Indira Gandhi
National Open University
School of Social Sciences

BLI-224
ICT Fundamentals

Block

3

NETWORK FUNDAMENTALS

UNIT 8

Network Topology 5

UNIT 9

Communication Protocols and Network Addressing 31

UNIT 10

Protocol Architecture 67

UNIT 11

Network Applications and Management 86

UNIT 12

Network Security 114

Programme Design Committee

Prof. Uma Kanjilal (Chairperson)
Faculty of LIS, SOSS, IGNOU

Prof. B.K.Sen, Retired Scientist, NISCAIR
New Delhi

Prof. K.S. Raghavan, DRTC
Indian Statistical Institute, Bangalore

Prof. Krishan Kumar, Retired Professor
Dept. of LIS, University of Delhi, Delhi

Prof. M.M. Kashyap, Retired Professor
Dept. of LIS, University of Delhi, Delhi

Prof. R.Satyanarayana
Retired Professor, Faculty of LIS, SOSS, IGNOU

Dr. R.Sevukan (Former Faculty Member)
Faculty of LIS, SOSS, IGNOU

Prof. S.B. Ghosh, Retired Professor
Faculty of LIS, SOSS, IGNOU

Prof. T. Viswanathan, Retired Director
NISCAIR, New Delhi

Dr. Zuchamo Yanthan
Faculty of LIS, SOSS, IGNOU

Conveners:

Dr. Jaideep Sharma
Faculty of LIS, SOSS, IGNOU

Prof. Neena Talwar Kanungo
Faculty of LIS, SOSS, IGNOU

Programme Coordinators

Prof. Jaideep Sharma and
Prof. Neena Talwar Kanungo

Course Coordinator

Prof. Uma Kanjilal

Programme Editor

Prof. Jaideep Sharma

Course Preparation Team

Unit No(s).	**Contributor(s)**
8-12	(Late) Prof. T. Viswanathan

Course Editor
Prof. Uma Kanjilal

Internal Faculty:
Prof. Uma Kanjilal
Prof. Neena Talwar Kanungo

Material Production

Mr. Manjit Singh
Section Officer (Pub.)
SOSS, IGNOU

Secretarial Assistance

Ms. Sunita Soni
SOSS
IGNOU

Cover Design

Ms. Ruchi Sethi
Web Designer
E Gyankosh
IGNOU

July, 2014

ISBN : 978-81-266-6772-7

"The University does not warrant or assume any legal liability or responsibility for the academic content of this course provided by the authors as far as the copyright issues are concerned"

Further information on Indira Gandhi National Open University courses may be obtained from the University's office at Maidan Garhi, New Delhi-110 068 or visit University's Website http://www.ignou.ac.in.

Printed and published on behalf of the Indira Gandhi National Open University, New Delhi by Director, SOSS.

Lasertypesetted at Graphic Printers, 204, Pankaj Tower, Mayur Vihar, Phase-I, Delhi-110091.

Printed at :

BLOCK 3 NETWORK FUNDAMENTALS

Introduction

The rapid development of Information Communication Technologies (ICT) has resulted in proliferation of computer networks in all walks of our life. The importance of computer networks and telecommunications has grown tremendously in the last one decade. Technological advances in communication have ushered in a new era not only of computer power but of access to information services through telecommunication networks. We are fast moving towards an information society which critically depends on networked environment for its sustenance. An understanding of concepts underlying telecommunications and networking has become essential for the LIS professionals in a networked society.

This Block has five Units:

Unit 8 on **Network Topology** deals with four basic aspects of data networks, viz. topology, media access control (MAC) protocols, address resolution and routing.

Unit 9 is **Communication Protocols and Network Addressing** which deals with two distinct but closely related aspects: Internet communication protocols and network addressing.

Unit 10 covers different aspects of **Protocol Architectures**

Unit 11 dealing with **Network Applications and Management** coveres main user applications that runs on Internet and network management. The applications discussed include non-real time ones like text and multimedia messaging and real time ones like interactive television and music on demand.

Unit 12 on **Network Security** discusses information security issues in a networked environment.

Blank Page

UNIT 8 NETWORK TOPOLOGY

Structure

8.0 Objectives

8.1 Introduction

8.2 Physical and Logical Topologies

8.3 Fully Connected Topology

8.4 Star Topology

8.5 Hubs and Switches

8.6 Bus Topology

8.7 Ring Topology

8.8 Mesh Topology

8.9 Tree Topology

8.10 Hybrid Topology

8.11 Media Access Control Protocols

8.12 Address Resolution

8.13 Routers

8.14 Routing Algorithms

8.15 Summary

8.16 Answers to Self-check Exercises

8.17 Keywords

8.18 References and Further Reading

8.0 OBJECTIVES

After going through this Unit, you will be able to understand and appreciate:

- What is meant by network topology;
- Difference between physical and logical topologies;
- Basic topologies like star, bus, ring, tree and hybrid;
- Why star topology is popular;
- Topology related network components like hubs and switches;
- Different logical topologies;
- Ethernet and token passing ring protocols;
- Merits and demerits of Ethernet and token passing ring protocols;
- Why address resolution is required and how it is performed;

- Purpose of domain name servers an address resolution protocol;
- The need for encapsulation;
- Router connectivity and the functioning of routers in Internet; and
- Routing algorithms and the associated performance parameters.

8.1 INTRODUCTION

As you are aware, computers world over are interconnected via the Internet. The connection to the Internet happens in a variety of ways. For example, a home computer is usually stand-alone computer connected to the Internet via a dial up telephone line. In homes where there is more than one computer, they may be interconnected to form a home computer network. In such cases, one of the computers acts as the Internet link. It is called a proxy Internet server. Other computers access the Internet via the proxy server. The home proxy server also accesses the Internet via dial up line usually. In some rare cases, a home may use a leased line to access Internet. Computers in organisations and offices are generally interconnected locally. The local network, called Local Area Network (LAN) is then connected to the Internet via a gateway using a leased line. The gateway may be a firewall or a proxy server. There are a variety of ways in which LAN computers may be interconnected.

Network topology refers to the study of geometric properties of the way in which the computers in a network are interconnected. A generalised network connection is shown in Fig. 8.1. Here four computers are attached to what is called a network cloud. The network cloud is a graphic symbol that denotes a network without specifying the geometry or other interconnection details. Network cloud is a black box that hides the interconnection details from the viewer.

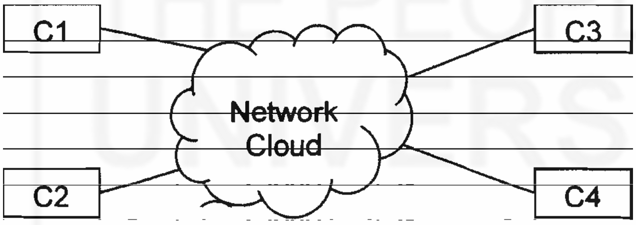


Fig. 8.1: A network cloud

There are a variety of ways in which these computers or nodes, as they are often called, can be interconnected physically inside the network cloud. For example, they may be interconnected as a daisy chain; say, C1-C2-C4-C3; or to a central switch as in the case of telephones being connected to an exchange. In a daisy chain connection, information moves from node to node in the order in which the chain is formed. In the above example, data moves from C1 to C2, C2 to C4 and C4 to C3 and vice versa for reverse flow. When a switch is used, a direct connection between two computers is established as in the case of calling and called subscribers in telephone communication. Network topology deals with the study of the way in which the computers are connected inside network cloud.

8.2 PHYSICALAND LOGICAL TOPOLOGIES

You may be aware that in data networks like the Internet, data moves in packets. A long string of text is spilt into small packets, say 1024 bytes long, and sent over the network hopping from node to node. Accordingly, these networks are called packet switched networks or packet data networks (PDNs). Interestingly, the way in which packets move in a network may not correspond to the way in which the computers are physically connected for a variety of reasons. One important reason is traffic management and routing. The idea is very similar to the vehicular traffic management. If a road is congested and there is traffic jam, one tends to take a different route even though the alternative route may be longer. Depending on the jam, the traffic may be diverted for quite sometime. Similarly, if a link is congested in a network, an alternative route may be chosen to forward the packets. Route is chosen independently for every packet depending on the traffic conditions. The path taken by a packet for traversing form a source computer to a destination computer is known as logical path. Obviously, there may be a number of logical paths in a network. The collection of all such paths is called the logical topology of the network. The physical links constitute the physical topology of the network. In essence, the logical topology refers to the way in which packets travel from a source to a destination, whereas the physical topology refers to the actual physical interconnection of the computers in the network. Physical topology is also referred to as *real* topology and the logical one as *virtual* topology.

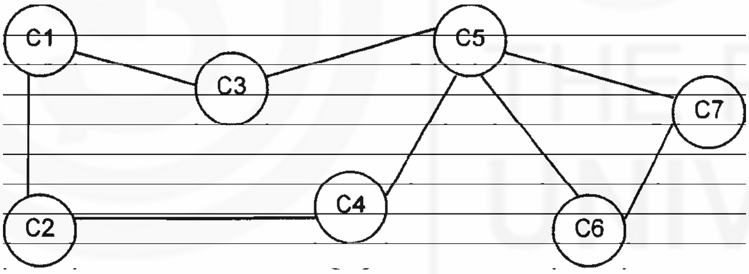
Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

1) A packet switched network uses a packet size of 2^{11} bytes. Determine the number of packets to be transmitted to transfer a file of size 1 MB.

2) Consider the physical topology given below:



Enumerate the number of logical paths between C1 and C7.

3) Can the packets of the same file travel via different logical paths? Do you foresee any problem in this case?

....

....

....

....

8.3 FULLY CONNECTED TOPOLOGY

Data communication involves computers in one part of the world being able to contact and communicate with computers in other parts of the world. For example, a home computer connects to different websites at different times. It is not just fixed one-to-one connection. It is multi-point communication connecting different to destinations at different times. For this purpose, a computer needs access to all the other computers that need to be contacted. One way of achieving this is to establish direct connections between the source computer and all the destination computers. In this case, we need as many links as there are destination computers. For example, if a home computer were to connect to 10 different web sites, then we would need 10 different links connecting the home computer to each web site. As you know, this obviously is not the case. However, if every computer in the world were to be connected to every other computer like this, then we need a very large number of links. A network formed this way for five computers is illustrated in Fig. 8.2. Here, every computer is directly connected to every other computer. Networks with this kind of connectivity are said to have fully connected topology. There are 10 links in Fig. 8.2. The links are assumed to be full duplex in the sense that they are capable of transporting information both ways. If the links were unidirectional (half-duplex) as in the case of optical fibres, we would need twice the number of links for two-way communication. The number of links required in a fully connected network becomes very large even with moderate number of computers. For example, we require 1225 links for fully interconnecting 50 computers.

In a general case with *N* computers, *N (N*- 1)/2 links are required as reasoned in the following.

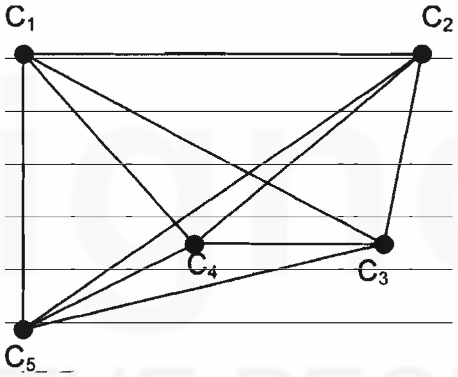


Fig. 8.2: Fully connected topology for five computers

Let us consider the *N* computers in some order. In order to connect the first computer to all other computers, we require *(N* - 1) links. With this, the second computer is already connected to the first. We now need *(N* - 2) links to connect the second computer to the others. For the third computer, we need *(N* -3) links, for the fourth *(N* - 4) links, and so on. The total number of links *N* works out as follows:

$$L = (N- 1) + (N-2) + \dots +1+0 = N(N-1)/2$$

Establishing separate and direct communication links connecting each computer to every other computer as shown in Fig. 8.2 is very expensive and is impracticable. Hence, this is just not done.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

4) How many half-duplex communication links are required for fully connected topology with 10 computers supporting full-duplex communication?

5) Draw a fully connected topology for two-way communication with four computers using fibre optic links. How many links are there in your network?

....

....

....

....

8.4 STAR TOPOLOGY

Star Topology is basically a physical topology. As the name implies, the topology looks like a star in the sky with rays emanating from the central point in all directions. It is a centralised topology where all computers are connected to a central point, which we call as *star point.* The topology is depicted in Fig. 8.3. This topology is easy to administer and maintain. The links can be tested and repaired from the central star point. This is one of biggest advantages of star topology. The topology is also a robust one. If a link or a computer fails, the rest of the network is not affected. As a result, this topology is very popular and is used extensively.

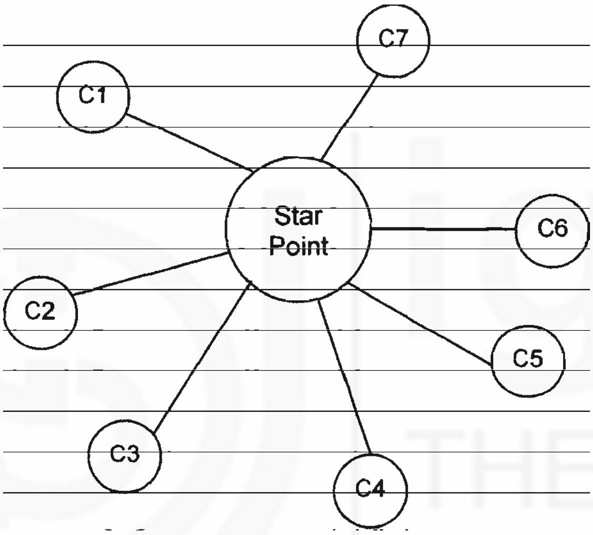


Fig. 8.3: Star topology

However, if the star point fails, the whole network fails. This is a disadvantage. Special care is taken at the time of designing the star point to make it very reliable.

Strictly speaking, physical star topology does not imply any logical topology. The logical topology is dependent on how the star point has been designed. This, in fact, makes this topology very attractive as it offers the flexibility of easy maintenance on the one hand and permits different logical topologies to be implemented on the other. Logical topology, as we mentioned earlier, defines the way in which a packet traverses form a source computer to a destination computer. Usually, there is set of rules that govern the exchange of packets between computers. Such a set of rules is called a protocol. There are many protocols used in networks. We learn more about protocols later in this unit. The star point of star topology can be designed to implement a variety of protocols such as Ethernet protocol, token ring protocol and a switch. Usually, either a hub that implements Ethernet protocol or a switch that permits switched connections is used as the star point. We learn about hubs and switches in the next section.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

6) Enumerate the advantages of star topology.

....

....

....

....

8.5 HUBS AND SWITCHES

As mentioned in the previous section, hubs and switches are used to implement logical topology in a physical star topology. Hub is a terminology used in several contexts in networks. In satellite networks, hub means a special ground station with which small satellite terminals communicate. In the context of LAN, hubs are used to implement two different logical topologies: Ethernet and token ring. But in most of the textbooks, the word hub is used to denote Ethernet hub, i.e. the hub that implements Ethernet logical topology. You must, however, be aware of the existence of different types of hubs. A hub is sometimes loosely called a concentrator as it connects to all computers in a network in star configuration. Ethernet is also implemented in bus physical topology that we discuss in the next section. Discussions on Ethernet in this section also apply to bus topology implementation. In fact, Ethernet was first implemented using bus topology. Later, it was implemented using hubs in star topology. At present, most of the Ethernet implementations are based on star physical topology using hubs.

Recall that logical topology corresponds to the way packets are transported within a network. Ethernet transports packets in much the same way information is exchanged in a group discussion among people. What happens in a group discussion? One who has something to say starts speaking and others listen. In the same way, in Ethernet topology, a computer starts transmitting whenever it has a packet to send. Other computers listen. Since any computer in the network can start a transmission Ethernet is called multiple access (MA) scheme. As happens in a group discussion, sometimes more than one computer may start a transmission simultaneously. What happens when two persons start speaking? Both of them go quiet and one among them starts afresh? Similar thing happens in Ethernet. When more than one computer start transmission simultaneously, we say that a collision has occurred. The computers that transmit simultaneously detect the collision, go quiet and follow a predetermined procedure to start the transmission afresh. Hence, Ethernet is a collision detection (CD) scheme. In a group discussion, if someone is already speaking, another person does not start to speak. Similar thing happens in Ethernet where detection of an ongoing transmission is called carrier sense (CS). On the whole Ethernet is a CSMA/CD scheme.

Let us now see how a hub helps implement Ethernet. Hubs come in 4, 8, 16, 24 and 48-port configurations. One computer can be attached to each port. Each port has provision for input/output and power connections. At the computer end there is a network interface card (NIC) that connects to the hub port. One of the ports is specially designed to be able to attach to another hub, thus allowing cascading of hubs. Cascading is useful when clusters of computer are located in nearby geographical areas. For example, an organisation spread over multiple floors of a multi-storey building, may use one hub per floor and cascade them so that computers in different floors can communicate with each other. Fig 8.4 shows a schematic of cascaded hubs with four ports each. Port 4 is specially designed to connect to another hub. Port 4 of the hub in Floor 1 is connected to Port 1 of the hub in Floor 2. Port 4 of the hub in Floor 2 is connected to Port 1 of the hub in Floor 3. Thus all the three hubs are cascaded in a daisy chain fashion. You may note that only two computers can be connected to the hub in Floor 2. If no cascading is used, the special port can be used to connect a computer. This is shown in the hub in Floor 3. Hubs that have provision for cascading are also called *stackable hubs.*

The internal mechanism of an Ethernet hub forwards any incoming packet from any computer to the output lines of all other computers as well as to the output line of the sending computer. In this sense, the hub acts as a broadcaster.

