific Digital Cellular (PDC)* – This is a Japanese standard and is very similar to IS 136 with around 50 million users
- *2G CDMA or Interim Standard 95 (IS 95) CDMA* – For every 1.25 MHz channel, there are up to 64 users which are orthogonally coded. This standard is also known as CDMAOne and is used in Australia, Korea North America, Japan, China and South America.

3.3.3 Evolution to Mid of Second Generation (2.5G) –

After the second generation comes 2.5G and introduced in the year 2000. 2.5G is intended for faster data rates which are required for supporting modern internet applications. Existing 2G equipment is modified (both hardware and software) to support 2.5G services for enhanced data rates. Enhanced rata rates are provided for services such as web browsing, mobile commerce, e-mail services and location based mobile services. 2.5G also supports web browsing technology like Wireless Application Protocol (WAP).

2.5 provides three TDMA upgrades which are as follows-

- *High Speed Circuit Switched Data (HSCSD)* – As the name suggests, this TDMA upgrade is a circuit switched technology and provides higher data speed rates as compared to 2G. This technology upgrade is the first attempt to provide better data rates for GSM. Rather than allocating a single time slot to a single user, the higher data rates are provided to users by providing consecutive time slots. This technology also takes care of error control coding algorithm.
- *General packet Radio Service (GPRS)* – Unlike HSCSD, this TDMA upgrade is a packet based technology. GPRS supports more users as compared to HSCSD. It uses 2G TDMA modulation format but redefines air interface for better packet data access. It is more suitable for non real time applications like retrieval of email, faxes and asymmetric web browsing. Installation of internet gateway and new routers at base station is mandatory for using this technology.
- *Enhanced Data rates for GSM Evolution (EDGE)* – EDGE provides better data rates than GPRS by using new digital modulation technique called 8-PSK (Phase Shift Keying).This is implemented by upgrading hardware and software at base station. This technology is also called as Enhanced GPRS. Nine different air interface formats are defined by EDGE known as Multiple modulation and Coding Schemes (MCS) with error control protection.

3.3.4 Third Generation (3G) –

3G is designed to provide higher data rates with much available wider bandwidth. It uses packet switched technology and users uses smaller bandwidth. This generation allows the identification of user's location. The 3G technology provides the services like transparent roaming, communication using Voice Over Internet Protocol (VOIP), receives live music, interactive live web sessions, better network capacity, multi mega-bit internet access, readily available internet access and simultaneous exchange of voice and data packets using a single cellular mobile.

The above discussed Wireless generations are compared below in the form of a comparison Table 1.

Table 1: Comparison between Wireless Generations – 1G, 2G, 2.5G and 3G

	1G	2G	2.5G	3G
Introduced in year	1980	1990	2000	After 2004
Communication Method	Circuit Switched	Circuit Switched	Both Packet and Circuit Switched	Packet Switched
Modulation Technique	Analog Frequency Modulation	Digital Modulation	Digital Modulation and Shift Keying	Digital Modulation and Shift Keying
Services	Voice service only	Both Voice and Data services	Both Voice and Data services with faster data rates	Both Voice and Data services with faster data rates
Channel Assignment	Dedicated Frequency Channel	Dynamic Channel Assignment	Dynamic Channel Assignment	Dynamic Channel Assignment
Standards	-	3 TDMA Standards – GSM, PDC, IS 136 and 1 CDMA Standard	3 TDMA Standards – HSCSD, GPRS and EDGE	EDGE and W-CDMA

3.4 INTRODUCTION TO CELLULAR MOBILE SYSTEMS - GSM

Now a day, everyone is dependent on a cellular phone (called mobile) to get connected to other person. This connectivity among users is wireless in nature. Such communication is furnished by the standards like GSM (Global System Mobile), CDMA (Code Division Multiple Access) etc. This wireless link is called the Radio Link. All the communication between users takes place through this radio link and in open wireless medium called the Common Air Interface (CAI). The concept of Global System for Mobile Communication (GSM) was introduced in 1990 by the European country. From then, this standard accepted widely and utilized by several countries.

GSM network consists of several components which are as follows:

Mobile Station (MS) - This is the device which is used by the GSM user and is portable, small, light-weight and hand-held device.

Base Transceiver Station (BTS) - It is the cell tower which is located on the roof by the service providers to provide network to its users. A BTS is connected to MS by wireless radio links.

Base Station Controller (BSC) - This controls one or more BTS and is connected to them. This connectivity is through wires. BTS and BSC together called Base Station (BS).

Mobile Switching Centre (MSC) – A MSC is connected to number of BSC and manages the call routing process.

Authentication Centre (AuC) – Authentication Centre is responsible for authenticating a legitimate user (subscriber) and also provides 128-bit authentication key to user. Home Location Register (HLR) – This is a database which stores the user's information and its location information. This provides user an IMSI (International Mobile Subscriber Identity) number to identify its user. In other words, the area to which a subscriber belongs is saved in HLR.

Visitor Location Register (VLR) – This database contains the information about subscriber who visited the area of a particular MSC and stores the IMSI (Internationl mobile subscriber indentity) number temporarily.

Operation Maintenance Centers (OMC) – The operation of each MS, BTS, BSC and MSC is monitored and maintained by this centre.

Subscriber Identity Module (SIM) – This is a removable 16k or 32k chip (or a small smart card) which a service provider provides to its subscriber. It is used in MS to access the GSM services like calling, messaging etc.

Public Networks – This consists of networks like PSTN (Public Switched Telephone Network), Data Network, ISDN (Integrated Services digital Network) to which MSC is connected.

Below given Figure 11 is the architecture of GSM containing all the above described components. GSM communication operates on 900 MHz/1800 MHz standards and uses techniques like FDD (Frequency Division Duplexing) and TDMA (Time Division Multiple Access).

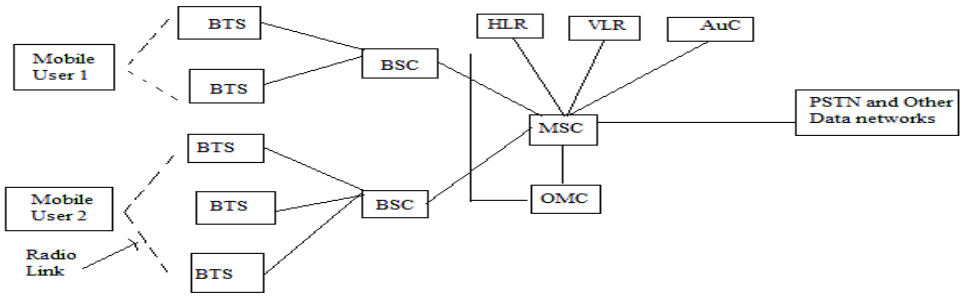


Figure 11: Global System Mobile (GSM) Architecture

Several generations like 1G (First Generation), 2G (Second Generation), 3G (Third Generation) in GSM has evolved during the past years. Even though the GSM network is utilized by almost every country these days but this standard has some vulnerabilities which are exploited by an intruder to get the access into the network or disturb its operation. The radio link between the MS and BTS is the most crucial point where an intruder takes advantage. Such vulnerabilities are listed below.

Vulnerabilities in GSM Communication

The GSM standard has some principles of security like subscriber identity confidentiality, use of a SIM as security module, subscriber identity authentication, use of triplets and stream ciphering of user traffic & user control data. An intruder takes an unfair advantage between a legitimate subscriber and the wireless radio link and breaches the security principles of GSM. This breach of principle is due to the

Following vulnerabilities present in GSM network:-

- *Wireless Radio Link* – All the communication is taking place through the medium of air. An intruder can easily intercept the communication between two subscribers or between a subscriber and its connected BTS.
- *Insecure A3/A5/A8 Algorithm* – GSM standard uses three algorithms. A3 algorithm is used for authenticating the subscriber through a 128-bit authentication key. A5 algorithm is used for encryption and decryption process and A8 algorithm is used for generating random keys. Many intruder attacks these three algorithms to know about the whole procedure. Every service provider keeps these algorithms confidential. But most of the intruder's targets the algorithm of GSM.
- *One-way Authentication* - In GSM network, only a BTS can authenticate a subscriber but a subscriber cannot authenticate a BTS. The problem arises when an intruder compromises a BTS and imposes attack through this BTS on legitimate subscriber.
- *Cloning of SIM Card* – An intruder can clone (or make a copy of a SIM card) by just deriving a 128-bit authentication key from the legitimate subscriber's SIM card. This results in misusing the SIM for fraudulent purpose.
- *No Integrity of Data* - In GSM standard, the authentication and confidentiality of a subscriber is maintained but there is no security provided for integrity of the data. An intruder can easily change the data with some fake data.

Advantages of GSM:

- GSM is already used worldwide with millions of subscribers.
- International roaming allows subscriber to use a single mobile phone throughout Western Europe. CDMA works in Asia, but not in France, Germany, the U.K. and other popular European destinations.
- GSM is mature which started in the mid-80s which is more stable network with robust features. CDMA is still building its network.
    i) GSM's maturity means engineers cut their teeth for the technology to create an unconscious preference.
- The availability of Subscriber Identity Modules, which are smart cards that provide secure data encryption which gives GSM mobile commerce advantages.

Disadvantages of GSM:

- Lack of access to American market.