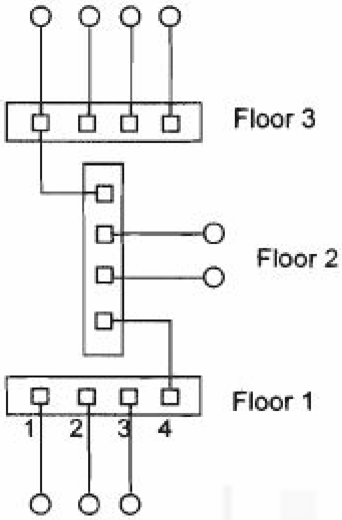


Fig. 8.4: 4-port cascaded hubs

This broadcast allows all the other computers and the transmitting computer to listen to the ongoing transmission. The transmitting computer is able to detect a collision by monitoring its own output line. If a bit received on its output line is not the same as the one sent by it, the computer knows that a collision has occurred. The Ethernet NIC in the computer senses carrier and detects collision. The hub enables multiple access feature as all the computers are connected to it and anyone can start a transmission. This is Ethernet hub implements CSMA/CD protocol. An important requirement of a hub-based design is that all computers connected to the hub must operate at the same speed.

The star point of star topology could be a switch. A switch is like a telephone exchange. The switch examines the destination address in an incoming packet and routes the packet to the appropriate outlet much as the telephone exchange examines the dialled number and routes the call to the appropriate destination. Much as the way telephone exchanges are interconnected, switches can be interconnected to route packets to computers that are not local. In fact, this is how most of the Internet connections work. An important advantage of the switch when compared to hub is that different computers can operate at different speeds. Of course, the source and the destination computer pair must operate at the same speed.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

7) What are the different logical topologies that can be implemented by hub?

....

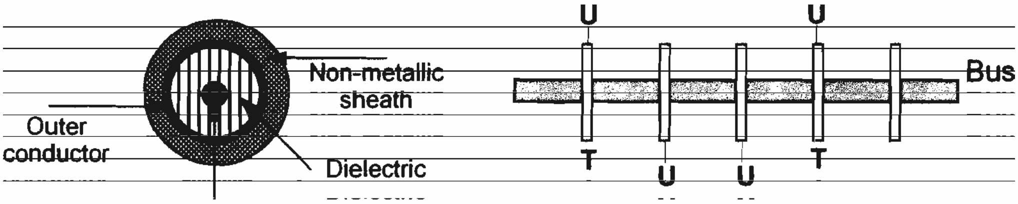
....

....

....

8.6 BUS TOPOLOGY

Bus is a cable laid linearly. Imagine a coil of cable unrolled, stretched and laid from end-to-end in a linear fashion and it becomes a bus. The cable is of coaxial type. Coaxial cable has one central conductor with a surrounding metallic shield. The central conductor and the shield are separated by a dielectric medium. Dielectric, as you may know is an insulator that electrically separates the inner conductor and the outer metallic shield. A thick non-metallic sheath further protects the central conductor and the shield. Fig 8.5(a) depicts a coaxial cable. The outer conductor (the metallic shield) is usually grounded and acts as an electromagnetic shield.



Inner central conductor

U = User node T = Tap

(a) Coaxial cable

(b) Bus topology

Fig. 8.5: Coaxial cable and the bus topology

With the outer conductor grounded, the cable essentially has one single conductor that carries current. When laid out as bus, the central conductor is like broadcast medium. Since the central conductor acts as a broadcast medium, it is called ether, the space, and the network as Ethernet. The bus topology is shown in Fig. 8.5(b). The bus cable is open at both ends and needs to be terminated with suitable terminations for proper electrical operation. Taps are essentially screws driven halfway into the central conductor. Taps are connected to the NICs of user nodes or computers. Thus, all computers are connected to the central conductor. Tap connection to the central conductor is passive and failure, if any, in one of the computers does not affect others in the network. But if the cable breaks at any point, the entire system is affected. It is not easy to detect and locate the cable break and a special device called *time domain reflectometer* is used for this purpose.

The NIC implements CSMA/CD protocol described in the previous section. Since all NICs are connected to the central conductor, there is multiple access (MA). They are able to sense (Carrier Sense) and listen to the transmission that is taking place on the bus. Any collision can be detected (CD) and necessary action taken to resolve the same.

The first version of Ethernet was implemented using a thick coaxial cable specified as 10Base5. The specification 10Base5 means the cable operates at 10 Mbps speed, uses what is called baseband modulation and can have a maximum length of 500 meters. This Ethernet version was called thick Ethernet. True to its name, the cable was thick and difficult to handle. Later another version called thin Ethernet was introduced with the cable specification as 10Base2. This cable was more flexible and easier to handle. Bus based Ethernet is the forerunner of hub based Ethernet. Because of the problems faced in the maintenance of thick and thin Ethernet, hub based solution was invented. The cable used in hub based system is a twisted pair and the specification is 10BaseT. Twisted pair cables cover a maximum length of 100 m. They are like telephone cables and are easier to handle. Some high-speed implementations of Ethernet used optical fibre. Then, the system specification is like 100BaseF. Optical fibres cover a distance of about 2000 m.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

8) What is the operating speed and the maximum distance covered by 20Base3 Ethernet system?

....

....

....

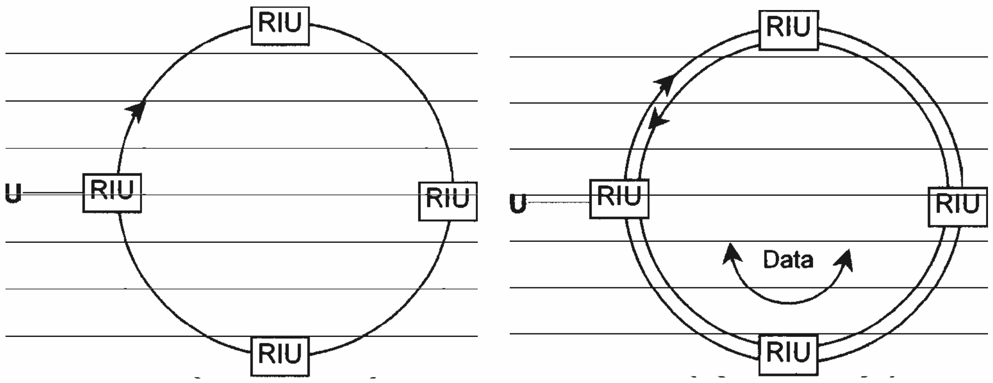
8.7 RING TOPOLOGY

In ring topology, a physical ring is formed by making point-to-point connection between computers. The computers themselves may not physically appear to be in the form of a ring, but electrically they form a ring. For example, two computers placed in adjacent rooms may be part of the ring. There is a circular communication path. Ring topology may be built around a single ring or two rings (dual ring).

Ring topologies are depicted in Fig. 8.6. The equivalent of tap in bus topology is Ring Interface Unit (RIU) in ring topology. User computers are attached to the RIUs. Unlike passive bus taps, RIUs are active units. Being active units, their failure rate is higher than passive taps. If a bus tap fails, only the concerned computer is affected. But if a RIU fails, the entire ring operation is affected. Hence, special considerations are required in ring topology to handle failures. In fact, failure management complicates the ring design and for this reason ring topology is not very popular.

Single ring topology shown in Fig. 8.6(a) usually uses bi-directional medium like copper wire. In case a segment of the ring or a RIU fails, the ring is folded back by the two end RIUs and the ring form of functioning continues. Dual ring configuration shown in Fig. 8.6(b) is generally adopted in the case of optical fibre design. As you may know, optical communication is naturally unidirectional as light that acts as the carrier of information is launched at one end of the fibre and received at the other. In dual ring, information travels in opposite directions in the two rings.

As in the case of bus topology, it is difficult to implement and maintain a ring structure physically. Hence, logical ring structure is often implemented using a ring hub in physical star topology. The logical topology in a ring network is called token passing ring or simply token ring.



(a) Single ring (b) Dual ring

Fig. 8.6: Ring Topology

A token circulates on the ring when it is idle. Token is a special bit pattern like 01111111 (zero followed by 7 ones). It is ensured that this pattern does not occur in the data transmitted by the computers. A special coding scheme is used to prevent the occurrence of token pattern in the data. A computer that has data to transmit seizes the token and starts transmitting the data. Imagine a circular formation of people around a round table participating in a group discussion. A ball is given to the group. The ball is passed around from one person to the next. The rule is that a person can speak only when he/she holds the ball. When a person wants to speak, he/she retains the ball when received and starts speaking. Once finished speaking, he/she passes the ball to the next person in the circular formation. The token ring scheme functions in a similar fashion. Two observations are important. First there is no collision in this scheme. Hence, the token ring functions more efficiently than Ethernet. However, Ethernet is more popular. Second, every station gets a chance to transmit as the token goes around. In Ethernet this is not the case. An unlucky station may keep on colliding again and again and may get a chance to transmit for quite some time.

On the ring, when a station starts transmission, the stations (computers) downstream listen to the transmission and monitor the destination address. If the destination address does not match one's own address, it passed to the next station as it is. When the data reaches the destination computer, the station copies and drains (takes away) the data from the ring. Thus a connection is established between the source and destination stations. After the source station has transmitted all data, it reintroduces the token on the ring. The token may now be seized by another station that has data to send. Obviously, the station next to the source station is the first one that can seize the token. Because of the use of token, the logical topology is called token passing ring. In optical fibre implementation, the logical topology is called Fibre Distributed Data Interface (FDDI). Ring networks operate at speeds of 10 Mbps to 1000 Mbps.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

9) Discuss the problem that would arise if a station on the ring seizes the token but fails to reintroduce the same on the ring after completing the data transmission. Suggest a mechanism to overcome such a problem.

....

....

....

....

8.8 MESH TOPOLOGY

Mesh is a complex interlaced structure realised by a bunch of point-to-point interconnections. Computers are interconnected without any geometric shape in mind. They are connected depending on the demand. They are somewhat like fully connected networks with some links missing. Therefore, they are sometimes referred to as partially connected topology. Mesh topology is illustrated in Fig. 8.7. In this topology, some of the nodes of the network are connected to more than one node by point-to-point links.

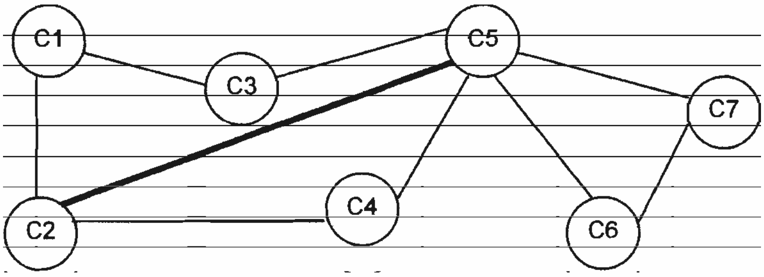


Fig. 8.7: Mesh Topology

Mesh is the actual topology that connects the wide area Internet all over the world. The nodes on the Internet invariably have mesh connectivity, i.e. multiple point-to-point links. Each node is connected to more than one node. The point-to-point links are established based on two considerations:

- Traffic between nodes
- Redundant routes.

Wherever heavy traffic is envisaged between two nodes, a direct link is established. Such links are sometimes called high usage routes. For example in Fig. 8.7, the link between node 2 and 5 is a high-capacity link to cater for heavy traffic between the two nodes. In addition, there must be at least one alternative route between a given source and destination i.e. a primary route and a secondary route. For example, the secondary route between C3 and C5 is via C3-C1-C2-C5. The secondary route may be used when C3-C5 link is broken. Often, there are many alternative routes between a source and destination. In such a case, the source and intermediate nodes must have some kind of intelligence to make a routing decision to select the best route at a given time. Most of the nodes on the Internet are routers that are capable of selecting the best possible route. Routing decisions take a definite amount of time. Hence, more is the number of intermediate nodes more is the time taken for the information to reach the destination. Usually, the shortest path with minimum number of intermediate nodes is chosen as the primary route. Only when that route is heavily loaded with traffic or unavailable for some other reason, an alternative route is chosen. We discuss routing and routing algorithms in more details in Sections 8.13 and 8.14 respectively.

8.9 TREE TOPOLOGY

Tree topology is also called hierarchical topology. In this topology, there is a clear hierarchy amongst the nodes. This is similar to a hierarchical structure in an organisation where there is a Chief Executive Officer (CEO) at the top, many senior level executives under him/her, junior executives reporting to seniors and so on. There are levels of responsibility and a clear reporting structure. Tree topology is modelled along the same lines. It is shown in Fig. 8.8. Strictly speaking, the structure is an inverted tree with the root node at the top and the branch and leaf nodes below the root.

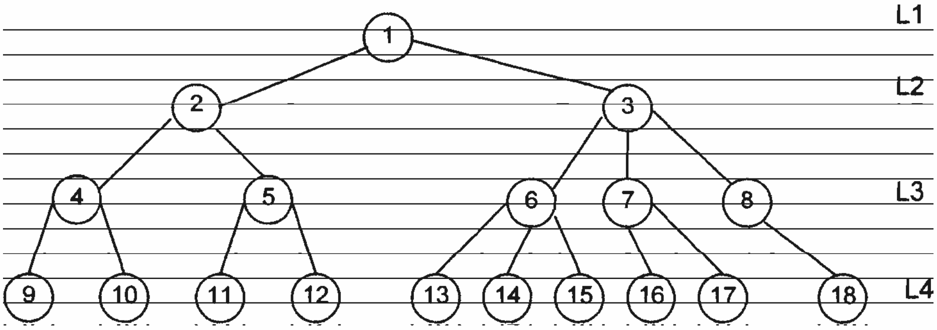


Fig. 8.8: Tree Topology

Four levels are shown in Fig. 8.8. The number of levels in a tree is called its depth. The top level node (Level 1) is called the root node and the bottom level nodes (level 4) as leaf nodes. The intermediate level nodes are called branch nodes at the designated levels. Each node except the root node has one point-to-point link connecting itself to the higher level node. Each node except leaf nodes has as many point-to-point links as there are braches attached to it. Leaf nodes, being at the bottom level have no branches. There is an interesting relationship between the number of nodes and the number of point-to-point links in a tree. The number of links is always one less than the number of nodes. Verify this in Fig. 8.8.

The number of branches that emanate from a node is called the branching factor (BF) of that node. If the branching factor is uniformly two for all the nodes, then the tree is called a binary tree. In Fig. 8.8, the left portion of the tree is shown to be binary. The tree itself is not binary as there are portions with branching factors that are not two. If all the nodes (except leaf nodes) in a tree have the same branching factor, then the same may be called the BF of the tree. If the BF of a tree is one, then the tree reduces to linear topology. The extreme right portion of Fig. 8.8 comprising nodes 8 and 18 represents the linear topology.

There is strict hierarchy of interaction amongst the nodes. Two nodes at the same level emanating from the same branch node above interact through that branch node. For example, nodes 6 and 7 in Fig. 8.8 communicate via node 3. If the nodes are attached to different branches, then the communication proceeds by traversing up the tree as much as required. For example, the communication between nodes 9 and 12 takes place via the route 9-4-2-5-12.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

10) What is the branching factor of a leaf node?

11) How many nodes and point-to-point links are there in a binary tree of depth 5?

12) Why is it a tree topology is also called a hierarchical topology?

....

....

....

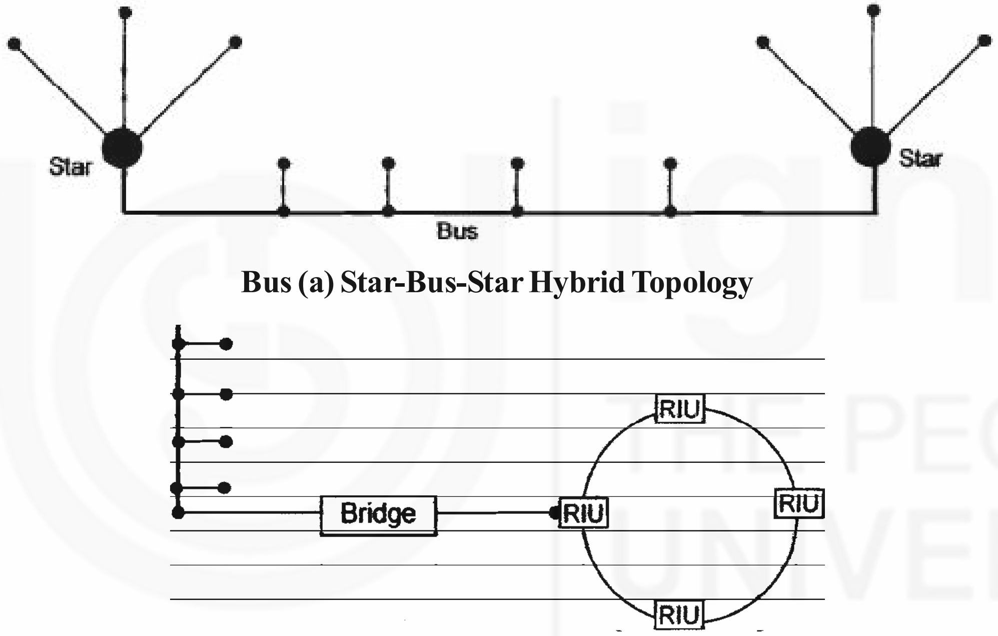
....

8.10 HYBRID TOPOLOGY

A hybrid topology is a combination topology in which two or more of the topologies discussed above coexist and work together. Figure 8.9 shows two example hybrid topologies. A large variety of hybrid topologies are possible. In Fig. 8.9(a), two star topologies are interconnected by a bus topology. This implementation is typical in campus networks like in a university. Each department may have star implementation while a bus or ring network may interconnect the departments. In general, such an implementation is adopted wherever the facilities that need to be interconnected are dispersed. For example, an office that is situated in different floors of a multi-storey building may use a hybrid structure. Computers in each floor may use Ethernet hub based star structure and the different floors may be interconnected by a bus structure. It is important to note that the entire set up works on the logical topology of Ethernet. Hence, interconnection of star with bus is seamless with no intermediate device.

Figure 8.9(b) shows a hybrid topology interconnecting a bus topology and a ring topology. In this case an intermediate device called a bridge is required. This is because the two topologies implement different logical protocols, viz. Ethernet and token ring. Bridge is an intelligent device. It implements two different logical topologies or protocols. In Fig. 8.9(b) it implements Ethernet on one side and token ring on the other. It appears like Ethernet NIC for the bus topology and as a token ring station attached to a RIU on the ring topology side. It is capable of recognising addresses at the logical level and mule *packets 1mm* one LAN to the other, it converts (reformats) packets of one protocol to that of another.

In some organisations, same topology is implemented in segments in geographical locations that are far apart All segments together are considered as one LAN.



Bus (a) Star-Bus-Star Hybrid Topology

(b) Bus-Ring Hybrid Topology

Fig. 8.9: Hybrid Topologies

In such cases, the signal from one segment to another traverses a long distance. In the process the signal level may be attenuated and may become too low to be recognised properly at the destination. To avoid this, devices called repeaters are used in between the segments. Repeaters are non-intelligent devices. They just amplify the signal level and make the signals strong so mat destination may recognise them property. In some cases, bridges may be used in place of repeaters for better management of LAN segments. Since, the bridges transform packets from one topology level to another, they automatically amplify the signals.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

13) In a university, LIS department has implemented star Ethernet LAN and the computer science (CS) department bus Ethernet LAN. The two departments are

in two different buildings that are far apart. How would you interconnect the two LANs?

14) In Question 13, how would your interconnection strategy change if the CS department had ring LAN? Explain the function of the device used.

15) Compare the features of bridges and repeaters.

....

....

....

....

8.11 MEDIA ACCESS CONTROL PROTOCOLS

Ethernet and token ring are two media access control (MAC) protocols that we have already studied briefly. In this section, we learn more details about them. Bus is a medium and so is ring, single or dual. Many computers are connected to these media. Any of the computers can access the medium that it is connected to. Since many computers access the same medium, we need some kind of protocol (a set of rules) so that there is an orderly access to the medium. Actually the access takes place in a controlled manner. Hence, the nomenclature MAC protocol is used.

In bus topology, Ethernet protocol is used. In ring topology, token passing ring protocol is used. Both these protocols are called multiple access protocols as they define sets of rules for multiple computers to access a medium. In hub based star implementations, the protocol depends on the type of hub used. The hub may be Ethernet hub or token ring hub.

Let us first see more details of Ethernet. As you already know, the Ethernet protocol rules are summarised in the acronym CSMA/CD. Consider the case when there is an ongoing transmission on the bus. The protocol must ensure that no new transmission starts at this stage. If it does, it will collide with the ongoing transmission and both transmissions will fail. No new transmission while there is an ongoing transmission is ensured by carrier sense (CS) mechanism. You may be aware that carrier refers to a high frequency transmission that carries the information signal. We have carrier frequencies in AM/FM radio broadcast. For example, 92.5 MHz is the carrier frequency of a FM radio station. The music signal is superimposed on this frequency and broadcast. Similarly, in LANs, information bits are superimposed over a carrier. The presence of carrier on the bus implies that there is a live transmission on the bus. Hence, any computer that has data to transmit will first sense the bus to see if the carrier is present. This process is aptly described as 'Listen before talking'. The rule specifies that a station can transmit only if there is no carrier present on the bus, i.e. the bus is idle or free.

There may be more than one station ready to transmit at any point of time. All such stations will start transmitting as soon as they find the bus idle. There is multiple access (MA) that results in collision. Let us what a collision is and how it is detected. If a computer is transmitting bit '0' and that gets changed to bit T on the bus or vice versa (bit T changes to bit '0'), then a collision is said to have occurred. The transmitting station is continuously listening to its own transmission on the bus and detects such a collision. When it finds a T changed to '0' or vice versa, it knows that a collision has occurred. Continuous comparison of what is transmitted and received on the bus is the collision detection (CD) mechanism. This process is often called as 'Listen while talking'.

Once a collision is detected, what do the stations do? They wait for random times and retransmit again. One station may wait for one millisecond, another for two and so on. Since the wait time is random for each station, it is likely that each station waits for different time and then attempts retransmission. The collision is resolved in this manner. Since the wait time is random for each station, it is possible that two or more stations wait for the same random time. In such a case, there will be a collision again during the retransmission attempt. If this happens, the same process of waiting for random times is repeated until the collision is finally resolved.

Now let us look at the details of token passing ring protocol. This protocol is relatively simple when compared to Ethernet. However, certain types of failures need to be taken care of in this protocol. As mentioned earlier, a token circulates on the ring whenever there is no data transmission on the ring. A token is a particular bit pattern and is recognised by this pattern. When a station has data to send it seizes the token and starts its own transmission. By seizing we mean that the station changes the token bit pattern such that it is no longer recognised as token. Instead, the pattern corresponds to one that indicates the beginning of transmission of data. Following this pattern, the destination and source addresses are sent. Whenever a station sees the beginning pattern, it examines the destination address to determine if the data is destined for itself. If so, it copies the data. When the data transmission is complete, the source station reintroduces a token on the ring. If any other station has data to send, it follows a similar procedure. To avoid a station holding the ring for a very long time, an upper limit is set for the size of data packet that can be transmitted at a time. If a station has large data to send, it needs to break down the same into a number of packets and transmit. After sending one packet, the station will have to wait for its turn to get the token. Only when it gets the token again, it can transmit another packet. This ensures that every station gets a fair chance to transmit.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

16) What is a carrier in LAN?

17) Find out and record the carrier frequency of a nearby AM radiobroadcast station.

18) What are the reasons due to which an Ethernet station experiences a collision during a retransmission attempt?

19) What is a token in token ring protocol?

20) In some token ring implementation, the destination station, instead of source station, reintroduces the token. What difference does it make?

....

....

....

....

....

....

8.12 ADDRESS RESOLUTION

In data networks, destination addresses have different formats at different levels. This is required for easy implementation of a complex system. For example, at user level, we need easy to remember addresses like names. Such user addresses are not transmittable as such. The network needs numerical addresses specified in bits. The name addresses provided by the user are decoded and the user data packet is encapsulated with decoded numerical addresses. The process of converting the addresses from one format to another is known as address resolution. Encapsulation takes place at several levels in a hierarchical structure for packet transmission. Three addresses and two levels of encapsulation are important to understand although there may be as many as six addresses encapsulated often. We study the important ones now.

Consider the case of a user sending e-mail. He/she uses a destination address something like *james_bond@mgm.co.uk.* The sending computer cannot use this address as it is because every computer on the Internet is addressed by a 32/128-bit number. The server mgm.co.uk is known on the network by a number assigned to it and not by its alphabet description. The 32/128-bit number is called Internet Protocol (IP) address. IP address has two versions: IPv4 and IPv6. IPv4 uses 32-bit address and IPv6 128-bit. IPv4 has been in use for a very long time, over 30 years, and most of the computers on the Internet have IPv4 addresses as of now. IPv6 has been introduced recently. Over the years, IPv6 is expected to replace IPv4 addresses.

The first step in packet transmission is to resolve the string address to numerical IP address. This is done with the help of Domain Name Servers (DNS) that are located in a hierarchical structure throughout the Internet. DNS have a table of string addresses with the corresponding numerical IP address. Usually, the table in one DNS is only partial and the entire set of addresses is covered by the complete hierarchical structure of the DNS. To start with, the sending computer accesses the nearest DNS by sending it the character string address provided by the user. If the DNS has the particular string stored in its table, it returns the numerical IP address. Otherwise it accesses another DNS that is in the hierarchy. This process is continued until a DNS is found that has the particular string address and its corresponding numerical address in its table. This process constitutes the first level of address resolution.

Once the sending computer receives the numerical IP address, it encapsulates the user message with numerical addresses. The numerical address received from DNS is used as the destination address and its own numerical address as the source address. This is the first level of encapsulation. The next level takes place in LANs.

You are aware of the use of NIC in bus LANs and RIU in ring LANs. These interface units have their own unique addresses assigned by the manufacturer. They are accessed by these addresses only. These addresses are 48-bit long. The destination and source stations are identified by 48-bit interface addresses on the LAN. The computers in which the interface units are housed are identified by their IP addresses. IP addresses are not recognised by the interface units. We now have two addresses: 32/128-bit IP address for the computer and the 48-bit interface address. We need to resolve the destination IP address to destination interface address before the data transmission can take place on the LAN. This is the second level of address resolution, in bus LAN, this address resolution is done by using a protocol called Address Resolution Protocol (ARP). Let us now see how ARP works. In all LANs, there is provision to broadcast information. This is usually done by reserving a special broadcast address. On bus LANs using ARP, the sending computer broadcasts the destination IP address received as part of the first-level encapsulated packet. All other stations (NIC) read this broadcast. Whichever NIC is attached to the computer that has this IP address responds in reply. The sending computer now knows the NIC address to which the user information should be forwarded. It now encapsulates the user packet with the received NIC address as destination and its own NIC address as source and transmits the packet. This is the second level of encapsulation.

Thus address resolution and encapsulation are two important functions carried out in data networks at different levels.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

21) Why do we need address resolution in data networks?

22) What is the function of DNS?

23) How does ARP resolve IP addresses?

....

....

....

....

8.13 ROUTERS

A router is a device that forms one of the basic building blocks of Internet. Internet cannot function without routers. You are already familiar with repeaters and bridges. Router is a higher-level device that performs the functions of a bridge and a repeater and more. The primary function of router is to direct the user packets encapsulated with IP addresses in the direction of the destination. In this sense, the routers are much like telephone exchanges for the data networks. Telephone exchanges route the phone calls to the appropriate destination. Similarly, routers forward the data packets towards the destination. The telephone exchanges examine the number dialled to determine the destination. Routers examine the destination IP address in the incoming packets to decide the destination route. For this purpose, routers maintain what are called routing tables that contain entries relating to a destination addresses and the corresponding output that should be used to route the incoming packet.

Consider a user connected to a bus LAN in an institution in Delhi wanting to send a file to a user connected to a bus LAN in an institution in New York. Both LANs are connected to Internet via routers. The routers are connected to the LAN on one side and to the internet on the other as shown in Fig. 8.10. They are recognised both as a local machine and an Internet machine. They have unique Internet IP addresses as wet as local LAW addresses.

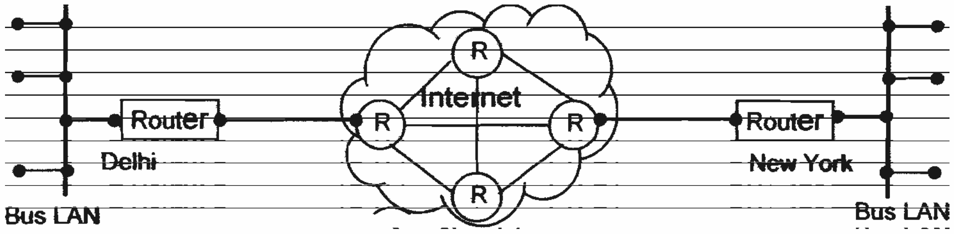


Fig. 8.10: Routers Use

There are also routers on the Internet that connect to other Internet routers in a mesh fashion. Whenever the destination IP address in a packet originating on the LAN does not point to a machine on one's own local network, as is the case in our example of Delhi sending to New York, the packet is forwarded to the router. This is done by a protocol called Transmission Control Protocol (TCP) mat runs on every LAN machine. We team about TCP HI Unit 9. The router forwards the same to another appropriate router on me Internet The packet travels via a number of routers on the Internet until it finally reaches the concerned router in New York that semis the packet to me appropriate machine on the LAN.

As you may observe, on me internet we have a number of possible routes that can be taken to reach the packet to New York. A typical route in our case is Delhi-Mumbai-Arnsterdam-London-New York. At every stage, the concerned router makes a routing decision as to which of the output ink must be chosen to forward the packet. The routing decision is taken according to the routing algorithm (software program) used by the router. Routing algorithm are discussed in the next section.

The routers shown at Delhi and New York institutions may also have multiple output links connecting to different routers on the Internet. It means that the concerned institutions have more than one Internet link. For example, if the packet were destined to Japan, the route chosen could be different if multiple paths were available.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

24) "Routers to data networks are Bee exchanges in telephone networks" Discuss.

....

....

....

....

8.14 ROUTING ALGORITHMS

As you may be aware, an algorithm is a step-by-step procedure to execute a task especially in a computer. Software programs implement algorithms to perform various tasks. Routing algorithms are procedures to make routing decisions. Routers execute routing algorithms to make routing decisions. Routing algorithms may be placed under two broad categories:

- Adaptive or dynamic algorithms
- Fixed or static algorithms

Dynamic algorithms adapt themselves to changing traffic conditions and network availability. For example, traffic may suddenly increase in a particular segment. As a result, tong queues of packets may build up stowing down the delivery to the destination. An adaptive algorithm may find an alternative route that may be longer but faster. Static algorithms use fixed routes for relatively long time durations. There could be more than one fixed route defined for a given destination in order of priority. Static algorithms do not monitor traffic conditions or the time to deliver. They are relatively easier to implement when compared to dynamic algorithms. They also need less processing power, i.e. CPU time.

Routing algorithms are designed to satisfy certain performance parameters. Some of the important parameters are:

1) Minimum delay for delivery
2) Minimum number of hops to reach the destination
3) Robustness
4) Stability
5) Fairness

Minimum delay may be local or global. By local we mean that the packet does not stay in the router for long. The output queue small and the packet leaves the router quickly. By global we mean tile delay in reaching tire destination. A packet may leave a router quickly but may get stuck later, tit such a case, it is better to route the packet by an alternative route even though the local delay may be longer.

A packet traversing a router is said to have done a hop. In other words, a packet hops from router to router on its way to the destination. As you are aware, every router examines the destination address in the header portion of every incoming packet. There is computational time overhead associate with this activity. Hence, it is desirable to have minimum number of routers or hops on the way to the destination.

Robustness means the ability to reach tile destination even when part of the network fails. A router should not forward a packet to a dead router on the way. In that case, the packet would never reach the destination. A robust routing algorithm would ensure that a packet is delivered to the destination at all costs. In other words there is guaranteed delivery.

Stability refers to the ability to deliver the packet as quickly as possible without the packet wandering here and mere. Sometimes loops may be formed in a network that packets may go round and round without moving forward towards the destination. A loop in the network is an unstable condition. Routing algorithms must ensure that no loops are formed. And if formed, they must be detected quickly and remedial action taken.

Fairness means delivery in a reasonable time for all types of packets. Sometimes, networks may receive high priority packets. In such case, other packets are delayed and the high priority ones are forwarded first. But such an action should not result in low priority or normal packets being delayed indefinitely. This criterion is called fairness.

There are many routing algorithms that are designed to implement one or more of the performance parameters discussed above. Some of the important algorithms are:

- Shortest path routing
- Flooding
- Hierarchical routing
- Broadcast routing

A shortest path may be determined based on one or more of the following factors:

- Link length
- Minimum local delay

- Minimum global delay
- Minimum cost
- Minimum number of hops.

Shortest path algorithm is one of the most popularly used ones. Flooding is a robust algorithm and is often used in military applications where delivery is critical. It is also very simple to implement. In flooding, an incoming packet is forwarded on all outgoing links except the one on which it arrived. The idea is that the packet will definitely be delivered to the destination via one of all the possible available routes. Hence, the algorithm is robust. Flooding generates a vast number of duplicate packets and can choke me network unless controlled. It can also cause loops easily. One of the reasons why a packet is not forwarded on the incoming link is to avoid looping. A variation of flooding is called controlled flooding. Here, a router remembers at the packets that it has forwarded. If the same packet returns to it, it is discarded straight away or after forwarding one or two more times. This is the control exercised.

Hierarchical routing maps the network in a hierarchy and forwards the packets via the appropriate hierarchical route. Broadcast routing is like flooding where the packet is sent even on the incoming route.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

25) Distinguish between dynamic and static routing.

26) What is robustness in routing? Which algorithm is designed to meet robustness requirement?

....

....

....

....

8.15 SUMMARY

This unit deals with four basic aspects of data networks, viz. topology, media access control (MAC) protocols, address resolution and routing. Data networks are designed in different geometrical shapes. The geometrical shape of the network is called the topology of the network. There are a variety of topologies in use. This unit discusses the main topologies. These include star, bus ring, mesh, tree and hybrid topologies. The way data travels in a network need not necessarily correspond to the physical topology of the network. Data travel often defines its own topology and this is called the logical topology of the network. Logical topology is generally defined by way of set of rules called protocols. Two important protocols, viz. Ethernet and token passing ring have been discussed in detail. The merit of implementing these protocols using star physical topology has been explained. Certain bask: devices that are required to build topologies and interconnecting them have been described. These include hubs, switches, repeaters, bridges, and routers.

All data packets that traverse a network carry the source and destination addresses.