☞ Check Your Progress 2

1. **State True or False**

    i) Enhanced Data rates for Gsm Evolution (EDGE) provides better data rates than General Packet Radio Service (GPRS). ☐

    ii) Global System Mobile (GSM) arrived in 2.5 G. ☐

    iii) Home Location Register (HLR) and Visitor Location Register (VLR) are the components of Mobile Switching Center (MSC). ☐

    iv) 2G provides data services only. ☐

    v) Interim Standard 136 (IS 136) is introduced in 3G. ☐

2. Compare 1G, 2G, 2.5G and 3G generations.

…….

…….

…….

3. Explain GSM architecture with a diagram.

…….

…….

…….

3.5 CODE DIVISION MULTIPLE ACCESS (CDMA)

Code Division Multiple Access (CDMA) started in 1993 when the first CDMA standard IS-95 issued. In 1995, CDMA technology put into commercialization in Hong Kong and America on large scale. In April, 2001, China Unicom began to construct CDMA networks—the largest in the world. At present, CDMA commercial networks are established in about 40 countries or area which is approximately 20% of all users in the world.

Code Division Multiple Access is a multiple access based technology which provides 1.25 MHz bandwidth per carrier. Its reuse factor is 1 (Figure 12) where as GSM reuse factor is 7, CDMA is available on operating frequency 450, 800, 1900 MHz. It provides inherently superior receive sensitivity (approx. -121 dB). In CDMA, there is a tradeoff between Capacity, Coverage and Quality. It uses precise power control algorithms which minimizes interference. It has multiple diversities like it receives spatial diversity through two receive antennas, path diversity through rake receivers, frequency diversity through spread spectrum and time diversity through interleaving. In CDMA, each user has a unique PN (Pseudo Noise) code. Each user transmits its information to other users by spreading with unique code. CDMA technology uses Direct Sequence Spread Spectrum (DSSS) Unlike other cellular technologies like GSM, each user is separated by a code not by time slot and frequency slot. Moreover, each user share the same bandwidth as the PN code separates and isolates each user and therefore prevents form interference.

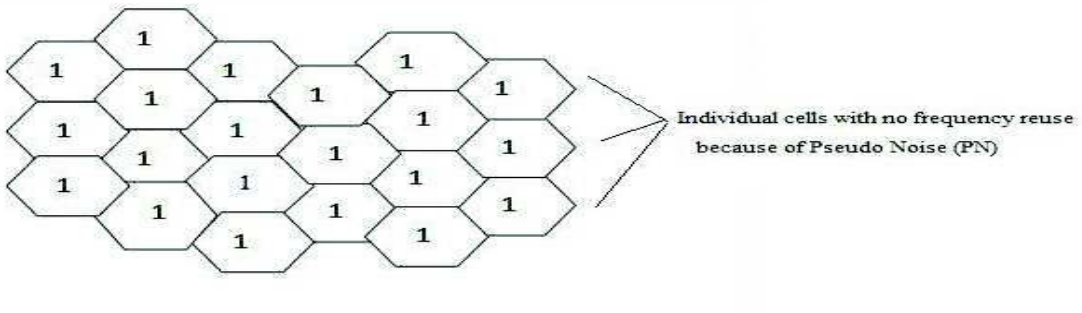


Figure 12: Code Division Multiple Access (CDMA) Frequency Allocation

CDMA technology can be used for implementing WLL (Wireless Local Loop). Existing landline operators can extend their network with WLL. Cellular operators can capitalize on their current network to deliver residential service with WLL. New service providers can quickly deploy non-traditional WLL solutions to rapidly meet a community's telephony needs.

Advantages of CDMA include:

- Increased cellular communications security.
- Provides simultaneous conversations.
- Increased efficiency so that the carrier can serve more subscribers.
- Smaller phones.
- Low power requirements and little cell-to-cell coordination needed by operators.
- Extended reach - beneficial to rural users situated far from cells.
- Uses Direct Sequence Spread Spectrum (DSSS) technology
- Provides soft & softer handoff of a user crossing between cellular region
- Uses rake receiver
- Provides high quality voice to its users
- Has power control
- Gives better coverage area network
- Has a very simple network planning of cells
- Provides smooth migration to 3G and the operator's benefit is protected.

Disadvantages of CDMA include:

- Due to its proprietary nature, all of CDMA's flaws are not known to the engineering community.
- CDMA is relatively new, and the network is not as mature as GSM.
- CDMA cannot offer international roaming, a large GSM advantage.
- Higher spectrum requirement.

☞ Check Your Progress 3

1. Explain how frequencies are allocated in Code Division Multiple Access (CDMA)?

   ……

   ……

   ……

2. List all advantages and disadvantages of Code Division Multiple Access (CDMA).

   ……

   ……

   ……

3.6 CELLULAR SYSTEM DESIGN FUNDAMENTAL

This section discusses about the basic fundamentals of cellular systems which are important in designing a cellular system. Such fundamentals concepts are frequency reuse, hand-off and signal strength threshold, Interference, and Coverage & Capacity improvements. We will discuss these basics of cellular design fundamentals one by one

3.6.1 Frequency Reuse

GSM technology involves the concept of frequency reuse. As the name suggests, the given frequencies are used again and again in different cells so as to serve more users at a time. A geographic area is divided into large hexagonal cells (not in practice) and then the frequencies are allotted to each cell. Each cluster has total of 7 small hexagonal cells (Figure 13). In this, each of the cell in cluster uses different frequencies so as to avoid interference and reuses them in cell in different clusters in order to provide service to all users. Each large cell has frequency starting from 1 to 7. The smaller inner most of each large cell is allotted the frequency 1 and then frequency 2 to its small cell which is at the top of the inner most (or central) small cell. Now, the next frequency is 3 which is clock wise next and so on.

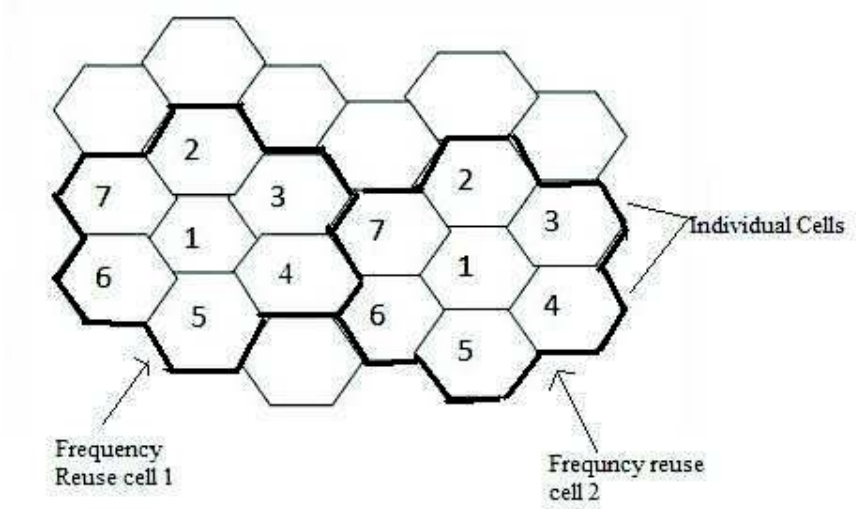


Figure 13: Frequency Reuse Concept

This is the way how the frequencies are reused in order to serve more and more users. Moreover, the reason of dividing the area into hexagonal cells than dividing it into triangle, circle or rectangle is that the hexagon has largest area for a given radius. Also the area of unit is proportional to number of base stations which is equal to the proportional to setup cost of base stations and the number of neighbors to a single unit is a way of hand-off which equals to the proportional to base station networking and control complexity.

3.6.2 Hand-off and Signal Strength

Hand-off is a way to transfer a user's calls one from one cell to another. It is also known as "Hand-over" of a user from one base station to another. There are two types of hand-offs. One is *Hard hand-off* in which the channel in the existing cell which the user is about to leave is released first and only then the channel in the target cell is engaged. Therefore, the connection to the existing cell is broken before or 'as' the connection to the target is made—for this reason such handovers are also known as *break-before-make*. Hard handovers are intended to be instantaneous in order to minimize the disruption to the call. When the mobile is between base stations, then the mobile can switch with any of the base stations, so the base stations bounce the link with the mobile back and forth. This is called *ping-ponging*.

A *soft handoff* is one in which the channel in the existing cell is retained and used for a while in parallel with the channel in the target cell. In this case the connection to the target is established before the connection to the existing is broken, hence this handover is called *make-before-break*. The interval, during which the two connections are used in parallel, may be brief or substantial. Soft handovers may involve using

connections to more than two cells: connections to three, four or more cells can be maintained by one phone at the same time. When a call is in a state of soft handover, the signal of the best of all used channels can be used for the call at a given moment or all the signals can be combined to produce a clearer copy of the signal. The latter is more advantageous, and when such combining is performed both in the downlink (forward link) and the uplink (reverse link) the handover is termed as *softer*. Softer handovers are possible when the cells involved in the handovers have a single cell site.

The question arises here is when to make a handoff or a handover? The answer to this question is based on the *signal strength* and the *minimum threshold value* of the strength required. Consider a simple scenario in which a user is moving from A place to B place. The user is on call. Now as the user is moving, the cell phone is constantly linked with the base station with the full signal strength. As the user moves away from the existing base station, gradually the signal strength keeps on decreasing with the distance. Now the point will come where the strength becomes so low that the minimum threshold value which is maintaining the links with existing base station has reached zero level. And this threshold value is increasing in correspondence with the new or target base station which is enough to maintain the call through the radio links.