These addresses use different formats at different levels. When moving from one level to another, addresses have to be translated from one format to another. This function called address resolution is required at different levels in networks. Three important address formats and the techniques for their resolution are discussed in detail. User level name addresses are resolved to IP addresses by domain name servers (DNS). IP addresses are resolved to hardware interface unit addresses by address resolution protocols (ARP). After addresses are resolved, encapsulation is required before data transmission.

Finally, the functioning of routers and the features of some of the important routing protocols have been discussed. A router is one of the basic building blocks of Internet. Internet cannot function without routers. Local routers connect to LANs on the one side and the Internet on the other. Internet routers connect to other routers in the Internet. The connectivity is such that there are at least two routes to reach a destination. Routers execute routing algorithms to make routing decisions.

Routing algorithms may be dynamic or static. Dynamic algorithms adapt themselves to changing traffic conditions and network availability. Static algorithms use fixed routes for relatively long time durations. Static algorithms are relatively easy to implement. Routing algorithms are designed to satisfy certain performance parameters. They include minimum delay, minimum number of hope, robustness, stability and fairness. Routing protocols that satisfy one or more of the performance have been discussed briefly. They include shortest path routing, flooding, controlled flooding and hierarchical routing algorithms. Shortest path algorithm is the most widely used one. It takes into factors like minimum local and global delay, minimum number of hops to the destination etc.

8.16 ANSWERS TO SELF-CHECK EXERCISES

1) Packet size is 2^{11} bytes. File size is 1 MB, i.e. 2^{20} bytes. Therefore the number of packets to be transmitted is $2^{20}/2^{11} = 2^{9}$, i.e. 512 packets.

2) The logical paths through which packets can travel from C1 to C7 are:

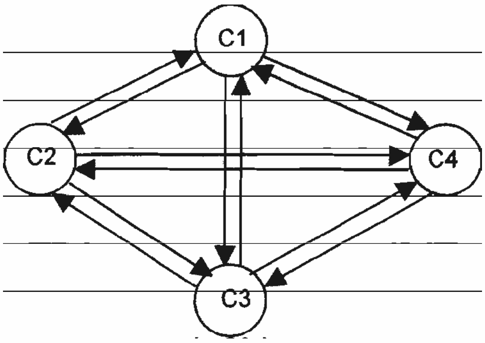
   - C1-C3-C5-C7
   - C1-C3-C5-C6-C7
   - C1-C2-C4-C5-C7
   - C1-C2-C4-C5-C6-C7

   There are a total of 4 logical paths constituting the logical topology of the network. The packet should not travel from C5 to C4, as it would result the packet going back to C1 and thus looping forever.

3) Yes. The packets of the same file may travel via different logical paths, as routing decision is taken for every packet independently. The problem that may arise is that the packets may arrive out of sequence at the destination. For example, consider a file having 10 packets. Packet 4 may be routed via longer route and Packet 5 via a shorter route. In such case, it is possible that Packet 5 arrives at the destination before Packet 4. The destination will have to take care of proper sequencing of the packets. For this purpose, whenever two or more packets of the same file are transmitted, the packets are tagged with a sequence number at the source so that the destination can sequence them properly. Packet numbering at the source and sequencing at the destination are taken care of high level protocols.

4) Half-duplex links carry information in only one direction. To support full-duplex communication, i.e. communication in both directions, we need two half-duplex links between every pair of computers. For a fully connected network with N computers, we need a total of N(N - 1)/2 full-duplex links or 2 x N(N -1)/2 half-duplex links. For 10 computers we need:

2 x 10(10 -1)/2 = 2 x (10 x 9)/2 = 90 half-duplex links.



5) Fibre optic links are one-way (half-duplex) communication links. For two-way (full-duplex) communication, we need two optical fibre links between ever source and destination pair. The fully connected topology for four computers is shown in the Figure below. There are 12 links in the network. The formula

N(N -1) links applies, i.e. 4 x 3 = 12.

6) The advantages of star topology are:

- Ease of implementation
- Ease of maintenance
- Ease of administration: connection, disconnection etc.
- It is a robust topology. If a link or a computer fails, the rest of the network is not affected
- Flexibility for implementing different logical topologies, i.e. protocols including Ethernet protocol, token ring protocol and a switch.

7) The different logical topologies (or protocols) that can be implemented by a hub are Ethernet and token ring.

8) 20Base3 Ethernet system means an operating speed of 20 Mbps (mega bits per second) and the maximum distance covered is 300 meters.

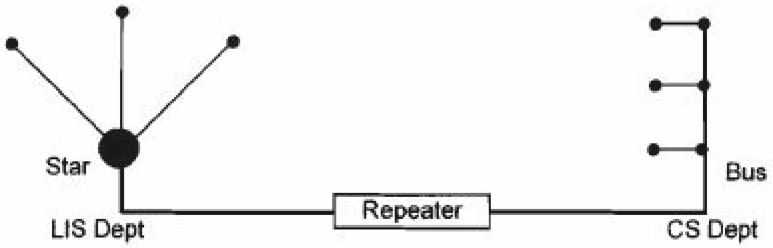
9) A station that seizes the token is expected to reintroduce the token on the ring after completing the data transmission. But it fails to so. In such a case, there will be no token on the ring. The token is said to have been lost. No other station can now transmit data and the ring is as good as 'dead'. Such a situation can be handled in the following way. When a station has data to send and finds no token or traffic on the ring for quite sometime, say 5 seconds, can introduce a new token on the ring. The new token can now be seized and new data transmissions can start.

10) Leaf nodes have no branches. Hence, the branching factor of leaf node is zero.

11) The number of nodes in a binary tree of depth 5 is (23 - 1) = 31. The number of point-to-point links is one less than the number of nodes, i.e. 30.

12) In tree topology, communication between adjacent nodes takes place by following a strict hierarchy. Hence, it is also called a hierarchical topology.

13) Both LIS and CS departments have Ethernet LAN, i.e. the same logical topology but different physical topologies. Since the departments are geographically far apart, signals will be attenuated. Hence, a repeater may be placed between the two buildings to boost the signal level as shown below.



14) If the CS department has a ring LAN, then the two protocols or the logical topologies are different. We need an intelligent device to transform packets from one format to another. Hence, we use a bridge instead of a repeater to interconnect the two departments. The bridge implements Ethernet protocol for the LIS department and the token ring protocol for the CS department. It performs necessary format conversion.

15) Comparison of repeaters and bridges:

Features	Repeater	Bridge
Signal amplification	Yes	Yes
Intelligent Device	No	Yes
Address Recognition	No	Yes
Packet Reformatting	No	Yes
Same Multiple Interfaces	Yes	Yes
Different Multiple Interfaces	No	Yes
Multiple Protocols	No	Yes

16) Carrier is a high frequency signal over which the data signals are superimposed.

17) AM broadcast frequencies lie in the range of 550 kHz to 1500 kHz. The AM Rajdhani channel in Delhi has a broadcast frequency of 666 kHz. Student is required to find out the broadcast frequency of an AM station nearby his/her city/town and record the answer.

18) There are two reasons as to why a collision may occur during retransmission attempt in Ethernet:

- The random wait time generated by two or more colliding stations might have been the same. For example, if two colliding stations generate random wait time as 0.1 sec by chance, then a collision will occur during retransmission attempt.

- A new station may join and start transmission at the same time when an old station is making a retransmission attempt.

19) A token is a specific bit pattern, say 7 ones and one zero (11111110). This pattern is unique such that it is not allowed to appear in the data.

20) Consider a token ring with 10 stations serially numbered. Let Station 2 seize the token and transmit data to Station 6. If Station 2 reintroduces the token, then the next station that would get the opportunity to transmit is Station 3. But if Station 6 reintroduces the token, then the next station that would get the opportunity to transmit is Station 7. Thus, the next station that gets opportunity to transmit changes in the two cases.

21) In data networks, destination addresses have different formats at different levels. This is required for easy implementation of a complex system. Address resolution is required to change addresses from one format to another.

22) Domain Name Server (DNS) resolves name string address provided by the user into a numerical IP address.

23) In all LANs, there is provision to broadcast information. On bus LANs using ARP the sending computer broadcasts the destination IP address of the packet. All other stations (NIC) on the LAN read this broadcast. Whichever NIC is attached to the computer that has mis IP address responds in reply. The sending computer now knows the NIC address to which the user information should be forwarded. Thus an IP address is resolved to NIC address.

24) The primary function of a router is to direct the user packets encapsulated with IP addresses in the direction of the destination. In this sense, the routers are much like telephone exchanges for the data networks. Telephone exchanges route the phone calls to the appropriate destination. Similarly, routers forward the data packets towards the destination. The telephone exchanges examine the number dialled to determine the destination. Routers examine the destination IP address in the incoming packets to decide the destination route.

25) Dynamic algorithms adapt themselves to changing traffic conditions and network availability. For example, traffic may suddenly increase in a particular segment. As a result, long queues of packets may build up slowing down the delivery to the destination. An adaptive algorithm may find an alternative route that may be longer but faster. Static algorithms use fixed routes for relatively long time durations. They do not monitor traffic conditions or the time to deliver. Static algorithms are relatively easier to implement when compared to dynamic algorithms. They also need less processing power, i.e. CPU time.

26) Robustness means the ability to reach the destination even when part of the network fails. A router should not forward a packet to a dead router on the way. In that case, the packet would never reach the destination. A robust routing algorithm would ensure that a packet is delivered to the destination at all costs. In other words there is guaranteed delivery. Flooding and controlled flooding are the routing algorithms designed to implement robustness.

8.17 KEYWORDS

Address Resolution : Given an address in one format, the process of obtaining the equivalent in another format

Algorithm : A step-by-step procedure for performing a task in a computer program

ARP : Address Resolution Protocol used to resolve IP addresses to NIC addresses in Ethernet LAN

Bridge : An intelligent device capable of interconnecting two dissimilar networks

Bus : A laid out open cable to which LAN computers are connected

Carrier Sensing : The process of checking whether a transmission is in progress on the bus of Ethernet LAN

Coaxial cable : A cable with an inner and an outer conductor placed coaxially and separated by insulating material. The overall structure is covered by a sheath

Collision Detection : The process of finding out if two or more stations are transmitting at the same time on one common medium

DNS : Domain Name Server used to resolve name string addresses into IP addresses.

Ethernet : A protocol used for exchange of information in bus or hub LAN

Flooding : A routing protocol that is robust

Hub : A device used in star configuration implementing one of the LAN protocols

Hybrid Topology : A network topology comprising two or more dissimilar network segments

LAN : Local Area Network

Media Access Control : A technique or protocol for controlling access to a common medium by multiple computers

Mesh : A network topology where computers are interconnected without any particular geometric shape in mind.

Multiple Access : Refers to where multiple computers access a common medium

NIC : Network Interface Card used in Ethernet LAN

Packet : A segment of information with specified length and structure

Protocol : A set of rules that govern the exchange of packets between computers

Repeater : A passive device that amplifies the signal level

Ring : LAN topology where computers are connected in the form of a ring

RIU : Ring Interface Unit used in ring LAN

Robustness : The ability of a routing algorithm to reach the destination even when part of the network fails

Router : A fundamental networking device used extensively in Internet to route packets towards their destination

Stability in Routing : The ability of routing algorithm to deliver packets as quickly as possible without the packets wandering here and there

Star : LAN topology configured with a central unit called hub

Switch : Networking device that switches incoming packets to output lines leading to their destinations

Token : A specific bit pattern use in token ring LAN

Token Ring : A LAN protocol for multiple access in the common ring

Tree : Network topology configured in the form of an inverted tree with the root at the top and the branches spanning downwards.

8.18 REFERENCES AND FURTHER READING

Mansfield, Kenneth C and Antonakos, James L. *An Introduction to Computer Networking*. New Delhi: Prentice Hall of India, 2002. Print

Tanenbaum. A. S. *Computer Networks*. 4th Ed. New Delhi: Prentice Hall of India, 2002. Print

Viswanathan. Thiagarajan. *Telecommunications Switching Systems and Networks*. New Delhi: Prentice Hall of India, 2008. Print

www.en.wikipedia.org/wiki

UNIT 9 COMMUNICATION PROTOCOLS AND NETWORK ADDRESSING

Structure

9.0 Objectives

9.1 Introduction

9.2 What are Protocols?

9.3 Computing Protocols

9.4 Communication Protocols: General Concepts

9.5 Common Communication Protocols

9.6 Basic Communication Protocols: IP, UDP, TCP

9.6.1 Internet Protocol (IP)

9.6.2 User Datagram Protocol (UDP)

9.6.3 Transmission Control Protocol (TCP)

9.7 Client-Server Architecture

9.8 Application Level Communication Protocols: FTP, Telnet

9.8.1 File Transfer Protocol (FTP)

9.8.2 Remote Login (Telnet)

9.9 Switching Level Convergence Protocol: ATM

9.10 Multi Protocol Label Switching: MPLS

9.11 Telephone and Mobile Numbering

9.11.1 Landline Telephone Numbering

9.11.2 Mobile Phone Numbering

9.12 Number Portability

9.13 IP Addressing: IPv4, IPv6

9.14 Web Communication Protocols: HTTP, WAP, LTP

9.15 Summary

9.16 Answers to Self-check Exercises

9.17 Keywords

9.18 References and Further Reading

9.0 OBJECTIVES

After going through this Unit, you will be able to understand and appreciate:

- What are protocols;
- Difference between computing and communication protocols;
- Need for communication protocols;

- Difference between connection-oriented and connectionless protocols;
- Basic packet transfer protocols like IP, TCP and UDP;
- Most widely used network computing architecture: Client-Server;
- File transfer and remote login application protocols like FTP and Telnet;
- Convergent cell switching in ATM networks;
- Fast routing technique using labels: MPLS;
- Numbering schemes used for landline and mobile phones;
- How network computers are addressed world over;
- Details of IPv4 addressing scheme;
- IPv6 features briefly; and
- Web access protocols for both wired and wireless networks.

9.1 INTRODUCTION

Quest for new knowledge is the central theme of human existence. All of us, whether we realise or not, are in the process of acquiring new knowledge all the time. When we ask a question, we are seeking knowledge. When we answer a query, we give information to the person posing the question. When a person assimilates the given information, we say that the person has acquired knowledge. Knowledge is spread via information that is communicated from one person to another in some form: oral, writing etc. Thus, knowledge, information and information communication are three entities that are closely inter-related.

It is often said that we are in the information age. In the last about six decades, information in the world has been growing at an exponential rate, i.e. doubling every 10 years. Information Communication Technology (ICT) has grown leaps and bound in the last 30 - 40 years. Instant transfer of information from one part of the world to any other part is a reality today. Underlying this development is the convergence of computer and communication technologies. This convergence process started in late 1960s and has led to the development of worldwide computer network that is now known popularly as **Internet.** A large number of home and office local area networks (LANs) and innumerable personal computers all over the world have been interconnected to form Internet. Hence, it aptly said that Internet is a **network of networks.** Information travels in the form of data packets on Internet and hence it is also called a **data network.** Data packets are of fixed length, say 2048 bytes, i.e. 2" bytes. Long messages are broken into as many packets as required before transmission. Because of packet-based transmission, the Internet also carries the nomenclature **Packet Data Network** (PDN). Since Internet is an open public network, another related nomenclature that is used sometimes is **Packet Switched Public Data network** (PSPDN). Internet is not limited to its presence only on the land but is also in ships at the seas and in planes in the air.

United Nations today has 192 countries of the world as its members. Almost all these countries have Internet connection in place. About 200,000 LANs are connected to the Internet. Over 1.5 billion people, i.e. a quarter of the world population has access to Internet. With the evolution of Internet, our life-style is changing. A number of our day-to-day activities are being carried out on the Internet. Clearly, the society is evolving towards a networked community with electronic information as the central commodity.

One might term the society of the 21st century as the **Networked Electronic Information Society** (NEIS). It is a society in which activities are centred on networks and the main commodity on the networks is electronic information in digital form.

It is important to realise that with alt its massive presence, Internet is still evolving. Today's Internet services are predominantly text and data oriented with only sprinkles of graphics, still pictures and slow motion video. Experience shows that Internet is slow for many network applications. Internet is basically designed for data transport. Real time services like voice and video transmissions experience serious quality problems. The key to the solution of current Internet problems lies in building **Global Information Infrastructure** (GII) that would have adequate capacity and efficiency to support full-scale services including high quality audio and motion video and high-resolution graphics envisioned for NEIS.

Information exchange between computers that are connected to a massive worldwide network cannot happen without standard procedures and sets of rules that govern such an exchange. A comprehensive collection of such standard sets of rules and procedures are called **communication protocols.** Furthermore, every entity on the Internet needs to be identified uniquely. This is done by assigning a **network address** to each entity. Communication protocols and network addressing are the subject matter of this unit.

9.2 WHAT ARE PROTOCOLS?

Let us start understanding protocols. The word protocol has different connotation under different circumstances. In governments, protocol means a strict official procedure in state affairs and diplomatic occasions. For example, in India the President is the Head of the State and there are protocols that govern his/her participation in state functions. These protocols specify how and where the President will be seated, who would accompany him/her, how would the dignitaries be introduced etc. In other words, they specify the accepted code of behaviour in particular situations. They may cover aspects about appearance (dress code), ways of greeting, conversation, and eating manners. All these rules help people successfully communicate and work together.

In inter-governmental dealings, the word protocol is used to denote the original draft of a diplomatic document containing especially terms of a treaty agreed to in a conference and signed by the parties concerned. You might have heard of Kyoto Protocol, a document that spells out the terms to be adhered to by the signatories for controlling and reducing carbon emissions in the world.

In science, a formal record of scientific experimental observations is often called a protocol. Procedures for carrying out scientific experiments or a record of the course of any medical treatment are also known as protocols.

The word protocol is used extensively in computers and communications as well. In computers, protocols deal with interaction between processes, exchange of messages etc. A process, as you may know, is a program in execution. In communication, protocols deal with signalling, switching, routing, forwarding, error control, monitoring, and recovery procedures in the exchange and transmission of information across entities in a network. A fundamental difference between the two is as follows. While computing protocols define rules for communication among processes within a computer, the communication protocols define rules for communication among computers. Both of them, however, deal with exchange of information. Protocols are generally software programs that implement the rules for communication. Some protocol functions are implemented in hardware, particularly those dealing with the movement of bits and bytes. In Section

9.3, we briefly discuss computing protocols and study communication protocols in greater detail in later sections.

9.3 COMPUTING PROTOCOLS

Computing and communication protocols together define sets of rules and procedures that govern all the information management functions. With electronic information being the central commodity in NEIS, information management functions become the core of technological capability in networks. There are seven functions of electronic information management that are important:

1) Generation

2) Acquisition

3) Storage

4) Retrieval

5) Processing

6) Transmission

7) Distribution

Computing and communication protocols govern all these functions. In general, information is generated by human thought process, human acts and by happenings in nature. Human intellectual activity is creative and intuitive and hence may not be amenable to protocols. Whether technology generates information is a debatable point. When data is processed in a computer, the output is considered as information. In that sense, it may be said that computers generate information. But the basic data comes from nature or human activity. However, machine generation of information can be governed by protocols.

Among the other functions, storage, retrieval and processing fall in the realm of computing protocols. The remaining functions, viz. acquisition, transmission and distribution fall in the class of communication protocols. Transmission and distribution functions may be collectively called as information *dissemination.* Transmission refers to bulk transfer between two main points. Distribution refers to transfer to end points like user computers or terminals.

Computing protocols are relatively a recent development. As you are aware, information processing, storage and retrieval are functions performed by application processes. You are familiar with applications like word processing, spread sheet, power point presentation and data base management. Computing protocols deal with information storage, retrieval and exchange among these applications. For example, how do we import information from word files into spreadsheets or vice versa? Or how do we import information from word files to power point presentation slides and vice versa? Computing protocols are being evolved to make such imports fairly easy. Some of the well-known computing protocol functions include message passing, process synchronisation and process switching, simple object access and object communication and data portability.

The idea of computing protocols is to encourage what are known as open systems design. Open systems follow industry standards and are capable of running on variety of platforms. For example, open office is an innovation in computing protocols. Many Java products use open computing protocols. Microsoft has recently announced a number of open computing protocols and has made them available in public domain. Open computing protocols offer greater opportunity and choice for freelance developers as they conform to industry standards. In contrast, closed protocols are proprietary in nature and are vendor specific.

Interoperability in computer systems is the main goal of computing protocols. By interoperability we mean the ability of different applications to interwork with each other using common data. User does not have to reformat and copy data from one application to another. Interoperability principles include:

- Ensuring open connections
- Promoting data portability
- Enhancing support for industry standards
- Driving open approach across competitors.

Although the open approach is currently limited to application packages from the same vendor, increasingly computing protocols are addressing issues for interoperability across different vendor products and platforms. Interoperability concept is also applicable to networked computers.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

1) Differentiate between computing and communication protocols.

....

....

....

....

9.4 COMMUNICATION PROTOCOLS: GENERAL CONCEPTS

Communication protocols deal with all aspects of communication functions that are required for information exchange among computers in a network or across networks. They are designed especially in the context of Internet. We have already discussed in Unit 8 the protocols that are used for information exchange in LANs. On the Internet, communication related functions include:

- Breaking up messages into packets
- Packet sequencing and reassembly
- Synchronisation or handshaking for information exchange
- Signalling: start and end of messages
- Switching: routing or forwarding of messages towards their respective destinations
- Connectionless and connection oriented transfers

- Message encapsulation and de-capsulation
- Format conversion
- Error detection and correction
- Setting up and termination of sessions
- Recover from unexpected loss of connection.

We discuss each one of the above functions in order in the following paragraphs.

Packet formation: As mentioned earlier, in data networks information is transferred in the form of packets. Packets are of fixed length with an upper bound on the size. For example, the maximum size of a packet in Ethernet is 1500 bytes. User messages longer than the maximum permissible packet size need to be broken into multiple packets.

Packet sequencing: As explained in Unit 8, routers take forwarding decisions independently for each packet. Even static routing algorithms may have more than one route defined for the same destination. Depending upon the path taken, the packets may arrive out of sequence at the destination. If the packets belong to the same message, they cannot be delivered to the user program unless they are properly sequenced. The concerned communication protocol needs to perform this function of sequencing the packets of the same message in proper order. Breaking up a message into multiple packets at the source and reassembling them at the destination are complementary functions performed by communication protocols.

Synchronisation: A packet transmission cannot start unless the receiving station is ready. The sending and the receiving stations exchange handshake signals and synchronise their transfer process. Handshake signals are like sending a query 'Are you ready?' and receiving a response like 'OK, go ahead'. The synchronisation process includes agreeing on transfer speeds and the required buffer sizes. When the transfer is in progress, the receiving station may want a pause for reasons like buffer full. Handshake signals are exchanged to enforce 'pause' and 'resume' actions.

Signalling: Once synchronisation is achieved, the actual transmission starts. At this stage, the sending station must signal to the receiving station the start of the packet. This is usually done by sending 'start of text (STX)' bit pattern. Similarly, the end of the packet is indicated by 'end of text (ETX)' bit pattern.

Routing: We have discussed this function in detail in Unit 8.

Transfer modes: There are two fundamental ways in which information transfer takes place in our life: connectionless and connection oriented. These transfers are analogous to postal communication and telephone communication respectively. In postal system, we write a letter, post the same and expect it to reach the addressee. The postal system delivers on the best-of-efforts basis. While the letters are delivered most of the time, some get lost somewhere. In telephone communication a connection is first established between the parties concerned and then the communication takes place.

Encapsulation: Consider a packet traversing six routers before reaching the destination. Let the source station and the first four routers be on the same network. The last two routers and the destination station belong to another incompatible network. The fourth router will now have to encapsulate the user packet to make it compatible to the destination network. Encapsulation is like putting one envelope (user packet) into another and writing the addresses differently on the outer envelop. At the destination, the outer envelope (encapsulated packet) is discarded and the original information obtained. This is termed as de-capsulation.

Format conversion: Sometimes when moving packets between incompatible networks, pack formats may have to be changed. An example is moving packets between Ethernet and Token ring LANs, which calls for format conversion.

Error handling: Errors occur in data transmission. These have to be detected and corrective action taken. Error detecting codes are used to detect errors. There are two basic techniques available for error recovery. One is when an error is detected in a packet, it is discarded and the sending station is requested to retransmit the packet. This technique is called automatic repeat request (ARQ). Handshake mechanism is used to request retransmission of the packet. The other is to use forward error correction codes (FEC) that are capable of both detecting and correcting errors at the receiving end.

Sessions: A variety of tasks are performed on the networks by establishing sessions between a server and a client computer. Online search of databases, remote job entry, remote login to a time sharing system and fie transfer between two systems are examples of different types of sessions. Different sessions have different requirements. For example, a dialogue may be two-way simultaneous or one-way alternating. A large file transfer session may call for establishing roll back points to recover from connection failures. Session protocols perform functions required to establish, successfully execute and terminate properly different types of sessions.

Packet loss: It is not unusual to experience unexpected loss of connections in networks. You might have had this experience while accessing Internet. Some Internet browsers including Microsoft's Internet Explorer have provision to resume a session that was terminated unexpectedly, say due to a power failure. Many communication protocols have features to recover from unexpected connection failures. This is particularly so in sessions related protocols.

In this section, we have studied the general features that are required in communication protocols. In the next section, we look at the details of some of the commonly used communication protocols.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

2) We use signalling as a matter of fact in our daily life. Give any four examples of such signalling.

3) Is SMS a connectionless or connection oriented service?

4) When you are typing on a computer terminal, you make a mistake. Then you correct it. Which one of the techniques, ARQ or FEC you are using? Give reasons.

5) Many word processing packages have auto correct features. Which one of the techniques, ARQ or FEC is used there? Give reasons.

....

....

....

....

9.5 COMMON COMMUNICATION PROTOCOLS

The field of ICT is replete with protocols. Hundreds of protocols have been defined for various purposes. Many are very specialised, some are rarely used and some are defunct. You are already familiar with computing and communication protocols. There are other classes of protocols such as *data (bits & bytes) transmission protocols, routing protocols, access protocols, services protocols* and *applications protocols.* As a user of networks, you need to be concerned with only about a dozen protocols. This is much like a language dictionary having over 100,000 words and the average vocabulary of a person being about 4000 words.

Extensively used communication protocols include:

- Internet Protocol (IP)
- User Datagram Protocol (UDP)
- Transmission Control Protocol (TCP)
- File Transfer Protocol (FTP)
- Remote Login Protocol (Telnet)
- Internet Control Message Protocol (ICMP)
- Dynamic Host Configuration Protocol (DHCP)
- Post Office Protocol 3 (POP3)
- Simple Mail Transfer Protocol (SMTP)
- Internet Message Access Protocol (IMAP)
- Cell Switching Protocols (ATM)
- Muti Protocol Label Switching (MPLS)
- HyperText Transfer Protocol (HTTP)
- Wireless Application Protocol (WAP)
- Lightweight Transport Protocol (LTP)
- General Packet Radio Service (GPRS)
- Simple Network Management Protocol (SNMP)

Of the above, the first three protocols, viz. IP, UDP and TCP are basic protocols used by a variety of Internet services and applications. We discuss them in Section 9.6. FTP and Telnet are most extensively used service or application level Internet protocols. A large number of applications on the Internet use what is known as Client-Server architecture. FTP and Telnet and web browsers also use this architecture. We present this architecture in Section 9.7. We discuss FTP and Telnet in Section 9.8.

Routers use ICMP to report any abnormal event on the network. An example of an abnormal event that a router may discover is the outage of the network in some segment. Such an event may be reported to all other routers on the network as welt as the to the network management centre. ICMP is also used to monitor the functioning of Internet. ICMP, however, is not discussed in this course. DHCP is used for managing IP address allocation in local networks. This is an advanced protocol meant for network administrators and as such we do not discuss the same. POP3, SMTP and IMAP are all e-mail related protocols. They are not discussed in this unit.

At the data transmission level, viz. transfer of bits and bytes a new convergent switching technique has emerged in the 1990s. This technique is known as cell switching and the associated transfer mode is called Asynchronous Transfer Mode (ATM). There is a set of protocols associated with ATM. We present an introduction to ATM in Section 9.9.

Multi Protocol Label Switching (MPLS) is router-based technique for routing IP packets fast. An introduction is given to MPLS in Section 9.10.

HTTP is the widely used web access protocol designed to work with desktop and laptop computers. WAP and LTP are wireless access protocols designed to work with small portable devices like mobile phones. These protocols are discussed in Section 9.14.

A related protocol is GPRS that is used to send packets over slow-speed wireless links. GPRS is not discussed here. SNMP is discussed in Unit 11 that deals with network management.

9.6 BASIC COMMUNICATION PROTOCOLS: IP, UDP, TCP

9.6.1 Internet Protocol (IP)

Internet protocol (IP) is fundamental to the operation of Internet. All services on the Internet use IP for sending or receiving packets. No computer can be connected to the Internet without the IP running on it. Hence all computer operating systems like Windows provide IP bundled with them. IP software is usually memory resident. IP specifies exactly how a packet must be formed and how an Internet router should deal with the packet.

Packet and packet switching are generic terms used in a variety of contexts in ICT. For example, a network not conforming to Internet standards may use packet switching and define its own packet structure. In order to distinguish from other packets and to uniquely identify IP packets, the term *IP datagram* or simply *datagram* is used. We use the term datagram to mean an IP packet in this course module.

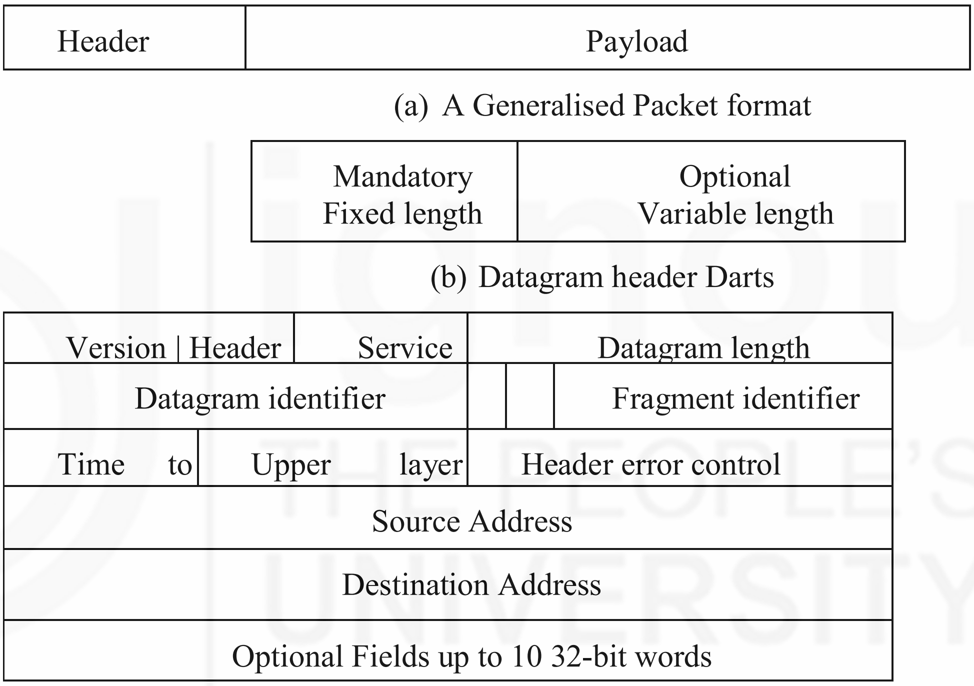
Although IP is a communication protocol defining datagram formats and transfer details, it serves an important purpose that goes almost unnoticed. Once on the Internet, a user can create and send datagrams to any computer on the Internet irrespective of where the destination computer is located. The user is unmindful of the presence of different component networks and a host of routers that interconnect them. Thus, IP makes a network of networks appear as one giant seamless data network.

An important aspect of IP is that it delivers datagrams on the best-of-efforts basis. The delivery is not guaranteed. The central idea in IP design has been internetworking and fast transfer of datagrams and not aspects like recorded delivery. Such aspects are taken care of higher-level protocol like TCP that we discuss later.

In general, a packet has the structure shown in Fig. 9.1. We use the term payload to denote the sum total of data handed over to IP for transmission. The payload may be pure user data or user data encapsulated with other information by any of the communication protocols.

The datagram header has a mandatory fixed length part and an optional variable length part as shown in Fig. 9.1(b). The fixed length is 20 bytes and the variable length can be up to 40 bytes making the maximum size of the header as 60 bytes. The different fields of the fixed part are illustrated in Fig. 9.1 (c) where each row is 32-bit or 4 bytes long. The source and destination addresses are 32-bit each corresponding to IPv4 address format.

IP addresses have two versions: Version 4 and Version 6 abbreviated as IPv4 and IPv6 respectively. IPv4 uses 32-bit address and IPv6 128-bit. IPv4 has been in use for a very long time, over 30 years, and most of the computers on the Internet have IPv4 addresses as of now. IPv6 has been introduced recently. Over the years, IPv6 is expected to replace IPv4 addresses. IP addresses are discussed in detail in Sections 9.13. The 'version' field in the header specifies the version to which the header belongs. Version information in each datagram permits the coexistence of different versions and smooth transition from one version to another.



(a) A Generalised Packet format

(b) Datagram header Darts

(c) Mandatory Fields of IPv4 datagram header

Fig. 9.1: IPv4 datagram formats

The maximum size of IP datagrams can be up to 64 k bytes including the header and the text part. But rarely such a big size is used. Different networks are allowed to set their own limit for the maximum size of the datagram well below the theoretical limit of 64 k. This maximum size set by a network is called the *maximum transfer unit* (MTU) of that network. This provision further complicates processing of datagrams. If a datagram is delivered to a network with a size greater than the MTU of the network, then the datagram needs to be fragmented for transportation within that network and reassembled at the exit point of that network. In such a case, we need a provision to identify the datagram and its fragments. IP header fields, datagram identifier* and 'fragment identifier' in Fig. 9.1(c) are provided for this purpose. While reassembling the fragments, IP must know the original protocol from which the fragments came. This is specified in the field 'Upper layer protocol'. The one-bit 'M' field when set to 'V implies 'More fragments to come'. This bit is set to '1' in all but the last fragment. The last fragment will have this bit set to '0'. There may be certain applications where fragmentation may not be acceptable. The one-bit 'D' field, if set to '1' would mean Do not fragment'. In such cases, the route will be so chosen that no fragmentation occur.

You may recall that sometimes packets may wander indefinitely without getting delivered to the destination due to routing errors. The field Time to live' is used to exercise control over such malfunctioning. The field 'service type' addresses issues like priority etc. Other fields in the header are self-explanatory.

9.6.2 User Datagram Protocol (UDP)

UDP provides connectionless service at the user level. It uses IP for this purpose. In that sense, UDP is a higher-level protocol when compared to IP. Here, a user submits his/her entire message to UDP with a request for transfer to the specified destination. User message is a payload for UDP. In turn, UDP encapsulates this with its own header and passes the same to IP as payload. User datagrams are different from IP datagrams. User datagrams do not conform to IP standard. They are just chunks of information of any size. UDP encapsulates the user datagram with its own header to form UDP datagram. UDP may split a user datagram into multiple UDP datagrams conforming to IP standards.