3.6.3 Interference

Interference is the disturbance caused in the medium to degrade the quality of service. The reason behind this interference can be a call in neighboring cells, base stations operating on same frequency, or any other mobile in the same cell. Such interference is the consequence of cross talk where a caller gets connected to another unintended called party. In cellular system, interference can be a *Co-Channel Interference* or *Adjacent Channel Interference.*

Co-Channel Interference is caused due to the frequency reuse phenomenon as the base stations which are operating at the same frequency causes interference. As described above the concept of frequency reuse, each cell has base station which is operating on a frequency. This interference degrades the receiver performance as the signal arrives from both intended transmitter and undesired transmitter which are operating on same frequency.

Adjacent Channel Interference is caused by extraneous power from a signal in an adjacent channel and is caused due to the base stations which are operating at adjacent frequencies. The reason behind this interference can be inadequate filtering, improper tuning or poor frequency control. This can be handled by applying the technologies like proper channel assignment and careful filtering.

3.6.4 Coverage and Capacity Improvements

The cellular system constantly needs improvements in order to better service (in terms of signal strength- coverage and readily available service - capacity) to its users. This can be achieved by two technologies – *Cell Splitting* and *Cell Sectoring* for capacity improvements and *Repeaters* for coverage improvement.

Cell Splitting – Just like the name, this technology splits a single cell into number of small cells. One cell may be divided into three smaller cells so that the capacity of users can be handled easily and all users get served simultaneously. Moreover, the all splitted cell (as shown in red color in Figure 14) has its own base stations.

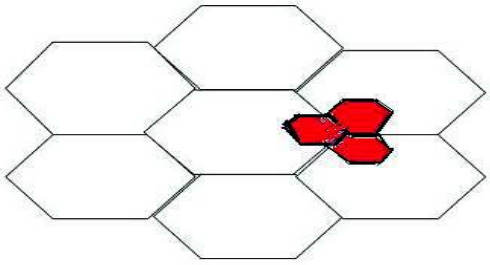


Figure 14: Cell Spelitting

As the number of base station has increased, the transmitted power of base station is reduced. This is because of the reason of small coverage area which is the region of splitted cell. This is done to handle varynig mobility. faster user are handle by larger sells and slow moving users are dealt by small cells.

Cell Sectoring – As the name suggests, a single cell is divided into small sectors at angle of 120 degree or 60 degree. When a cell is sectored into six small cells, this sectoring is called 60 degree sectoring. When a cell is sectored into three small cells, this sectoring is called 120 degree sectoring (Figure 15 & 16). This is an another way of improving capacity of a particular cell. Moreover, 120 degree sectoring reduces co-channel interference as the antennas used in the technology are directional antennas and not omni-directional antenna. Directional antenna signals are directed in a particular direction where as an Omni-directional antenna signals are directed in all directions equally.

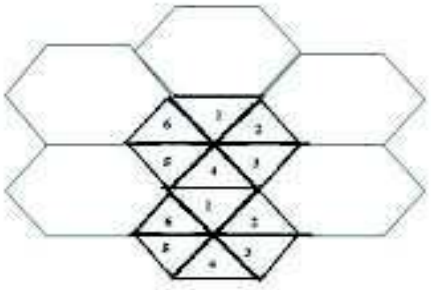


Figure 15: Cell 60 Degree Sectoring

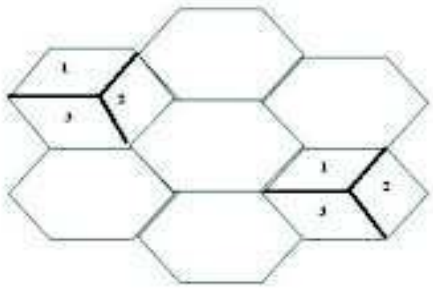


Figure 16: Cell 120 Degree Sectoring

Repeaters – This technology is employed in order to improve coverage of a cellular site. Radio repeaters are used to provide extended range at the places where the signals face obstacle and are difficult to reach like in buildings, basements etc. As repeaters are bi-directional and has range extension capability, the signal reaches the target places easily.

☞ Check Your Progress 4

1. **State True or False**

   i) Total of 8 small cells are needed to make a one large cell in order to reuse the frequency. ☐

   ii) Threshold level decreases as the user moves away from existing base stations. ☐

   iii) Hard Hand-off relies on the concept of break before make. ☐

iv) Adjacent Channel Interference is caused due to frequency reuse concept. ☐

v) Both Cell Splitting and Cell Sectoring are the solution for coverage improvement. ☐

2. Explain the difference between Adjacent Channel Interference and Co-Channel Interference?

……

……

……

3. What is Cell Sectoring? State its type?

……

……

……

3.7 SUMMARY

This completes our discussion on the Wireless Communication Networks which includes Independent Mode and Ad-hoc Mode. Further, we discussed various wireless communication systems such as Paging System, Cordless Telephone Systems, Cellular Mobile Systems, Global System Mobile (GSM), and Code Division Multiple Access (CDMA). Also, we discussed the various wireless generations from 1G (First Generation), 2G, 2.5G and 3G and compared these with each other in the form of a table. At the end of a unit, various cellular design fundamental have been discussed which covers concepts like frequency reuse, hand-offs, Coverage and Capacity improvements and Interference.
The information given on various topics can be supplemented with additional reading. However, wireless technology is very popular and useful these days and provides mobility to the users flying regularly from one place o another.

3.8 SUGGESTED READING

1. Rappaport, Theodore S. 2005. *Wireless communication – Principles and Practice. Second Edition,* Pearson Prentice Hall of India (PHI)

2. Smith, Richard Keith. 2006. *Mobile and Wireless Communications: An Introduction*. Tata McGraw-Hill Publication

3. Palanivelu and Nakkeeran. 2009. *"Wireless and Mobile Communication"* PHI Learning Pvt. Ltd

4. Schiller, Jochen H. 2003. ***Mobile** Communication*. Addison- Wesley Publications

5. Schwartz, Mischa. 2005. *Mobile and Wireless Communications,* Press Syndicate of the University of Cambridge

6. Vijay K. Garg, etl. Wireless Communication , Pearson

7. www.wikipedia.org

8. Stallings, William. 2009 *Wireless Communications and Networks*, Pearson Education India

3.9 SOLUTIONS / ANSWERS

☞ Check Your Progress 1

1. i) False
   ii) True
   iii) True
   iv) False
   v) False
   vi) False

2. Bluetooth works on ad-hoc mode in which the network is formed quickly and is of temporary basis. Bluetooth technology is created by a telecom company called Ericsson in 1994. It was developed in order to connect two devices without wires. The range in which Bluetooth technology works is of 10 meters (or 30 feet approx) only. The name "Bluetooth" is after tenth-century king Harald I of Denmark and parts of Norway who united Danish tribes into a single kingdom. The implication is that Bluetooth does the same with communications protocols, uniting them into one universal standard.

   Bluetooth works on 2.4 GHz ISM band (Industrial, Secientific and Medical band) which divides the data into parts and sends it on up to 79 bands. It uses Frequency Hopping Spread Spectrum (FHSS) with Time Division Duplexing (TDD) technique at the rate of 1600 hops/sec. Moreover, the modulation technique employed is Guassian Frequency Shift Keying (GFSK) which was the only available modulation technique at the time of Bluetooth. Data rate is around 128 Mbps (Mega Bits per Second) and can support up to 8 devices simultaneously in Master-Slave mode. Bluetooth has versions started from version 1.0 to version 4.0.

3. Cordless Telephone system consists of a landline telephone which is a fixed port. This landline is connected to the telephone exchange called Public Switched Telephone Network (PSTN). The landline telephone has a wireless (or cordless) handset which is connected to the landline telephone through a radio link. Therefore, through this cordless system, the user has the freedom to move while on a call. But this has a range or distance limitation which is around few tens of meters only.

☞ Check Your Progress 2

1. i) True
   ii) False
   iii) True
   iv) False
   iv) False.

2. Comparison between 1G, 2G, 2.5G and 3G

	1G	2G	2.5G	3G
Introduced in year	1980	1990	2000	After 2004
Communication Method	Circuit Switched	Circuit Switched	Both Packet and Circuit Switched	Packet Switched
Modulation Technique	Analog Frequency Modulation	Digital Modulation	Digital Modulation and Shift Keying	Digital Modulation and Shift Keying
Services	Voice service only	Both Voice and Data services	Both Voice and Data services with faster data rates	Both Voice and Data services with faster data rates
Channel Assignment	Dedicated Frequency Channel	Dynamic Channel Assignment	Dynamic Channel Assignment	Dynamic Channel Assignment
Standards	-	3 TDMA Standards – GSM, PDC, IS 136 and 1 CDMA Standard	3 TDMA Standards – HSCSD, GPRS and EDGE	EDGE and W-CDMA

3. GSM network consists of several components which are as follows:

   Mobile Station (MS) - This is the device which is used by the GSM user and is portable, small, light-weight and hand-held.

   Base Transceiver Station (BTS) - It is the cell tower which is located on the roof by the service providers to provide network to its users. A BTS is connected to MS by wireless radio links.

   Base Station Controller (BSC) - This controls one or more BTS and is connected to them. This connectivity is through wires. BTS and BSC together called Base Station (BS).

   Mobile Switching Centre (MSC) – A MSC is connected to number of BSC and manages the call routing process.

   Authentication Centre (AuC) – Authentication Centre is responsible for authenticating a legitimate user (subscriber) and also provides 128-bit authentication key to user.

   Home Location Register (HLR) – This is a database which stores the user's information and its location information. This provides user an IMSI (International Mobile Subscriber Identity) number to identify its user. In other words, the area to which a subscriber belongs is saved in HLR.

   Visitor Location Register (VLR) – This database contains the information about subscriber who visited the area of a particular MSC and stores the IMSI number temporarily.

   Operation Maintenance Centers (OMC) – The operation of each MS, BTS, BSC and MSC is monitored and maintained by this centre.

Subscriber Identity Module (SIM) – This is a removable 16k or 32k chip (or a small smart card) which a service provider provides to its subscriber. It is used in MS to access the GSM services like calling, messaging etc.

Public Networks – This consists of networks like PSTN (Public Switched Telephone Network), Data Network, ISDN (Integrated Services digital Network) to which MSC is connected.

Below given Figure 11 is the architecture of GSM containing all the above described components. GSM communication operates on 900 MHz/1800 MHz standards and uses techniques like FDD (Frequency Division Duplexing) and TDMA (Time Division Multiple Access).

☞ Check Your Progress 3

1. Code Division Multiple Access is a multiple access based technology which provides 1.25 MHz bandwidth per carrier. It reuses factor 1 (Figure 12) where as GSM reuses factor of 7, CDMA is available on operating frequency 450, 800, 1900 MHz. It uses RUIM Card and provides inherently superior receive sensitivity (approx. -121 dB). In CDMA, there is a tradeoff between Capacity, Coverage and Quality. It uses precise power control algorithms which minimizes interference. It has multiple diversities like it receives spatial diversity through two receive antennas, path diversity through rake receivers, frequency diversity through spread spectrum and time diversity through interleaving. In CDMA, each user has a unique PN (Pseudo Noise) code. Each user transmits its information to other users by spreading with unique code. CDMA technology uses Direct Sequence Spread Spectrum (DSSS) is used. Unlike other cellular technologies like GSM, each user is separated by a code not by time slot and frequency slot. Moreover, each user share the same bandwidth as the PN code separates and isolates each user and therefore prevents form interference. User axis shows cumulative signal strength of all users.

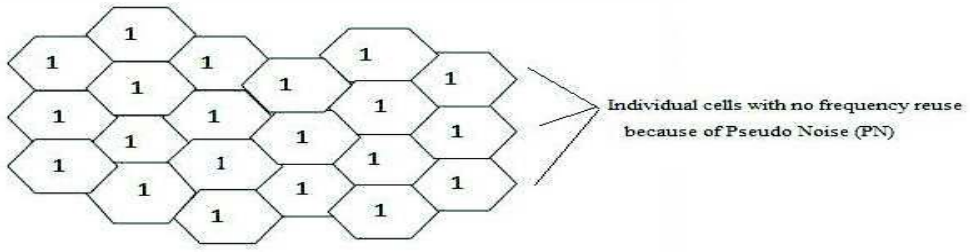


Figure 12: Code Division Multiple Access (CDMA) Frequency Allocation

2. Following are the advantages and disadvantages of CDMA -

   Advantages of CDMA include:

   - Increased cellular communications security.
   - Simultaneous conversations.
   - Increased efficiency, meaning that the carrier can serve more subscribers.
   - Smaller phones.
   - Low power requirements and little cell-to-cell coordination needed by operators.

- Extended reach - beneficial to rural users situated far from cells.
- Uses Direct Sequence Spread Spectrum (DSSS) technology
- Provides soft & softer handoff of a user crossing between cellular region
- Uses rake receiver
- Has a variable rate vocoder
- Provides high quality voice to its users
- Has power control
- Gives better coverage area network
- Has a very simple network planning of cells
- Provides smooth migration to 3G and the operator's benefit is protected.

Disadvantages of CDMA include:

- Due to its proprietary nature, all of CDMA's flaws are not known to the engineering community.
- CDMA is relatively new, and the network is not as mature as GSM.
- CDMA cannot offer international roaming, a large GSM advantage

☞ Check Your Progress 4

1. i) False
   ii) True
   iii) False
   iv) False
   iv) False.

2. *Co-Channel Interference* is caused due to the frequency reuse phenomenon as the base stations which are operating at the same frequency causes interference. As described above the concept of frequency reuse, each cells has base stations which are operating on frequencies. This interference degrades the receiver performance as the signal arrives from both intended transmitter and undesired transmitter which is operating on same frequency.

   Adjacent Channel Interference is caused by extraneous power from a signal in an adjacent channel and is caused due to the base stations which are operating at adjacent frequencies. The reason behind this interference can be inadequate filtering, improper tuning or poor frequency control. This can be handled by applying the technologies like proper channel assignment and careful filtering. Data Encryption Standard (DES) is a symmetric-key block cipher published by the National Institute of Standards and Technology (NIST) in 1977. DES has been the most widely used symmetric-key block cipher since its publication.

3. *Cell Sectoring* – As the name suggests, a single cell is divided into small sectors at angle of 120 degree or 60 degree. When a cell is sectored into three small cells, this sectoring is called 120 degree sectoring. When a cell is sectored into six small cells, this sectoring is called 60 degree sectoring (Figure 15 & 16). This is an another way of improving capacity of a particular cell. Moreover, 120 degree sectoring reduces co-channel interference as the antennas used in the technology are directional antennas and not omni-directional antenna.

Directional antenna signals are directed in the particular direction where as an Omni-directional antenna signals are directed in all directions equally.

UNIT 4 NETWORK SECURITY

Structure **Page Nos.**

4.0 Introduction to Security 70
4.1 Objectives 71
4.2 Types of Security 71
4.2.1 Application Security
4.2.2 Computer Security
4.2.3 Data Security
4.2.4 Information Security
4.2.5 Network Security
4.3 Need of Security 72
4.4 Security Services 73
4.4.1 Confidentiality
4.4.2 Availability
4.4.3 Integrity
4.4.4 Authentication
4.4.5 Non-Repudiation
4.4.6 Other Services
4.5 Authentication and Privacy 74
4.6 Block Cipher and Stream Cipher 77
4.7 Public and Private Key Cryptography 79
4.8 Introduction to RSA, DES and MD5 81
4.9 Summary 84
4.10 Suggested Reading 84
4.11 Solutions/Answers 85

4.0 INTRODUCTION TO SECURITY

Use of technology among people is increasing day by day. Such technologies are Computers, Internet (or Network), Mobile phone, Laptops, Tablets, Hard-disk etc. These technologies have internal and external memory which contains electronic data. This data can be confidential, public or private. Now, the security of such data becomes mandatory for all users so as to prevent it from any form of attack which can make this data corrupted. Therefore, Security is a very essential part of day-to-today activities. Now, we will start with defining the term “Security”.
Security can be defined by the following statements –

- the state of being secure
- precautions taken to ensure against theft, espionage, etc
- protection of assets
- free from danger or attack or threat
- form of protection

Overall, Security ensures that all processes work as expected. It is the most critical factor and has minimal standard which should be maintained by an individual or organization. This brings reliability, safety and assurance of being protected.
In this unit, you will be introduced to types of security and its services like Confidentiality, Availability, Integrity, Authentication, and Non-Repudiation etc. In addition, you will be introduced to the concepts of Cryptography and Cryptology which further define the way in which encryption and decryption can be done. Also, Public and Private Key Cryptography are introduced at later stages. Finally, we will discuss about the Public and Private Key Cryptography algorithms like RSA, DES and MD5.

4.1 OBJECTIVES

After going through this unit you will be able to:

- define the Security and its types;
- define the Security Services;
- discuss Block cipher and Stream Cipher;
- define the define the Cryptography and Cryptology
- define Public and Private Key Cryptography; and
- Define RSA, DES and MD5.

4.2 TYPES OF SECURITY

Information Technology (IT) Security – consists of following types:

1. Application security,
2. Computer security,
3. Data security,
4. Information security
5. Network security

4.2.1 Application Security

Application security prevents attack and vulnerabilities on an application. This application can be a mobile application or any other application such as web application etc. The security of an application remains throughout its lifecycle from initial phase to its running phase (or application phase) and on maintenance phase too.

4.2.2 Computer Security

Computer security is about securing a computer system (Desktop or Laptop etc) or a host. This type of security ensures a computer virus free with the help of an anti-virus software. Moreover, a computer should use genuine and updated software and hardware. Also it should be protected with a password. This type of security is a form of computer security.

4.2.3 Data Security

Data Security involves security of electronic data which is present on any hard-disks / secondary storage either of computer system or on network, on server, etc. Such security can be implemented by using passwords, cryptography (through encryption and decryption), biometric authentication, or through access control list etc.

4.2.4 Information Security

Information Security is defined as protection of information and information systems from unauthorized access, use, disclosure, disruption, modification, perusal, inspection, recording or destruction. This involves security of electronic data which is present on any database or file in any electronic memory. “Data Security” and “Information Security” are used interchangeably and are almost similar.

4.2.5 Network Security

Network Security takes care of a network, its associated processes and aims to secure it. This network can be an organizational/company internal network or any external network. All data which is coming inside the network and going outside the network is analyzed and monitored to keep the network danger free. Moreover, every process which is part of the network is also monitored.