UDP datagram is shown in Fig 9.2. Now let us see as to why UDP adds its own header to the user data. Many application processes or users on a computer may use UDP simultaneously. Hence, UDP needs to maintain an identity of individual process and its corresponding destination process. This information is kept in its header in the form of source and destination port numbers so that the datagram may be delivered to the correct destination process along with the source identification. In Fig. 9.2 each row is 4 bytes long. With two rows the UDP header is 8-bytes long. The port fields in the header identify the source and destination processes or applications.

Source Port	Destination Port
UDP Length	UDP Error
UDP Data or payload	

.

Fig. 9.2: UDP datagram structure

Each port field is 16-bit long. The destination port value is used to deliver the user datagram to the correct application. The destination application may use the source port value for sending a response to the source application. The port address feature is the one that distinguishes UDP from IP. Otherwise, the functional capability of UDP is the same as that of IP. As in the case of IP, UDP messages may be lost, duplicated and delivered out of sequence.

The value of the UDP length field specifies the total length of the datagram including data part and the header. The use of UDP error control field is optional. This field is used only for the header portion of the PDU, i.e. the error control is done only for the header. The UDP does not perform error control at the datagram level. This must be taken care of at the application level. The payload supplied by the user or an application program follows the header. The entire UDP datagram with its header and user data becomes the payload for IP.

UDP, being a connectionless service functions on the best-of-efforts basis. There is no delivery acknowledgement in UDP. There is no guarantee of delivery. But it is used extensively like the postal system. The protocol is simple, efficient and fast. There are a large number of applications where occasional non-delivery is acceptable. If the underlying network is reliable, UDP is very effective.

9.6.3 Transmission Control Protocol (TCP)

TCP is a connection-oriented service. It uses IP and is at a higher level. In fact, UDP and TCP are at the same level. TCP is guaranteed delivery service. TCP provides reliable and error free communication. It achieves this in four ways:

1) Detects errors in datagrams and uses ARQ technique for error recovery

2) Recognises duplicate datagrams and discards all but one

3) Detect lost datagrams and retransmit the same

4) Sequences datagrams received out of sequence

You are already familiar with error detection and the use of automatic repeat request. As you aware that some routing algorithms send out multiple copies of datagrams to achieve robustness. TCP checks for duplicate datagrams and accepts only the error free copy received first. Detection of lost datagrams is done using acknowledgement and timer mechanisms. Receipt of every datagram is acknowledged by the destination. At the time of sending a datagram the source initiates a timer with a value within which the acknowledgement must be received. If the timer expires and no acknowledgement has been received, the source concludes that the datagram is lost and despatches another copy. By adopting the above said four mechanisms TCP is able to provide reliable and error free transmission.

TCP being a connection-oriented service establishes a connection between two communicating programs before data transfer begins. The service progresses in four phases as in the following:

- Source requests TCP for a connection by giving the destination identity. TCP contacts the destination
- Destination responds with a positive acknowledgement
- Data transfer takes place
- Connection terminated

This is very much like what happens in a telephone conversation. Much as the way we use both telephone and postal systems extensively in our daily life, both TCP and UDP are used extensively on the Internet. Since TCP and IP are closely interlinked, vendors bundle both the software routines as part of the operating system. This is why you always hear of TCP/IP together.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

6) How many fields are there in the mandatory portion of IP header?

7) Why is fragmentation required while transferring IP datagrams?

8) What action is expected of a router if the field 'D' is set to '1' in the IP header?

9) What purpose the port fields in UDP header serve?

10) How does TCP detect lost datagrams?

....

....

....

....

9.7 CLIENT-SERVER ARCHITECTURE

As mentioned earlier, FTP, Telnet and Web browsers are based on client-server architecture. Client-Server architecture is the most widely used form of computation on data networks. It has evolved from interactive computing model that was prevalent in the 1960s and 1970s. In interactive computing, a user interacts with a mainframe computer via a terminal that may be dumb or smart. The interaction model follows a master-slave approach. The mainframe computer acts as the master and the terminal as the slave. The slave terminal is under the complete control of the master computer.

With the advent of personal computers and data networks, the master-slave model of interaction has given way to **peer-to-peer interaction model.** Peer-to-peer interaction permits arbitrary communication among computers on the network. No distinction is made among the computers. A PC may contact another PC or a large mainframe as easily. Similarly, a mainframe computer can contact another mainframe or a PC. Distributed computing has become the norm. Distributed computing means any form of computation between two or more computers communicating over network.

Computers called **servers** that provide different types of services are on the networks. The services are accessible to other computers that are treated as **clients** of the service-providing computers. This model of interaction is known as the **client-server architecture.** A computer on the network may act both as a server and a client. When it provides service, it is a server and when it accesses the services of another computer, it is a client. We may thus say that the client-server architecture is a form of distributed computing with peer-to-peer interaction.

The client-server configuration is depicted in Fig 9.3. There are two machines and a network in the configuration: a server machine, a client machine and a data network.

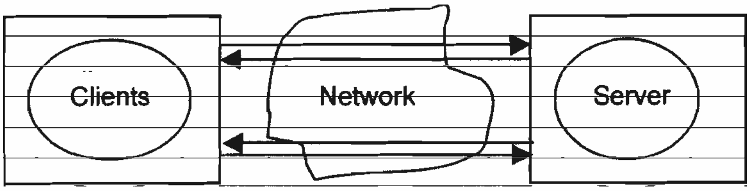


Fig. 9.3: Client -Server configuration

The server and the client interact via the data network. The server provides a set of information and computational services that are availed by remote clients. As shown in Fig. 9.3, usually many clients access one server simultaneously. It must be noted that the server and client machines do not actually interact. It is the server program and the client program that interact although we normally speak of server and client interaction. Support for multiple clients is possible only because of program-to-program interaction. A server creates as many processes of the same program as there are clients logged on to it. Use of multiprogramming and time-sharing features of the server operating system makes this possible. The server machine is one, but the instances of a server application program are many. This is how many users access one web site simultaneously. Client-server interaction may take place at one of the following three levels:

1) Human - Server Program

2) Human - Human

3) Client Program - Server Program

The first case is the most popular one with human client and a machine server. A typical example is that of user accessing information from a server, say searching a database. An example of the second case is student - teacher chat session. In online learning, student is tutored by a teacher. The student is the client and the teacher is the server. Timed periodic file transfer or email transfer between two or more machines are examples of the third case.

In all client-server interactions, it the client that always initiates a session. The server is ready and waiting without doing anything. When a client request comes, the server program responds. This is like a shopkeeper ready to sell with his shop open but the actual transaction takes place only when a customer arrives. The server service must be available on 24 x 7 (24 hours a day, 7 days a week).

Server systems are generally more powerful than client systems. They fall in one of the following categories:

- PC servers
- Workstation servers
- Mainframe servers

PC servers typically use standard 32-bit microprocessors. They have large RAM and hard disk capacity. They are ruggedised for continuous uninterrupted running with backup power systems and cooling systems where required. PC servers must have an operating system that can handle multiple users, as many client PCs may connect to the server at a time. Such operating systems are known as *network operating system (NOS).* Some of the popular NOS are MS Windows NT, Windows 2003, Novell Netware and Linux. All the servers are designed to support simultaneous access from many clients.

Workstation servers use high power or custom designed microprocessors. They are generally 64-bit or 128-bit microprocessor based systems. Workstation servers run under Unix like operating system that has a rich set of tools for supporting a wide variety of applications. Unix is a more reliable and secure operating system when compared to Windows. Linux is a recent addition to the world of operating systems and is considered a suitable substitute for both Unix and Windows. Linux is available in the open software domain. Some predict that in future, both PCs and workstations may run Linux instead of Windows or Unix. However, experience so far has not shown this to be true.

Mainframe computer based servers are even more reliable and powerful than Unix workstation servers. Mainframe based servers are often called enterprise servers to convey the fact that they are more powerful than PC servers or workstation servers.

Client systems are of two types:

- Desktop personal computers
- Mobile stations

The most popularly used desktop systems are Intel microprocessor based computers running Microsoft's Windows operating system. Such systems are sometimes called

Wintel systems signifying Windows operating systems and Intel microprocessor. The other class of desktop personal computer is Apple Macintosh. Mobile stations may be smart cellular phones (like Blackberry or iPOD), notebook computers and personal digital assistants or tablet PCs etc.

One of the powerful features of client-server architecture is its **scalability.** An application may start on a low-end PC and move in steps to a large PC, workstation and mainframe as the number of users rises. Interestingly, the upgradatton may happen without the user ever being aware of it.

In client-server architecture all applications have two program parts: a server program part and a client program part. The server program part is responsible for providing the specified services and the client part enables access to the services. Hence, anyone developing applications that would run on a network needed to develop both server and client parts of the software. The client software needs to be distributed to all client machines that may be spread all over the world. For example, you cannot access a server site that stores PDF (portable document format) files without an Acrobat Reader that is the client software for accessing PDF databases. In the early days, new server applications used custom-designed client software. Soon, it was obvious that roiling out client software to thousands of users all around the world is rather time consuming and expensive. With the arrival of World Wide Web (WWW), most of the network applications are designed to be **web-enabled** so that browsers now available with most of the PCs can access the applications without having to have special client software. In other words, the client software is embedded in the browsers. Two of the well-known browsers are Internet Explorer from Microsoft and Netscape Navigator from Netscape Communications.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

11) What are the differences between interactive computing and client-server computing models?

12) Can we say that Internet uses peer-to-peer communication? Why?

13) What mechanisms are used to support multiple clients on the same server?

....

....

....

....

9.8 APPLICATION LEVEL COMMUNICATION PROTOCOLS: FTP, TELNET

9.8.1 File Transfer Protocol (FTP)

FTP is used to transfer files from one computer to another on the Internet. FTP works in an interactive mode. A repertoire of FTP commands is available for interaction. A user invokes FTP client application on his/her computer as the first step. The user then enters the identity of the remote computer from which files are to be transferred using 'open' command of FTP. The FTP client then invokes TCP to establish connection with the remote computer. Once the connection is established, FTP server is activated at the remote computer.

At the next step, the user is authorised to access the remote computer by inputting a valid user name and password. A valid user account is required for this purpose. When the authorisation is successful, the user may examine and select a file on the remote computer by using the list command 'Is'. He/She then uses FTP 'gef command to transfer the file to his/her own computer. FTP client allows a user to transfer a file to the remote computer from the local computer as well. For this purpose, user invokes the 'send' command of FTP. The FTP client application is closed by 'bye' command.

FTP recognises only two types of files: text and binary. Any non-text file is treated as binary file. Examples include audio, computer programs, spread sheets and graphics data. Text files have to be strictly according to one of the standard character encoding schemes like ASCII or EBCDIC. If in doubt about the nature of the fife, it is best to specify binary format. Binary format will transfer a text file as well successfully. However, transfer of text files is more efficient and faster. Where known, it is a good idea to specify 'text' as the file type. However, If an incorrect type is specified, the resulting file may be malformed.

There are server systems on the Internet, which make available files to general public. Examples include servers providing government circulars or legal judgements. Such public files can be accessed without the user having an account on the server. FTP client makes this possible by providing an account called anonymous' with password as 'guest'.

Since FTP application runs on client-server model, the FTP server must run under multiprogramming and time-sharing operating system to enable multiple clients to access the server simultaneously.

9.8.2 Remote Login (Telnet)

Telnet allows an Internet user to log into a remote time-sharing computer and access and execute programs on the remote machine. For this purpose, the user invokes a Telnet client on his/her machine and specifies the identity of the remote machine. Telnet client makes a connection to the remote computer using TCP. Once the connection is established, the remote computer (Telnet server program) takes over the user's display and issues a login command. The user follows the regular login procedure by giving his/her account name and password. From then on the user computer behaves exactly like a terminal on the remote system. When the user logs out, the remote computer breaks the Internet connection and the Telnet client on the local machine exits automatically.

Remote login is a general access feature. The generality makes it a powerful tool on the Internet. It enables the programs on the remote computer accessible without having to make any changes to the programs themselves. The installation of the Telnet server on the time-sharing system is ail that is required. The telnet client and server together make the user computer appear as a standard terminal on the remote system. Hence, no changes are required as far as the application on the remote system is concerned. In view of this generality, different arbitrary brand of computers can be connected to the remote system. In effect, any computer on the Internet can become a Telnet client to any Telnet server on the Internet. Unlike FTP or e-mail, Telnet allows the user to interact dynamically with the remote system. Due to this, Telnet service is very popular.

Telnet sessions may run into occasional problems. The application program on the remote computer may malfunction or freeze. The local computer then hangs. We need a mechanism to come out of this situation. Remember that during a Telnet session, two programs are running: one the program on the remote computer and the other the Telnet client on the local machine. Telnet makes a provision to switch between these two programs. Once a Telnet session is established, every keystroke by the user is passed on to the remote computer. A special combination keystroke, like *Ctrl* +], is reserved to revert to the local program. The Telnet client examines every keystroke of the user before passing on the same to the remote machine. If the special combination key is pressed, it stops communication with the remote machine and allows communication with the local client program. The user can then terminate connection with the remote computer, close the Telnet client and resume local operations.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

14) Let the IGNOU LIS course modules be available on a fictitious FTP server called "cm.lis@Jgnou.ac.in". Write down the FTP commands and responses to download this unit to your computer.

....

....

....

....

9.9 SWITCHING LEVEL CONVERGENCE PROTOCOL: ATM

There are three major forms of switching techniques used in telecommunication and networks:

1) Circuit Switching

2) Packet Switching

3) Cell Switching

Circuit switching is the oldest technique used in telephone networks and has been in existence for over 120 years. Packet switching is about 50 years old used in data networks like the Internet. Cell switching is the most recent one evolved during mid 1990s used in new telecommunication infrastructure.

Before we proceed to discuss these techniques, definitions of two terms are in order: *channel* and *circuit.* A *channel* is defined as an information pipe with some specified characteristics like bandwidth, capacity, level of attenuation and noise immunity. A channel is a one-way link. A *circuit* is a two-way link and comprises two channels that enable two-way information flow between two entities. The two channels of a circuit need not have the same characteristics. If they do, then the circuit is said to be *symmetric.* Otherwise, the circuit is said to be *asymmetric.* Some authors tend to use the term *channel* to mean a physical medium. This is incorrect. A physical medium like optical fibre may carry several thousand information channels in a multiplexed mode.

Circuit switching is connection oriented. A circuit comprising two channels for two-way communication is established between the two communicating entities before the information transfer begins. The circuit is established using dedicated physical resources. The physical resource may be copper wires, optical fibres, radio or satellite links or a combination of these media. The circuit remains dedicated for the communicating pair until it is released, it is unavailable to any other communication need while dedicated to the communicating pair.

The main advantage of circuit switching is that once the circuit is established there is a direct connection between the communicating entities and the network is transparent to them. The information flows smoothly over the circuit from one end to another. The information is delivered in proper sequence and there is no possibility of out of sequence delivery. There is no delay caused by network elements like routers. The main disadvantage of circuit switching is that the scarce network resources remain dedicated for the entire duration of the information transfer phase and are heavily under utilised. Be it a telephonic conversation or computer interaction, mere are pauses during the session and the dedicated resources remain idle during the pauses.

In packet switching, messages are split as packets at the source and delivered to the network. *The* network transports the packets to the destination. Packet switched networks adopt two different approaches for transporting packets from the source to the destination:

- Datagram approach
- Virtual circuit approach

You are already familiar with datagram transport and the associated problems of out-of-sequence arrival at the destination and datagram losses. Virtual circuit approach was conceived to take advantage of in-sequence delivery of circuit switching while better utilizing the physical resources. The virtual circuit approach draws upon the idea of circuit estabtishment as in circuit switching. Instead of establishing a dedicated circuit, it establishes a fixed route from the source to destination. Since the packets follow the fixed route, they are delivered in order to the destination. Need for re-sequencing does not arise.

Virtual circuit makes routing more efficient and reduce the header overhead. As soon as a virtual circuit (fixed route) is established between a communicating pair, the same is given a unique identifier called *virtual circuit number* (VCN). The VCN defines the source and destination addresses, the message and the route. Hence, VCN together with the packet number uniquely identifies the packet. VCN plus packet number is much smaller in size when compared to the elaborate identification described earlier. Thus the header size and the transmission overhead are reduced significantly. Routing is also made simpler as the VCN is used to index a table to find out the outgoing link. There is no analysis of destination address and route determination.

Although virtual circuit concept is a major step forward, it still suffers from possible loss of packets and non-smooth flow of information. Since routers are involved, queues may build up and packets may be discarded. Dynamic variation in queue lengths may result in packets being delivered with different delay times thus interrupting smooth flow and affecting real time services.

Cell switching is the most recent switching technique evolved during 1990s. The main objective of cell switching has been to minimise the problems experienced in virtual circuit switching. This is done in two ways:

- To redefine the packet as a cell that is very small in size
- To leap forward in the speeds of virtual circuit switching.

Cell switching is designed to cause minimal network delay ensuring at the same time efficient utilisation of network resources. You are aware of MTU and the associated problem of possible segmentation and reassembly. This problem is completely avoided in cell switching. The entire infrastructure uses a standard cell size of 53 bytes. The cell has 48 bytes of payload and 5 bytes of header. Now let us understand the merits of cell switching. Ceil switching is built on a very reliable and ultra fast network infrastructure. The reliable technology almost rules out cell tosses. Even if a cell or two is lost very rarely, the effect is unnoticeable in real time services like voice and video. The small size of the cell makes the loss imperceptible to hearing or viewing. In data services of course, recovery is required.