☞ **Check Your Progress 1**

1. **State True or False** ☐

   a) National Security is part of Monetary Security. ☐

   b) Network Security monitors data incoming inside a network as well as going outside the network. ☐

   c) Information and data security are almost the same type of security. ☐

   d) Application security, Computer security, Data security, and Network security – all these are part of Information security. ☐

   e) Application level security talks deals with all the application of mobile, web etc. ☐

   f) Information Security deals with information present at network only. ☐

2. Define Security in your own terms?

   ……

   ……

   ……

   ……

3. How computer security and data security differ from each other?

   ……

   ……

   ……

4.3 NEED OF SECURITY

The question which arises here is why there is a need of security? The following vulnerabilities protections are the answer to the question -

- To isolate data from getting hacked, corrupted, unauthorized access, unauthorized modified, unauthorized deleted etc
- To maintain confidentiality, availability and integrity of data
- To prevent electronic mail from getting hacked and unauthorized access
- To protect easy passwords and pins being cracked
- To eradicate vulnerabilities (weakness) in the system or data

4.4 SECURITY SERVICES

In order to overcome the above mentioned vulnerabilities of a system or data or network etc, there are 5 major security services (Figure 1) – Confidentiality, Integrity, Availability, Non-Repudiation and Authentication which are as follows:

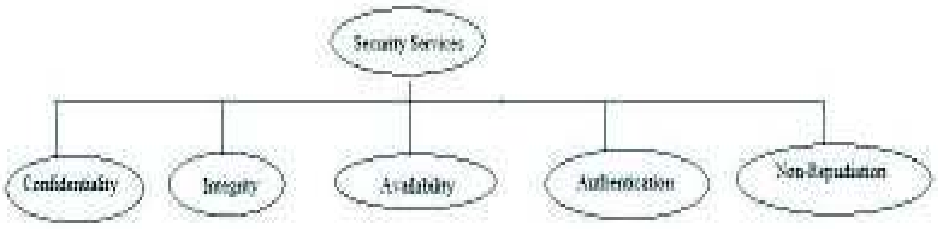


Figure 1: Security Services

4.4.1 Confidentiality

Confidentiality means keeping information secret from unauthorized access and is probably the most common aspect of information security. It is important to protect confidential information. An organization needs to guard against those malicious actions that endanger the confidentiality of its information. For example, an account user is authorized to see his account transaction online and no other account user can access this data as it is confidential.

4.4.2 Integrity

Information needs to be changed constantly. In a bank, when a customer deposits or withdraws money, the balance of their account needs to be changed. Integrity means that changes should be done only by authorized users and through authorized mechanisms. Moreover, the changes should get reflected at all the ends on which the changed information is accessed.

4.4.3 Availability

The third component of information security services is availability. The information created and stored by an organization needs to be available to authorized users and applications. Information is useless if it is not available o authorized users. Information needs to be changed constantly, which means that it must be accessible to those authorized to access it. Unavailability of information is just as harmful to an organization as a lack of confidentiality or integrity. Imagine what would happen to a bank if the customers could not access their accounts for transactions. Therefore, information should be accessible and useable upon appropriate demand by an authorized user and availability is the prevention of unauthorized withholding of information.

4.4.4 Authentication

Authentication is the process by which a person or other entity proves that it is who (or what) it says it is. For example, a bank authenticates a person or entity that deal before transferring something valuable, such as information or money, to or from, it. Authentication is achieved by presenting some unique identifying entity to the endpoint that is undertaking the process. An example of this process is the way you authenticate yourself with an ATM - here you insert your bank card (something you have) and enter your personal identification number (PIN –Personal Identification Number, something you know). Another example can be the authentication process

for email account. In this case, you have the email address and you know the corresponding account password to access the account.

4.4.5 Non-Repudiation

Non-repudiation is the prevention of either the sender or the receiver denying a transmitted message. A system must be able to prove that certain messages were sent and received. Non-repudiation is often implemented by using digital signatures. For example, a user A sent a message to user B. At later stage, user A should not deny of having sent the message to user B.

Other Security Service –

Access Control

Access control means control of access through identification and authentication. A system needs to be able to identify and authenticate users for access to data, applications and hardware. In a large system there may be a complex structure determining which users and applications have access to which objects. This is done through Access Control List (ACL). For example, an account holder while checking his data online can only view data but cannot modify it. This is because of the reason of access given to the user on the basis of his role and identity.

☞ **Check Your Progress 2**

1. **State True or False** ☐

   i) Confidentiality means to hide the data from everyone. ☐

   ii) Availability of resources or data defines the security service "Availability". ☐

   iii) Authentication is about "what you know" and "what you have". ☐

   iv) Unauthorized access is a type of vulnerability. ☐

   v) Maintaining confidentiality, availability and integrity of data are the one of the parameters for a requirement of security. ☐

2. Discuss all the possible vulnerabilities which can be a threat to information?

   ……

   ……

   ……

3. What do you understand by Security Services?

   ……

   ……

   ……

4.5 AUTHENTICATION AND PRIVACY

Authentication and Privacy refer to the problems of ensuring that communication takes place only between authorized and authenticated users or the right parties without disclosing information to unauthorized users. There is much needed security

infrastructure in place for authentication and privacy based on well known techniques in symmetric and asymmetric cryptography.
Authentication as explained in previous section is all about identifying the user and based on his identification, giving access and rights to the user. In this section, we will discuss about how an authentication can be done with the help of identification.

Authentication-Identification

Identification is all about being able to identify yourself to a computer and is absolutely essential -

- ATM, e-banking identifies a user with the help of PIN
- Access to e-mail, computer accounts, identifies a user with the help of a password
- Access to personal information (e.g., staff or student portal)

Non-computer identification

- Bank teller knows you by sight
- Bank teller checks your picture against a photo ID
- Bank back office compares cheque signature to one on record
- All examples of biometric identification.

Computer Identification

- How we identify a human to a computer?
- Username/Passwords (common),
- Token, e.g. ATM card,
- Cryptographic protocols,
- Combinations, e.g. token and password,
- Biometrics, e.g. face recognition, finger prints, and retina/iris scans

Privacy

Handling user privacy and maintaining user security are tough tasks to do. In most of the cases, it is done through a technique called “Cryptography”.
Cryptography is defined as a process of conversion of plain and readable text to cipher and (unreadable) text called encryption. For example, in Figure 2, the plain text “I am doing bca from ignou” is converted to cipher text “L dp grlqj efd iurp ljqrx” by using Caesar cipher cryptographic algorithm**.**

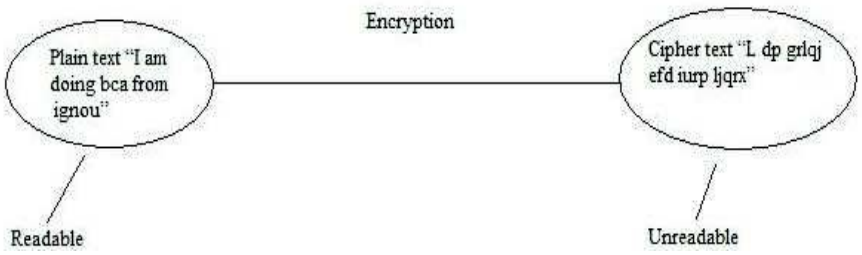


Figure 2: Process of Encryption

Decryption is the process of converting cipher and unreadable text to plain and readable text (called decryption). In given Figure 3, cipher text "L dp grlqj efd iurp ljqrx" is converted to plain text "I am doing bca from ignou" with the help of decryption process.

Please note – Both the process "Encryption" and "Decryption" are performed with the help of a key. Either the same key is used for both encryption (called symmetric or private key encryption) or separate keys (one for encryption and other one for decryption) are used called the asymmetric or public key encryption.

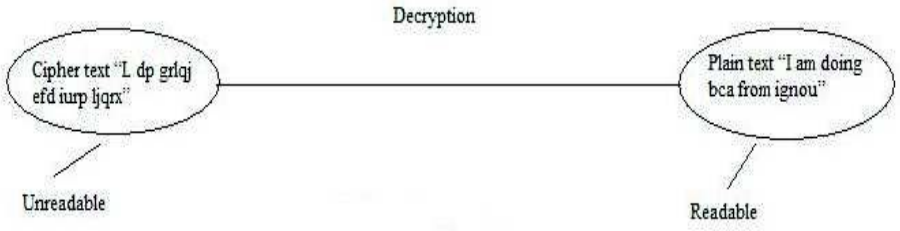


Figure 3: Process of Decryption

Cryptanalysis is the reverse process of cryptography. It means an attacker tries to find the plain text from captured cipher text. But the attacker does this without any key. The key is secured and attacker does not have any kind of access to the key. He only has the cipher text on which he applies reverse engineering.

Please Note – There is a very little difference between "Decryption" and "Cryptanalysis" as in both the cases, the aim is to know or to find the plain text behind cipher text. In decryption, the key is always available to the user who wants to decrypt the cipher text. But in case of cryptanalysis, there is no such key available to decrypt cipher text. In this situation, it is the attacker and not the user who wants to find the cipher text "without key" in order to break the cipher algorithm which is used to convert the plain text into an unreadable cipher text. The main motive is to attack the system with wrong intensions. In case of decryption part, the user uses the key to decrypt the plain text and there is no such wrong intension. The user with a key is always considered as right or authoritive person to decrypt the cipher text into its corresponding plain text.

Cryptology is the combination of Cryptography and Cryptanalysis (Figure 4).

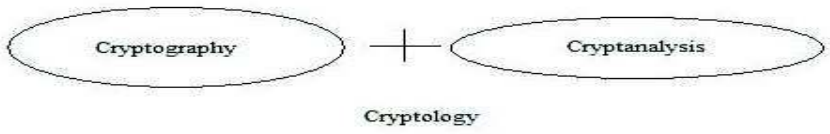


Figure 4: Cryptology

Cryptography - the process of encryption can be Symmetric (Secret Key or Private Key) and Asymmetric which will be discussed in detail in coming sections.

☞ Check Your Progress 3

1. How "Authentication" can be proved through "Identification"?

……

……

2. Difference between Cryptography and Cryptanalysis.

……

……

……

……

3. Define Encryption and Decryption?

……

……

……

……

4.6 BLOCK AND STREAM CIPHERS

Now we will discuss the method in which the plain text is converted into cipher text. In some methods, plain text is treated as numerous units or blocks and then it is converted into cipher text. But in some methods, plain text is divided into bits and these bits individually are given as input to the method which converts each single bit to the cipher text. So therefore, there are two cipher methods (Block Cipher and Stream Cipher) in which plain text is given as input in order to convert them to their corresponding cipher text.

Block Cipher, as the name suggests, takes input (i.e. plain text) and divides the plain text into number of units or blocks. After receiving input, plain text as a unit or block is encrypted with the key and converts it to a cipher text. For example, (Figure 5) the plain text "I am doing bca from ignou" is converted to cipher text "L dp grlqj efd iurp ljqrx" If this cipher text is produced by using Block cipher, then this cipher treats the plain text as "I" as first unit or block, "am" as second unit, "doing" as third unit, "bca" as fourth unit, "from" as fifth unit, and "ignou" as last and sixth unit. The corresponding cipher text produced as "L dp grlqj efd iurp ljqrx" where "L" is the cipher text for first unit, "dp" is the cipher produced for second unit, "grlqj" as the cipher for third unit, "efd" is cipher for fourth unit, "iurp" cipher for fifth unit and "ljqrx" cipher for last unit.

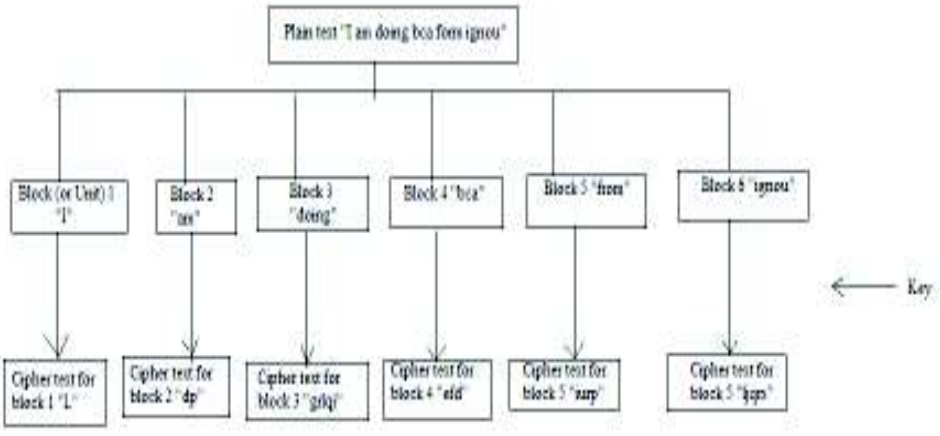


Figure 5: Block Cipher

Now we will discuss advantages and disadvantages of Block Cipher –

Advantages of Block Cipher -

- It is faster than stream cipher.
- If any block contains any transmission error then it will not have affect on other blocks.
- It is not sufficient in hardware but may be used to connect keyboard to CPU (central process unit)
- Block ciphers can be easier to implement in software, because there is no bit manipulation like in stream cipher which is time consuming process and treats data in computer-sized blocks
- Block Cipher is more suitable in trading applications.
- Short blocks at the end of a message can also be added with blank or zero status.

Disadvantages of Block Cipher -

- If two same unit or blocks of plaintext is there, then the cipher produces same cipher text for units or blocks.
- It is easy to insert or delete units/blocks .
- Block encryption may be more susceptible to cryptanalysis attack as compared to stream cipher as block cipher provides strong hints to attacker because two same units produce same cipher text.
- Block encryption is more susceptible to replay as compared to stream encryption.

Stream Cipher takes input (i.e. plain text) and divide this plain text into number of bits (combination of such bits is palin text). After receiving single bit which represents as a part of plain text is encrypted with the key and converts it to a cipher text. For example, (Figure 6) the plain text “I am doing bca from ignou” is converted to cipher text “L dp grlqj efd iurp ljqrx” If this cipher text is produced by using Stream cipher, then this cipher treats each alphabet as a single bit and converts each bit one after another to cipher text. “I” as first bit, “a” as second bit, “m” as third bit, “d” as fourth bit, “o” as fifth bit and so on. The corresponding cipher text produced as “L dp grlqj efd iurp ljqrx” where “L” is the cipher text for first bit, “d” for second bit, “p” is the cipher produced for third bit, “g” cipher text for fourth bit, “r” cipher for fifth bit and so on like this. Please note that we have taken this example for simplicity. Also we have used Caesar cipher cryptographic algorithm for both the stream.

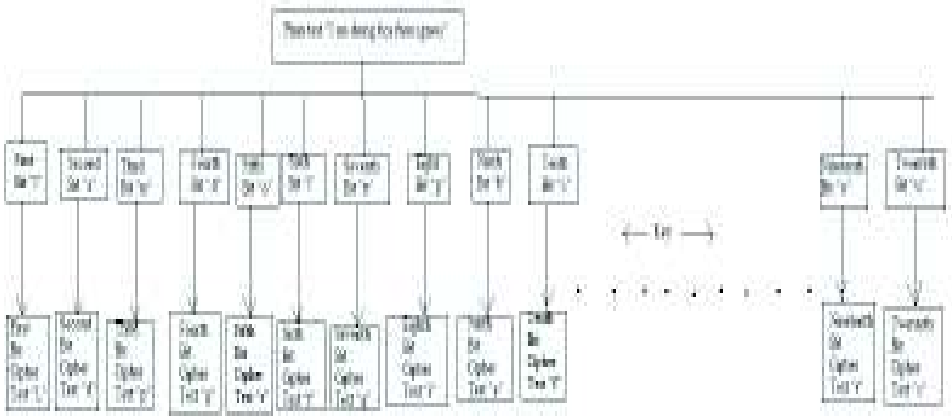


Figure 6: Stream Cipher

Now we will discuss advantages and disadvantages of Stream Cipher –

Advantages of Stream Cipher -

- Stream cipher is suitable for hardware implementation as only encryption and decryption data one bit at a time.
- Stream cipher is less susceptible to cryptanalysis than block cipher as block cipher provides strong hints to attacker because two same units produce same cipher text.
- Stream cipher is less to vulnerable to insertion or deletion of units.
- This cipher can be more easily analyzed mathematically.
- It is more suitable for military applications.
- It is less useful for attackers as same plain text is encrypted but in single individual bits and not in units.
- Self-synchronous stream ciphers are non-periodic because each key character is function dependent on the entire preceding message stream.
- Self – synchronous cipher protect against all type of authenticity threats because any change to the cipher text affects the key stream

Disadvantage of Stream Cipher -

- If during transmission, any bit is lost or become erroneous, then it is difficult to re-arrange and collect all the converted cipher text which in return may result in re-transmission of the entire plain text.
- It is slower than block but can be configured to make faster by implemented in special purpose hardware capable of encryption several million bits for second.
- It is not suitable for the software.

4.7 PUBLIC AND PRIVATE KEY CRYPTOGRAPHY

Encryption and Decryption:
Encryption is the conversion of data into a form, called a ciphertext, which cannot be easily understood by unauthorised entities whereas Decryption is the process of converting encrypted data back into its original form, so it can be understood.
Most security technologies rely, to some degree, on encryption of text or data. For example, encryption is used in the creation of certificates and digital signatures, for the secure storage of secrets or transport of information. Encryption can be anything from a simple process of substituting one character for another, in which case the key is the substitution rule, to some complex mathematical algorithm. It is to be assumed that the more difficult is to decrypt the ciphertext, the better. Trade-off - if the algorithm is too complex and it takes too long to use, or requires keys that are too large to store easily, it becomes impractical to use. There is a need a balance between the strength of the encryption; that is, how difficult it is for someone to discover the algorithm and the key, and ease of use. There are two main types of encryption in use for computer security, referred to as symmetric and asymmetric key encryption**.**

Symmetric Key

Symmetric key cryptography, also called private or secret key cryptography, is the classic cryptographic use of keys:

Here the same key is used to encrypt and decrypt the data. In given Figure 7, User A and User B both uses same secret/shared key to encrypt and decrypt the message.

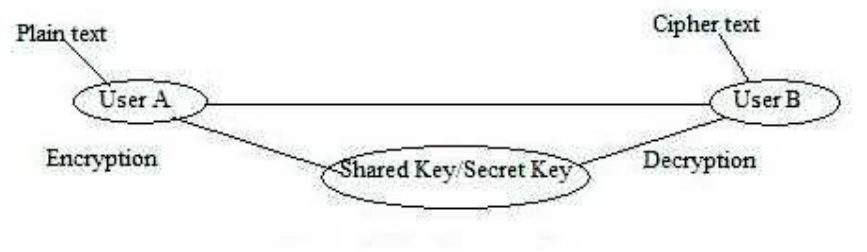


Figure 7: Symmetric/Private Key Cryptography

Asymmetric Key

In asymmetric key cryptography, different keys are used for encrypting and decrypting a message. In that case, one key can be made public called the public key while the other is kept private known as private key. There are advantages to this public-key–private-key arrangement, often referred to as public key cryptography. (1) The necessity of distributing secret keys to large numbers of users is eliminated, and (2) the algorithm can be used for authentication as well as for creating cipher text. In given Figure 8, User A takes plain text and encrypts it with public key of User B which is publically available. When User B receives cipher text, it decrypts the cipher text with its own (Private/ Secret Key).