Cell switching uses virtual circuit principle. The cells are guaranteed to be delivered in sequence. Virtual circuit reduces switching overheads significantly and makes switching extremely fast. For this reason, cell switching is sometimes called **fast packet switching.**

The networks that use cell switching are called **Asynchronous Transfer Mode (ATM)** networks. The reason for this is that the cells of a particular message are not switched in a fixed time frame say every millisecond. They are switched as they arrive. The arrival is a mixed bag of cells from different messages or services. They are switched in the order in which they arrive. Consecutive cells do not necessarily belong to the same message or service. In other words, the cells of a message or service are not continuous or synchronous in time. Hence, the term asynchronous transfer is used. Asynchronous transfer ensures effective utilisation of the network resources. The resources are not dedicated to one service.

In contrast, in the conventional circuit switched networks the information transfer is continuous and synchronous. In synchronous transfers, information pieces arrive in fixed time frame, say one byte every microsecond. In ATM, the cell arrival is not time synchronous. The time gap between the arrivals of two consecutive cells of the same message is not fixed but a variable one. However, the variability is very small because of the high-speed switching of ATM. For all practical purposes, the services perceive synchronous arrival. ATM is a technique that marks the convergence of both circuit and packet switching. Hence, ATM protocols are often referred to as convergence protocols.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

15) Why is cell switching superior to circuit and packet switching?

....

....

....

....

9.10 MULTI PROTOCOL LABEL SWITCHING: MPLS

As you know, virtual circuit makes routing more efficient and reduce the header overhead by using VCN. The VCN uniquely defines the source and the destination, the message and the route. Use of VCN reduces the header size and the transmission overhead.

Routing is made simpler as the VCN is used to index a table to find out the outgoing link. Multi Protocol Label Switching (MPLS) is an attempt to bring VCN concept to IP packets. Here, an IP packet is assigned a label that uniquely identifies the destination. In fact, the IP packet is encapsulated with the label header. The label is then used to index into a table to find out the outgoing link to be used for forwarding the packet. There is no examination of the destination address every time. This greatly simplifies routing overhead and makes the IP packets move faster through the network. This is particularly useful where large volume data transfers are involved as in the case FTP service.

MPLS is a router-based solution to improve the router efficiency. This is not a protocol that runs on any user machine. User machines run only the conventional communication protocols like TCP and FTP. We need MPLS-capable routers to implement MPLS. Only MPLS-capable routers can assign labels and handle MPLS packets. There are two ways in which the labels are assigned to IP packets: data-driven and control-driven assignments.

In data-driven assignment, when a packet enters a MPLS-capable router, it contacts the next MPLS-capable router and asks for a label for the destination address. The next MPLS-capable router in turn connects to the next one and the process is continued until the destination router is reached. Thus a fixed route is formed for all packets to the same destination. The first router now encapsulates the IP packet with the label supplied by the next router and forwards the packet. From then on, the label is used for routing. The name mufti protocol is used to signify the fact that the MPLS-capable routers can forward IP packets from a variety o f protocols like TCP and FTP.

In control-driven assignment, a destination router creates labels for all its host computers and passes them to its neighbours. The neighbours in turn create labels and contact other neighbours. The process is continued until all the routers acquire the path. Thereafter, the label is used for routing.

9.11 TELEPHONE AND MOBILE NUMBERING

Every entity in any network needs to be uniquely identified. Otherwise, the entity cannot be accessed. In telephone and mobile networks the entity is a phone instrument and it is number that uniquely identifies the entity. In Internet, the entity is a computer and it is IP address that uniquely identifies the entity. Although it is called an address, IP address is also a number. At the user level, the addresses are specified by string of characters on the Internet, (e.g. ignou.com). In a sense, similar character addressing is also available in telephone networks by way of directories where one looks up the number corresponding to a name. The addressing or numbering scheme follows a structure. We discuss the telephone and mobile addressing schemes in this section and the IP addresses in Section 9.13.

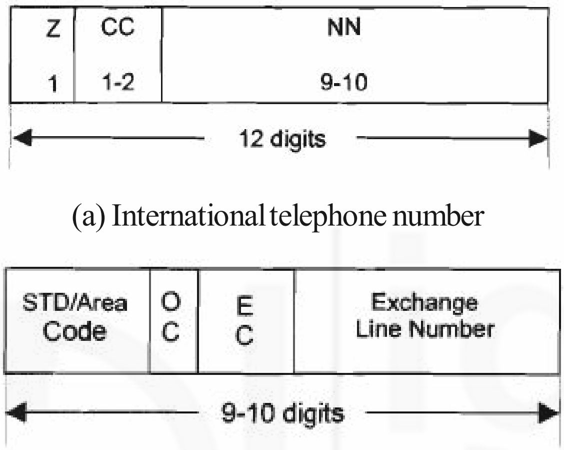
9.11.1 Landline Telephone Numbering

Telephone numbering worldwide follows an international standard set by International Telecommunications Union (ITU). The details are specified in the standards E.160 - E.163 of ITU. In ITU parlance, the numbering scheme is called **numbering plan.** As per the plan, the world is divide into 9 zones with each zone being identified by a zone code as indicated in Table 9.1. The zone names in Table 9.1 are representative. For exact delineation, one is advised to refer to the standards. Europe is given two codes, as there are many countries there.

Table 9.1: World zones for telephone numbering

Zone	Code	Zone	Code
North America	1	Australia	6
Africa	2	Russia	7
Europe 1	3	Far East	8
Europe 2	4	South Asia	9
South America	5	-	-

The structure of the number is illustrated in Fig. 9.4. The maximum size of the number is 12 digits. The first digit is the zone number. The remaining 11 digits are divided between country code (CC) and the national number (NN). The country code is one or two digits. With the zone code added, effectively the country code is 3-digit long. In common usage, zone code is not mentioned separately. It is included as part of country code. For example, the country code for India is mentioned as ‘91’. But to be precise, one should say that the zone code for India is ‘9’ and the country code is ‘1’. Together they make ‘91’. The country code is kept to be of variable length. The general principle adopted is that the countries with large population are assigned short codes of two digits (1 zone + 1 country). The countries with smaller population are assigned longer codes of three digits (1 zone + 2 country). For example, in zone 9, Maldives has a 3-digit country code ‘960’. With just over 200 countries in the world this coding would work for millenniums to come, in fact, forever. Countries with large population like India get 10 digits for national number and smaller countries tike Maldives nine digits. Ten digits provide for a maximum of 10 billion connections. For India with a population of 1.2 billion, this is adequate for times to come. You may appreciate that the telephone numbering plan has been designed with farsightedness.



(a) International telephone number

(b) National telephone number

EC = Exchange code OC = Operator code

Fig. 9.4: Telephone Number Structure

National telephone number has four parts in it as shown in Fig. 9.4(b). Subscriber Trunk dialling (STD) code may be further subdivided as one digit region code within the country and one or more digits of sub area codes within the region. In India, eight regions have been identified for telephone numbering. These are numbered 1 - 8 as shown in Table 9.2. The region descriptions are indicative and actual area covered is as per Department of Telecommunications guidelines.

Table 9.2: Indian regions for telephone numbering

Region	C	Region	C
Delhi, NCR, Haryana, Punjab	**1**	U.P. & Bihar	**5**
Mumbai, Maharashtra,	**2**	Orissa	**6**
Kolkata & North East	**3**	Central	**7**
Tamil Nadu & Kerala	**4**	Karnataka &	**8**

The sub area codes are kept to be of variable length. The general principle adopted is that the sub areas with large population are assigned short codes of two digits (1 region + 1 sub area). The sub areas with smaller population are assigned longer codes of three or more digits. For example, Delhi has a code '11' and Noida '120'. Similarly, Mumbai has a code '22' whereas Bopal, a town near Ahemadabad has the code '2707'.

Operator code (OC) is used when there is more than one service provider in an area. Until the early 90s, India had only the state operator, the Department of Telecommunications, providing telecom services in the country. But now telecom is opened up to private operators. We generally have more than one operator in major cities and towns. The operator code is used to identify the different service providers.

Every service provider has more than one telephone exchange in a city. Exchange code (EC) identifies the telephone exchange to which the subscriber is connected. Usually two or three digits are provided for EC. If the number of exchanges exceeds 99, we need three digits. This is the case in cities like Delhi and Mumbai. In smaller cities and towns, only two digits may be used.

The last part of the national telephone number is the line number assigned to the subscriber in the telephone exchange to which he/she is connected. Exchanges are usually designed to support 1000 or 10,000 subscribers. Accordingly, the line number may have 3 or 4 digits.

9.11.2 Mobile Phone Numbering

Technically speaking, there is no reason as to why mobile phone numbering could not follow the same numbering plan as the landline phone numbering. After all, the mobile is another telephone instrument except that it works on radio technology instead of landline (electrical or optical cable) technology. In fact, initially mobile phones in the United States used the same numbering scheme as the landlines. But, commercial considerations have led to a different scheme of numbering for mobile phones. In the beginning, the cost of mobile technology was relatively higher when compared to the landline technology. Mobile service providers needed to charge the customers higher. You may be aware that in the initial days of mobile communications, the incoming calls to mobile phones used to attract incoming call charges and the outgoing calls used to cost about six times the landline charges. The charge differential being so high, users needed to know whether they are calling a mobile phone or a landline phone. Further, roaming feature of mobile phones and the associated charging policies made the distinct identification of mobiles phones necessary. Hence, the need arose to distinguish a mobile phone from a landline phone. Thus was bom a different numbering scheme for mobile phones.

In general the series of numbers starting with '9' was reserved all over the world for future use while the landline numbering plan was evolved. When mobile technology came up and a need arose for distinguishing mobile phone numbering, it was decided to use the '9' series for mobile numbering. As you have learnt, India has 10-digit national number with its country code being '91'. The 10-digit national number starting with '9' was allotted to mobile phones. The '9' series provides for one billion numbers and it was considered adequate to meet the needs of mobile users in India, particularly because the cost being high not much penetration was expected. But, the history proved otherwise.

India has over 500 million mobile users in the country as of June 2010. Close to 750 million mobile numbers have already been allotted. If the rate of growth in mobile users continues at the present rate, we would run out of mobile number space soon. Hence, Telecommunications Regulatory Authority of India (TRAI) has opened up unused '8' series numbers for mobile users. An interesting fact is that India has only about 450 landline users and the growth rate here is not very significant. Considering this, it is likely over a period of time that number space for landline users may be reduced and the space thus freed may be allotted to mobile users. However, it is important to note that the potential for high-speed applications is much higher in the case of landline communications. This is because the radio bandwidth is limited whereas the landline bandwidth is unlimited.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

16) What is the maximum number of telephone numbers (both landline and mobile together) that can be assigned using the 10-digit national number?

17) Identify the different components of the international telephone number 911129534336.

....

....

....

....

9.12 NUMBER PORTABILITY

Number portability is a feature that allows a telephone user to retain his telephone number permanently. It is like the PAN (Permanent Account Number) allotted by the Income Tax Department or the Social Security Number assigned to individuals in the United States. Such numbers remain permanent for the lifetime of the individual irrespective of where the individual lives or works with. The number portability feature implements a similar concept. Imagine that you are given a telephone number once in your lifetime and that number remains valid for your complete lifespan. Would that not be very interesting? Number portability attempts to do just that. However, we have a long way to go in this regard.

Number portability needs to be considered in three situations from the users' point of view:

- Change of location
- Change of operator or service provider

- Change of service from landline to mobile or vice versa

Accordingly, three kinds of number portability are discussed from the telecom network point of view:

- Location portability
- Operator portability
- Service portability

Location portability implies that if a user moves his residence or place of work from one locality to another in the same city or moves from one city to another, his/her telephone number does not change. Both intra-city and inter-city movements have to be taken care of.

Operator portability implies that if a user moves from one operator to another, say from Airtel to Vodafone, his/her telephone number does not change. You may recall that the national number has a field (OC) that identifies the service provider. When number portability is introduced, OC may lose significance and may just be used by the old operator to redirect the call to the new operator.

Service portability implies that a user may move from one form of service to another and yet retain the original number. At present, three forms of service are available: landline telephone, mobile phone and voice-over-IP. Service portability must ensure mobility of the user among all these three services.

So far our discussions were limited to one portability requirement at a time. Portability requirements may occur in combinations as well. The following combinations may arise:

- Location + operator portability. A user may shift from one city to another and may want to change the operator also at the same time.
- Location + service portability. A user may shift from one city to another and may want to change from one form of service to another at the same time.
- Operator + service portability. A user may want to change the operator and the service as well.
- Location + Operator + Service portability. All aspects being changed at the same time.

Major changes may be required in the telecommunication equipments for implementing number portability. Hence, although number portability is being talked about for many years now, its implementation is not wide spread.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

18) A mobile user moves from the city of Hyderabad to Mangudi, a village in Tamil Nadu, which does not have mobile network coverage but has landline connectivity. What portability aspects would come into picture in this case?

....

....

....

....

9.13 IP ADDRESSING: IPV4, IPV6

As explained earlier, every entity in a network needs to be uniquely identified. Otherwise, the entity cannot be accessed. In telephone and mobile networks the entity is a phone instrument and it is number that uniquely identifies the entity. In Internet, the entity is a computer and it is IP address that uniquely identifies the entity. Although it is called an address, IP address is also a number. An IP datagram cannot be delivered to the destination unless the destination is uniquely and unambiguously identified and IP address does exactly that. Computers connected to the Internet are called **hosts** and hence the term host address is used extensively.

There are two versions of IP addresses under active use. The most widely used one is defined in Version 4 of Internet Protocol abbreviated as **IPv4.** IPv4 has been in successful use for about 30 years and it uses a 32-bit address. With 32 bits we can uniquely address 2^{32} hosts, i.e. approximately 4 billion (4 x 10^{9}) hosts. Remember that the world has a population of over 6 billion. If everyone were to have a computer in this world, we would not have enough IP addresses to assign to each one of these. Further, the IP address is a structured one having different address formats. Structuring has the effect of reducing the effective address space to a much smaller number than 4 billion. With the rapid growth of Internet over the last 30 years we are now on the brink of running out of addresses for new machines. It is in this context, a new version of IP, Version 6 abbreviated as **IPv6** has been recently standardised and is being introduced on internet. Addresses in IPv6 are 128 bits long. With 128 bits we can have approximately 256 x 10^{36} unique addresses. Such an address space is unlikely to run out in the foreseeable future and must serve the mankind at least for a few millenniums. The two addresses, IPv4 and IPv6 are interoperable and would coexist for many decades to come. IP addresses are assigned and managed by a non-profit corporation called *Internet Corporation for Assigned Names and Numbers* (ICANN) to ensure uniqueness in naming and numbering hosts. In this Unit, we discuss IPv4 in detail and IPv6 briefly. Readers interested in more details of IPv6 may refer to Further Reading material listed at the end of this unit.

IPv4 address is structured reflecting the objectives of Internet. You may recall that Internet is a network of networks. Hence, at the level of ICANN, the main interest concerns networks rather than hosts. ICANN, through its agents around the world assigns addresses to networks that contain many hosts. As you are aware, each network is connected to the Internet via a **router** or a **layer-3 switch.** The router has an Internet network address and is capable of forwarding datagrams towards destination networks. On its own network, it distributes the datagrams to the respective hosts. The host addresses are assigned by the respective network owners and maintained on the router.

For the purpose of assigning network addresses by ICANN, the networks are classified under three categories: large, medium and small signifying the number of hosts on the network. Corresponding to three network categories, there are three address classes: **Class A, Class B** and **Class C** respectively. The general structure of IPv4 address has three fields as shown in Fig. 9.5. Class A provides for large, Class B for medium and Class C for small networks. The ‘Class’ field is of variable length of 1 - 3 bits. A one-bit class field with a value ‘0’ specifies Class A addresses. The 2-bft field with value ‘10’ and the 3-bit field with value ‘110’ specify Class B and C addresses respectively. Often, the class field and the network number field together are called **network address** and the host field as **host address.** We also use the same convention in this unit.

Class	Network No.	Host

Fig. 9.5: IPv4 Address structure

In Class A address, 7-bit pattern following the first bit specifies the network number. Seven bits provide for $2^7 = 128$ bit patterns. Two of the 7-bit patterns are reserved for special purposes. They are all zeros ‘0000000’ and all ones ‘1111111’ patterns. The all zeros pattern implies the local network in which the host itself is located. The all ones pattern is used for loop back testing of protocols and applications. That leaves us with 126 Class A networks world over. The remaining 24 bits are used for host addresses supporting up to 16 million hosts on each network. With 16 million hosts on a single network, Class A represents the largest possible network on the Internet using IPv4 addresses.

In Class B address, 14-bit pattern following the first two bits specifies the network number. Leaving out the special patterns mentioned above, this means that up to $(2^{14} - 2) = 16, 382$ medium sized networks may exist on the Internet. Much as in the case of network addresses, special patterns are reserved for similar purposes in host addresses as well. Taking this into consideration, each medium sized network may have up to $(2^{16} - 2) = 64$ k hosts.

In Class C address, 21-bit pattern following the first three bits specifies the network number. Leaving out the special patterns mentioned above, this means that up to $(2^{21} - 2)$ H” 2 million small sized networks may exist on the Internet and each such network may have up to $(2^8 - 2) = 254$ hosts. Class C networks are ideally suited for small organisations. They are used extensively. Some small organisations may have more than 254 hosts, say ranging from 300 to 1000, but may not have as many as 16 k hosts to warrant a Class B address. This, in fact is the case with organisations like universities, research laboratories and large corporate houses. In such cases, the organisation is allotted as many Class C addresses as needed. For example, three Class C addresses can support up to 762 (3 × 254) hosts. This approach of using multiple Class C addresses helps in conserving Class B addresses. If Class B addresses were to be assigned to such organisations, the address space would remain heavily under utilised.