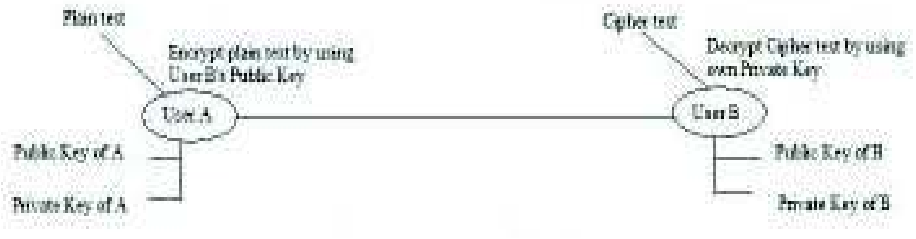


Figure 8: Asymmetric Key Cryptography

Comparison between Symmetric and Asymmetric Cryptography

- Symmetric Cryptography uses single key to encrypt and decrypt data whereas Asymmetric Cryptography uses public key to encrypt the data and private key to decrypt it.
- Symmetric key cryptography is much faster than asymmetric key encryption.
- Symmetric key cryptography does not require a lot of computer resources when compared to public key encryption which uses up more computer resources.
- In Symmetric Cryptography, secret key exchange is a problem. But in asymmetric cryptography, there is no such key exchange problem.
- In Symmetric Cryptography, origin and authenticity of message cannot be guaranteed whereas Asymmetric Cryptography provides method for message authentication, detection of tampering, non-repudiation.
- Symmetric Cryptography prevents widespread message security compromise but in Asymmetric Cryptography, widespread security compromise is possible.

☞ Check Your Progress 4

1. **State True or False**

i) Symmetric Key Cryptography uses two different key for encryption and decryption.

ii) Block Cipher is slower than Stream Cipher.

iii) Public key and Private key are the part of asymmetric key cryptography.

iv) Cipher text is the output of the process called “Encryption”.

v) Authenticity of messages is guaranteed by asymmetric key Cryptography.

2. Discuss advantages and disadvantages of Block and Stream Ciphers?

……

……

……

……

……

3. State the difference between Symmetric and Asymmetric Cryptography?

……

……

……

……

……

4.8 INTRODUCTION TO RSA, MD5 AND DES

Data Encryption Standard (DES)

Data Encryption Standard (DES) was developed as a standard for communications and data protection by an IBM research team, in response to a public request for proposals by the NBS - the National Bureau of Standards (which is now known as NIST). DES was developed in the 1970s by the National Bureau of Standards with the help of the National Security Agency. Its purpose is to provide a standard method for protecting sensitive commercial and unclassified data. IBM created the first draft of the algorithm, calling it LUCIFER. DES officially became a federal standard in November of 1976.
The Data Encryption Standard (DES) is a symmetric-key block cipher published by the National Institute of Standards and Technology (NIST) in 1977. DES has been the most widely used symmetric-key block cipher since its publication.

DES takes 64 bit plain text converts it into 64 bit cipher text with the help of 64 bit key (Figure 9) which is later reduced to 56 bit key as every 8th bit of 64 bit is discarded to form a key of 54 bit. As DES is a block cipher, it takes plain text as block of 64 bit.

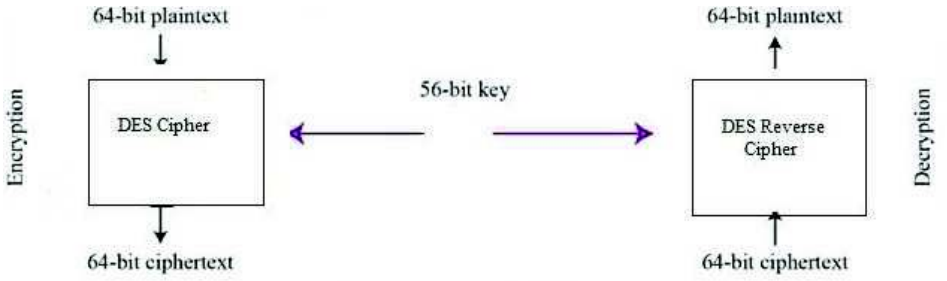


Figure 9: Data Encryption Standard (DES)

RSA

RSA is an asymmetric block cipher (as two different keys are used for encryption and decryption). It was developed by Ron **R**ivest, Adi **S**hamir and Leonard **A**dleman in 1977. A user of RSA chooses two large prime numbers and then calculates the product of two large prime numbers. The prime factors must be kept secret. Anyone can use the public key to encrypt a message, but with currently published methods, if the public key is large enough, only someone with knowledge of the prime factors can feasibly decode the message. Whether breaking RSA encryption is as hard as factoring is an open question known as the RSA problem. The following steps are involved in RSA to calculate encryption key and decryption key.

- Choose two large prime numbers p and q
- Multiply p and q together to get n
- Choose the encryption key e, such that e and $(p - 1)$ x $(q - 1)$ are relatively prime.
- Two numbers are relatively prime if they have no common factor greater than one $(1<e<((p - 1) x (q – 1))$
- Compute decryption key d such that
- $d = e \ mod \ ((p - 1) x (q - 1))$
- Construct public key as (e, n) and construct cipher text, $c = p^e \ mod \ (n)$
- Construct private key as (d, n) and construct plain text, $p = c^d \ mod \ (n)$

Now we will take two prime numbers to find the public and private key and cipher text and plain text.

- Choose two large prime numbers $p=61$ and $q= 53$
- Multiply p and q together to get $n =61*53=3233$
- Choose the encryption key e, such that e and $(p - 1)$ x $(q - 1)$ are relatively prime.
    - $(p-1) =(61-1)=60$
    - $(q-1) =(53-1)=52$
    - $(p - 1) x (q - 1) =60*52=3120$
    - Choosing a relatively prime number between $1<e<3120$ which is not a multiple of 3120. We can choose e=17
- Compute decryption key d such that
- $d = 17 \ mod \ (3120) = 2753$
- Construct public key as $(17, 3233)$ and construct cipher text, $c = 65^{17} \ mod \ (3233)=2790$

- Construct private key as *(2753, 3233)* and construct plain text, $p = 2790^{2753} \bmod (3233)=65$

Message Digest5

Before starting MD5, we will first discuss about has *Hash Functions* which takes input a plain text or a message and converts it to a hash value with the help of hash algorithm. Hash Functions are called "*One-way Functions"* as the hash value, which is the result of converted plain text, *cannot be converted back* to the plain text or message**.** Every message produces different hash value. No two different plain messages can have same hash value. Similarly, One hash value belongs to one plain text message only.

The **MD5 Message-Digest Algorithm** is a widely used cryptographic hash function that produces a 128-bit (16-byte) hash value. MD5 has been employed in a wide variety of security applications, and is also commonly used to check data integrity. MD5 was designed by Ron Rivest in 1991 to replace an earlier hash function, MD4. An MD5 hash is typically expressed as a hexadecimal number, 32 digits long. The following are the steps for Message Digest 5 algorithm –

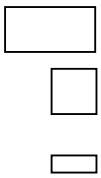
MD5 takes input of arbitrary length and gets broken into blocks of size 512 bits. It produces output of 128 bits.

- Append padding bits so *length ≡ 448 mod 512* (padded message 64 bits less than an integer multiplied by 512)
- Append length: a 64-bit representation of the length to the original message (before the padding) → total length of message k*512 bits
- Initialize MD buffer: 128-bit buffer holds intermediate and final results (4 32-bit registers, ABCD)
- Process message in 512-bit blocks
- 4 rounds of processing
- Similar structure but different logical function
- Each round takes the 512-bit input and values of ABCD and modifies ABCD
- Output: from the last stage is a 128-bit digest
- Every bit of plain text influences every bit of the the hash code
- Complex repetition of the basic functions → unlikely that two random messages would have similar regularities
- MD5 is as strong as possible for 128-bit digest (Rivest's conjecture)

Cryptographic checksum is just as a regular checksum protects the receiver from accidental changes to the message, a cryptographic checksum protects the receiver from malicious changes to the message. One-way function given a cryptographic checksum for a message, it is virtually impossible to figure out what message produced that checksum; it is not computationally feasible to find two messages that hash to the same cryptographic checksum.

If you are given a checksum for a message and you are able to compute exactly the same checksum for that message, then it is highly likely this message produced the checksum you were given.

☞ **Check Your Progress 5**



1. **State True or False**

   1. Data Encryption Standard (DES) is a symmetric-key block cipher.
   2. RSA was developed in 1978.
   3. RSA is an example of symmetric-key block cipher.
   4. RSA takes two large prime numbers as its input.
   5. Message Digest 5 (MD5) is a "One-way hash function".

2. Discuss about Data Encryption Standard (DES)?

   ……

   ……

   ……

   ……

3. Explain RSA with the help of an example?

   ……

   ……

   ……

   ……

4.9 SUMMARY

This completes our discussion on the introductory concepts of Security. The Security Services discussed in the unit are the basic mandatory services but there can be other services for security. There are many other services such as Accessibility, Authorization etc. Moreover, the security and various cryptography algorithms are introduced and designed in order to prevent passive and active attacks like Man-in-the-middle attack, Brute Force attack, Denial of Service (DOS), Distributed Denial of Service (DDOS), Virus, Worm, Trojan Horse etc.
The information given on various topics such as Cryptographic Algorithm, Block and Stream Ciphers, Security attacks, Vulnerabilities, RSA, DES, MD5 etc is exhaustive yet can be supplemented with additional reading. However, Security is an emerging field and implementation of security can be achieved by using various security tools like Intrusion Detection and Prevention Systems (IDPS), Encase, Process Viewer etc.

4.10 SUGGESTED READING

- Stallings, William 2006. *Cryptography and Network Security. Fourth Edition*, Pearson Prentice Hall Cambridge: Pearson Education Inc .
- Kahate, Atul. 2003. *Cryptography and Network Security*. Tata McGraw-Hill Publication.
- Schneier, Bruce. 2008. *"Schneier on Security"* Wiley Publications.
- Ferguson, Niels. Schneier, Bruce. and Kohno, Tadayoshi. 2010. *Cryptography Engineering,* John Wiley & Sons

- Kaufman, Charlie. Perlman, Radia. Speciner, Mike 2002. *Network Security: Private Communication in a Public World (2nd Edition)* . Prentice Hall
- Tipton, Harold F. and Krause, Micki. 2004. *Information Security Management Handbook,* Fifth Edition. Auerbach Publications.
- Rosenberg, Jothy. and Remy, David. *Securing Web Services with WS-Security: De-mystifying WS-Security, WS-Policy, SAML, XML Signature, and XML Encryption*
- Pfleeger, Charles P. and Pfleeger, Shari Lawrence. 2007 *Security in Computing*, Third Edition. Prentice Hall Publication
- Ellis, Juanita. Speed, Tim. and Crowell, William P. 2001. "*The Internet Security Guidebook: From Planning to Deployment,"* Academic Press
- Canavan, John E. 2001. *"The Fundamentals of Network Security"* Artech House.
- www.wikipedia.com

4.11 SOLUTIONS / ANSWERS

☞ Check Your Progress 1

1. a) False
   b) True
   c) True
   d) False
   e) True
   f) False

2. Security can be defined by the following statements –
   - the state of being secure
   - precautions taken to ensure against theft, espionage, etc
   - protection of assets
   - free from danger or attack or threat
   - form of protection

3. **Computer Security**

   Computer security is about securing a computer system (Desktop or Laptop etc) or a host. This type of security ensures a computer danger free and contains no virus by using anti-virus software. Moreover, a computer should use genuine and updated software and hardware. Also it should be protected with password. This type of security is a form of computer security.

4. **Data Security**

Data Security involves security of electronic data which is present on any file, folder, organization, network, computer system, electronic mail, hard-disk etc. Such security can be implemented by using passwords, cryptography (through encryption and decryption), biometric authentication, or through access control list etc.

☞ Check Your Progress 2

1. i) False

ii) True

iii) True

iv) True

v) False.

2. Following are the Vulnerabilities -

- To isolate data from getting hacked, corrupted, unauthorized access, unauthorized modified, unauthorized deleted etc
- To maintain confidentiality, availability and integrity of data
- To prevent electronic mail from getting hacked and unauthorized access
- To protect easy passwords and pins being cracked
- To eradicate vulnerabilities (weakness) in the system or data

3. In order to overcome all the vulnerabilities of a system or data or network etc, there are 5 major security services – Confidentiality, Integrity, Availability, Non-Repudiation and Authentication. If all these five basic security services are ensured then the system or network or data will be free of virus, danger etc.

☞ Check Your Progress 3

1. **Authentication-Identification**

Identification is all about being able to identify yourself to a computer and is absolutely essential -

- ATM, e-banking identifies a user with the help of PIN
- Access to e-mail, computer accounts, identifies a user with the help of a password
- Access to personal information (e.g., staff or student portal)

Non-computer identification

- Bank teller knows you by sight
- Bank teller checks your picture against a photo ID
- Bank back office compares cheque signature to one on record
- All examples of biometric identification.

Computer Identification

- How we identify a human to a computer?
- Username/Passwords (common),
- Token, e.g. ATM card,
- Cryptographic protocols,
- Combinations, e.g. token and password,
- Biometrics, e.g. face recognition, finger prints, and retina/iris scans

2. **Cryptography** is defined as a process of conversion of plain and readable text to cipher and unreadable text (called encryption). For example, in Figure 2, the plain text “I am doing bca from ignou” is converted to cipher text “L dp grlqj efd iurp ljqrx” by using Caesar cipher cryptographic algorithm**.**

Cryptanalysis is the reverse process of cryptography. It means an attacker tries to find the plain text from captured cipher text. But the attacker does this without any key. The key is secured and attacker does not have any kind of access to the key. He only has the cipher text on which he will apply reverse engineering.

3. **Encryption** is defined as a process of conversion of plain and readable text to cipher and unreadable text . For example, in figure 3, the plain text "I am doing bca from ignou" is converted to cipher text "L dp grlqj efd iurp ljqrx" by using Caesar cipher cryptographic algorithm

   Decryption is the process of converting cipher and unreadable text to plain and readable text (called decryption). In given Figure 3, cipher text "L dp grlqj efd iurp ljqrx" is converted to plain text "I am doing bca from ignou" with the help of decryption process..

☞ Check Your Progress 4

1. i) False
   ii) False
   iii) True
   iv) False
   v) True.

2. Following are the advantages and disadvantages of Block and Stream Cipher -

   Advantages of Block Cipher -

   - It is faster than stream cipher.
   - If any block contains any transmission error then it will not have affect on other blocks.
   - It is not sufficient in hardware but may be used to connect keyboard to CPU (central process unit)
   - Block ciphers can be easier to implement in software, because there is no bit manipulation like in stream cipher which is time consuming process and treats data in computer-sized blocks
   - Block Cipher is more suitable in trading applications.
   - Short blocks at the end of a message can also be added with blank or zero status.

   Disadvantages of Block Cipher -

   - If two same unit or blocks of plaintext is there, then the cipher produces same cipher text for units or blocks.
   - It is easy to insert or delete units/blocks .
   - Block encryption may be more susceptible to cryptanalysis attack as compared to stream cipher as block cipher provides strong hints to attacker because two same units produce same cipher text.
   - Block encryption is more susceptible to replay as compare to stream encryption.

   Advantages of Stream Cipher -

- Stream cipher is suitable for hardware implementation as only encryption and decryption data one bit at a time.
- Stream cipher is less susceptible to cryptanalysis than block cipher as block cipher provides strong hints to attacker because two same units produce same cipher text.
- Stream cipher is less to vulnerable to insertion or deletion of units.
- This cipher can be more easily analyzed mathematically.
- It is more suitable for military applications.
- It is less useful for attackers as same plain text is encrypted but in single individual bits and not in units.
- Self-synchronous stream ciphers are non-periodic because each key character is function dependent on the entire preceding message stream.
- Self – synchronous cipher protect against all type of authenticity threats because any change to the cipher text affects the key stream

Disadvantage of Stream Cipher -

- If during transmission, any bit is lost or become erroneous, then it is difficult to re-arrange and collect all the converted cipher text which in return may result in re-transmission of the entire plain text.

It is slower than block but can be configured to make more fast by implemented in special purpose hardware capable of encryption several million bits for second.

- It is not suitable for the software.

3. **Comparison between Symmetric and Asymmetric Cryptography**

- Symmetric Cryptography uses single key to encrypt and decrypt data whereas Asymmetric Cryptography uses public key to encrypt the data and private key to decrypt it.
- Symmetric key cryptography is much faster than asymmetric key encryption.
- Symmetric key cryptography does not require a lot of computer resources when compared to public key encryption which uses up more computer resources.
- In Symmetric Cryptography, secret key exchange is a problem. But in asymmetric cryptography, there is no suck key exchange problem.
- In Symmetric Cryptography, origin and authenticity of message cannot be guaranteed whereas Asymmetric Cryptography provides method for message authentication, detection of tampering, non-repudiation.
- Symmetric Cryptography prevents widespread message security compromise but in Asymmetric Cryptography, widespread security compromise is possible.

☞ Check Your Progress 5

1. i) True

   ii) False

iii) False

iv) True

v) True.

2. The Data Encryption Standard (DES) is a symmetric-key block cipher published by the National Institute of Standards and Technology (NIST) in 1977. DES has been the most widely used symmetric-key block cipher since its publication.

   DES takes 64 bit plain text converts it into 64 bit cipher text with the help of 64 bit key (Figure 9) which later reduced to 56 bit key as every 8th bit of 64 bit is discarded to form a key of 54 bit. As DES is a block cipher, it takes plain text as block of 64 bit.

3. RSA is an asymmetric block cipher (as two different keys are used for encryption and decryption). It was developed by Ron **R**ivest, Adi **S**hamir and Leonard **A**dleman in 1977. Now RSA with an example.

   - Choose two large prime numbers $p=61$ and $q= 53$
   - Multiply p and q together to get $n =61*53=3233$
   - Choose the encryption key e, such that e and $(p - 1)$ x $(q - 1)$ are relatively prime.
     - $(p-1) =(61-1)=60$
     - $(q-1) =(53-1)=52$
     - $(p - 1) x (q - 1) =60*52=3120$
     - Choosing a relatively prime number between $1<e<3120$ which is not a multiple of 3120. We can choose e=17
   - Compute decryption key d such that
   - $d = 17 \bmod (3120) = 2753$
   - Construct public key as $(17, 3233)$ and construct cipher text, $c = 65^{17} \bmod (3233)=2790$
   - Construct private key as $(2753, 3233)$ and construct plain text, $p = 2790^{2753} \bmod (3233)=65$