In addition to the above three primary classes of addresses, there are two special categories of addresses, Class D & E. Class D address is used for multicasting and Class E is reserved for future use. Multicasting is the distribution of datagrams to many hosts that have the same address group. Note that broadcasting is the distribution of datagrams to all the hosts. Hence, multicasting can be called as ‘limited broadcasting’.

For the sake of convenience and clarity, the 32-bit IPv4 addresses are presented in a dotted decimal notation. The addresses are viewed as four bytes (32 bits) and the decimal value of each byte is written with periods separating them. As you know, an 8-bit pattern can have values ranging from 0 —255. An example address in decimal notation is 183.41.235.7. The equivalent binary address is 10110111 00101001 11101011 00000111. With experience, it is felt that a hexadecimal representation would have served the purpose of clarity much better. As you know, hexadecimal representation uses a base of 16 using symbols 0 through 9 and A, B, C, D, E and F. The hexadecimal representation of this address is B7.29.EB.07.

Consider the case of an organisation that has 9,000 hosts. We would need to allocate a Class B address for this organisation to avoid allocating too many Class C addresses. Recall that one Class B address can support up to 16,382 computers. In this case, 7,382 addresses are wasted because the same network address cannot be used for another organisation. This is yet another example of how address space remains unutilised. The net result is the loss of address space. It was realised, though rather late, that a large segment of address space remains unused in IP class based address structure. While address space remained unused with the existing users, addresses for new users were becoming unavailable. A class definition with incremental number of hosts would have been far better. By the time this realisation came, the damage had been done. In order to contain further damage, Internet management introduced a classless addressing mechanism known as **Classless Inter Domain Routing** (CIDR) in late 1990s. The basic idea behind CIDR is to allocate the remaining IP addresses in blocks of contiguous addresses without any class consideration. With CIDR, an organisation can seek provision for hosts in powers of 2 such as 256, 512,1024, 2048 and so on. While network address length is fixed in Class A, B, C networks as 8, 16, 24 bits respectively, in CIDR the network address length may lie in the range of 8 - 31 bits theoretically. In practice, however, the range is 12 - 24. Note that the network address length of 8, 16, and 24 in CIDR automatically correspond to Class A, B and C networks respectively.

While CIDR makes more efficient utilisation of IP address space, it needed major changes in the routers all over the Internet, as the algorithm for routing has to undergo a sea change to handle both classed and classless addresses. Routing was relatively simpler and the router configuration was small with classed addressing. With CIDR the situation has changed. Routers have to maintain a much larger database including information about the length of the network address field. This is kept in a 32-bit field called **subnet mask** by setting as many higher order bits of the subnet mask to T as the length of the network address. For example, if the network address length is 20 bits, the subnet mask in binary notation is 11111111 11111111 11110000 00000000 and in decimal notation is 255.255.240.0. In decimal notation 255 means all the 8 bits of the byte, are set to '1' and 240 implies that four higher order bits of the byte are set to '1'. Therefore, twenty higher order bits of the subnet mask are set to '1' in this case.

IPv4 addressing is a classical example of how ad hoc and non-visionary decisions at the global level can turn out to be messy. It is worth noting that such a mess has never occurred in other fields like telephone and ISDN numbering. One of the major drawbacks of Internet is that many such shortsighted decisions exist and more and more ad hoc solutions are being found. This is clearly due to the lack of rigorous standardisation process such as the ones followed in ISO and ITU.

The introduction of CIDR and the concept of subnet mask led to a slight modification to the decimal notation for the IP address. Since the network address is now of variable length, its length is indicated by a number at the end of decimal notation after placing a 7. Example of new notation is 183.241.060.000/23. Here the network address is of length 23 bits. The subnet mask is 255.255.254.0

Let us now briefly discuss IPv6 addresses. As mentioned earlier, IPv6 addresses are 128 bits long. This is a very large address space, i.e. $2^{128} = 10^{38}$, i.e. approximately 10^{28} times the world population. This number is large enough to probably address almost every little thing on the earth. Therefore, it is impossible that this number space will ever get exhausted, certainly not for many millenniums. It is expected that in future items like refrigerators, air conditioners, cars, buses, ships, airplanes, and even bicycles will all be assigned IPv6 addresses so that they can be controlled and guided from the Internet.

IPv6 addresses are structured along the IPv4 addresses. The address space being very large, over 20 classes of addresses have been proposed. Five significant changes have been introduced in IPv6:

1) Large address space provided by 128-bit addresses

2) Flexible multiple header format. There is one base header that is mandatory and is of fixed size of 40 bytes. More extension headers can be introduced optionally.

3) Improved control options

4) Permits pre-allocation of resources

5) Provision for protocol extension.

There are only five fields in the fixed header portion other than the source and the destination addresses. These are *version, flow label, payload length, next header* and *hop limit.* You may recall that Ipv4 header has 11 fields in its header excluding the source and destination address.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

19) What are the lengths of network addresses in Class A, B, and C IP addresses?

20) How many hosts can be supported in one Class C address? Explain.

21) Write the following binary IP address in decimal and hexadecimal notations: 00011100 10101000 1100110011 00001100. What class of IP address is this? What is the network number (decimal) in this address?

22) Write the subnet mask in binary for /22 Classless address suffix?

23) How many hosts can be supported in classless address with suffix /27?

....

....

....

....

9.14 WEB COMMUNICATION PROTOCOLS: HTTP, WAP, LTP

As you are aware, World Wide Web (WWW) or simply Web is very popular on the Internet. These days, a large number of business houses, government agencies and many individuals have their own web sites. The number of web sites is growing day by day. Many business houses are currently upgrading their web sites to facilitate electronic commerce. In this context, it is necessary for you to learn about the communication protocol used for web access. Before we discuss the protocol let us briefly review the basics of web.

The concept of Web started in 1989 in France. In 1994, a consortium called World Wide Web Consortium (W3C) was established. W3C now has many countries as its members and is responsible for development and standards for Web and its access. From the users' point of view, web is a collection of documents scattered all over the world that are accessible over the Internet. The documents are often referred to as web pages. These documents are **hypertexts** as they contain embedded links to other documents. Embedded links are in the form of Uniform Resource Locator (URL) that contains three parts: a resource name, the identification of the server in which the resource is located and the protocol that can be used to access the resource. In effect, URL is unique identifier for a specific resource on the Internet anywhere in the world. An embedded URL is called **hyperlink.**

Web pages are of two types: static and dynamic. Static web pages are designed using a language called *Hypertext Markup Language* (HTML) that allows a developer to place text, graphics, sound, video and hyperlinks in a web page Being a mark up language, HTML defines how documents are to be formatted. In the process, it mixes the contents and format information. This poses serious problems while editing the pages. To overcome this deficiency, two new languages called *extensible Mark up Language* (XML) and *extensible Style Language* (XSL) have been developed. XML sets standards for structuring the contents and XSL for formatting the pages. Thus, the content and formatting are separated. These languages are being used increasingly these days.

As you are aware, web access is based on the client-server architecture that was discussed in Section 9.7. The web browser that acts as the client sends a web page request using URL and the server responds by returning the requested document. It is often necessary for the server to keep track of the user preferences for presenting information. The server does this by storing what are called **cookies** on the clients system. Cookies are short strings of data that a server sends along with a web page and uses the same later for meeting user preferences. The user, however, has the option of blocking cookies being stored on his/her system.

The protocol used for communication between the web browser client and the server is called *Hypertext Transfer Protocol* (HTTP). For secure applications, *Secure Hypertext Transfer Protocol* (HTTPS) is used. HTTPS is discussed in Unit 12 on Network Security. We discuss HTTP below.

HTTP is used universally to access web services all over the Internet. It specifies how a client may send requests to servers and how the servers may respond. The requests are sent in the form of ASCII (American Standard Code for Information Interchange) strings and the responses are received in the form of Multipurpose Internet Mail Extension (MIME) format. HTTP establishes a TCP connection on Port 80 of the server and uses the same for sending requests and receiving responses. More than one request may be sent without waiting for the responses. This is called pipelining of requests.

The first word of the ASCII string is one of the reserved words that specify the operation requested. For example, the reserved word GET signifies a read request for a web page and PUT for storing a page. PUT operation calls for authentication of the user. The authentication information usually follows the PUT request. The information that follows the operational request is called **request header** and need to be specified in a particular format.

In HTTP parlance, these operational requests are called **methods.** Other methods include POST, DELETE, and TRACE etc. POST is used append information to an existing page. It is used in the case of notice and news boards. POST method requires authentication so that only authorised users can post notice or news. DELETE is a request for removing the web page and obviously requires authentication. TRACE is a request for echoing the message that is being sent. This is used for diagnostic purposes.

The response from the server begins with a 3-digit status word that informs the client whether the request was successfully processed or not. It also indicates the reason in the case of unsuccessful processing. Typical failure messages include ‘no content found’, ‘page not found’, ‘page removed’ and ‘forbidden page’ etc.

With the advent of digital wireless access to Internet, considerable interest was generated in making small portable devices like mobile phones access web using wireless finks. Two access protocols **Wireless Application Protocol (WAP)** and **Lightweight Transport Protocol (LTP)** were developed for this purpose. These protocols were optimally designed to work with low bandwidth wireless links and wireless devices with a slow CPU, small amount of memory and a small screen. Obviously, such restrictions do not apply to desktop or laptop PCs. The device and link capability dictated the design of wireless protocols. In designing HTTP, these restrictions were not there. Over the time, the handsets have been made very powerful and wireless communication links have also become faster. The latest example in this category is iPhone4 from Apple Corporation. Accordingly, the later versions of WAP and LTP are also more sophisticated.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

24) How do the design considerations differ for wireless web access protocols when compared with HTTP?

....

....

....

....

9.15 SUMMARY

This unit has dealt with two distinct but closely related aspects: Internet communication protocols and network addressing. Protocols are generally software programs that implement the rules and procedures for communication. Some protocol functions are implemented in hardware too. There are protocols for computing purposes as well. Computing protocols are relatively a recent development. They define rules for information exchange among processes within a computer. Communication protocols define rules for exchange of information among computers. They deal with all aspects of communication functions that are required for information exchange among computers in a network or across networks.

An overview discussion of the communication protocols in Section 9.4 brings out their general functionalities like breaking messages into packets, packet sequencing and reassembly, message encapsulation and de-capsulation, error detection and correction and loss recovery. A list of commonly used communication protocols is given in Section 9.5. Then, the basic or fundamental protocols without which Internet cannot function are discussed in detail. They include IP, TCP and UDP. Internet Protocol (IP) is responsible for transporting packets from source to destination. Transmission Control Protocol (TCP) provides assured quality services that ensure errorless and lossless data transmission. User Datagram Protocol (UDP) is a low overhead, fast and simple protocol that delivers user messages on best-of-efforts basis.

The unit then covers Client-Server architecture that is fundamental to running remote applications on the Internet. It is the most widely used form of computation model on data networks. It has evolved from interactive computing model of yesteryears. Thereafter, two application level protocols that use client-server model for communication are discussed. File Transfer Protocol (FTP) is used for transferring files from one computer to another on the Internet. FTP works in an interactive mode using a repertoire of commands. Remote login protocol (Telnet) allows an Internet user to log into a remote time-sharing computer and access and execute programs on the remote machine.

The unit then focuses on two switching level communication protocols ATM and MPLS. Asynchronous Transfer Mode (ATM) is the new communication protocol used in basic telecommunication infrastructure. It uses the principle of cell switching that combines the advantages of both circuit and packet switching techniques. ATM is extremely reliable and fast. Routers use Multi Protocol Label Switching (MPLS) to speed up the process of routing packets across networks.

The unit then turns its attention to addressing entities uniquely on the networks. First, the numbering plans for telephone and mobile networks are discussed. Both international numbering and national numbering in India are elaborated. Another important issue, viz. number portability is then discussed. Number portability needs to be considered at three levels: Location portability, Operator portability and Service portability.

Addressing in data networks is then discussed. Version 4 of IP address (IPv4) is discussed in detail bringing out its limitations and merits. Developments in IPv6 are then briefly presented.

Finally, the unit discusses the web communication protocols. The universally used HyperText Transfer Protocol (HTTP) is described in detail. Brief features of wireless web protocols Wireless Application Protocol (WAP) and Lightweight Transfer Protocol (LTP) are then presented.

9.16 ANSWERS TO SELF-CHECK EXERCISES

1) Computing protocols define rules for communication among processes within a computer. Communication protocols define rules for communication among computers connected to the same or different networks.

   Computing protocols are concerned with storage, retrieval and processing functions of information management. Communication protocols are concerned with acquisition, transmission and distribution functions of information management.

2) Examples of signalling from our daily life:

   - A bus conductor's whistle to stop and start the bus
   - Flagging of a sport event like running race
   - Indicator lights in cars
   - Caller tunes in mobile phones.

3) Small Messaging Service is a connectionless service. One prepares a message and sends it across expecting it to be delivered. The service is provided on the best-of-efforts basis.

4) ARQ is the technique used here. You observe (detect) an error, erase (discard) it and input the right character (retransmit).

5) FEC is the technique here. The package detects the error and corrects it automatically. There is no retransmission by the user. This is forward correction at the receiving end.

6) Thirteen including source and destination addresses.

7) Internet is a network of networks. Each component network has its own maximum transfer unit (MTU) defined. If a datagram is delivered to a network with a size greater than the MTU of the network, then the datagram needs to be fragmented for transportation within that network and reassembled at the exit point of that network.

8) There are certain applications that cannot run with fragmentation. For such applications, the 1-bit 'D' field is set to '1'. This would mean 'Do not fragment'. If a router finds this bit set to '1', it must route the datagram via such networks that have MTU equal to or greater than the datagram size. Then no fragmentation will occur.

9) The port fields in UDP header identify the source and destination processes or applications. Using the information in these fields, UDP is able to deliver the datagram to the correct destination application. The source port identifies the source application that is sending the datagram, tf the destination so desires, it can send a reply datagram to the source application by using the source port address.

10) In TCP, detection of lost datagrams is done using acknowledgement and timer mechanisms. Receipt of every datagram is acknowledged by the destination. At the time of sending a datagram the source initiates a timer with a value within which the acknowledgement must be received. If the timer expires and no acknowledgement has been received, the source concludes that the datagram is lost and dispatches another copy.

11) In interactive computing, a user interacts with a mainframe computer via a terminal that may be dumb or smart. The interaction model follows a master-slave approach. The mainframe computer acts as the master and the terminal as the slave. The slave terminal is under the complete control of the master computer.

12) In client-server architecture, a computer on the network may act both as a server and a client. When it provides service, it is a server and when it accesses the services of another computer, it is a client. During interaction, the client is not under the control of the server. Client can do its own computing. Both the client and the server act independently and hence share the status of being peers. We may thus say that the client-server architecture is a form of distributed computing with peer-to-peer interaction.

13) Yes. Internet uses peer-to-peer communication. Any computer can contact any other computer. No computer is under the control of another. All computers are considered autonomous and can function independently. Hence, we say Internet uses peer-to-peer communication.

14) Machines do not communicate in client-server interaction. The interaction is between the server program and client program running on the server machine and the client machine respectively. As many instances of server program are activated as there are clients accessing the service. One instance is dedicated to one client. This is made possible by using the multi-programming and time-sharing features of the server operating system.

```
    ftp>open
    (to) cm.lis@ignou.ac.in
    Connected to cm.lis@ignou.ac.in
    LIS Course Module Services at IGNOU
    cm.lis@ignou.ac.in FTP server ready
    Name: yourusername
    Password:…………..
    Login OK
ftp> Is
PORT command successful.
Opening ASCII mode data connection for file list
Block 1: Basics of ICT
Block 2: Middleware Technologies
Block 3: Network Fundamentals
Block 4: Internet Tools and Services
Transfer complete
Xxx bytes received in xx seconds
ftp> Is Block 3: Network Fundamentals
PORT command successful.
Opening ASCII mode data connection for file list
Unit 8: Topology
Unit 9: Communication Protocols and Network Addressing
Unit 10: Protocol Architecture
Unit 11: Network Applications and Management
Unit 12: Network Security
Transfer complete
Xxx bytes received in xx seconds
ftp> get
(remote file): Block 3: Network Fundamentals/ Unit 8: Topology
(local-file): mydocuments/topology
PORT command successful.
Opening ASCII mode data connection for file list
Transfer complete
Local: mydocuments/topology
Xxx bytes received in xx seconds
ftp>bye Goodbye.
```

15) In packet switching, there are problems like out-of-sequence arrival at the destination and datagram losses. There is also the problem of segmentation and reassembly due to maximum transfer unit (MTU) limitation. Routing overheads are high in packet switching. In circuit switching, physical resources remain dedicated leading to their inefficient use. Cell switching overcomes these problems.

    Cell switching is designed to cause minimal network delay ensuring at the same time efficient utilisation of network resources. The physical resources do not remain dedicated.

    Cell switching is built on a very reliable and ultra fast network infrastructure. The reliable technology almost rules out cell losses. Even if a cell or two is lost very rarely, the effect is unnoticeable in real time services like voice and video. The small size of the cell makes the loss imperceptible to hearing or viewing. In data services of course, recovery is required.

    Cell switching uses virtual circuit principle. The cells are guaranteed to be delivered in sequence. Virtual circuit reduces switching overheads significantly and makes switching extremely fast. It is for these reasons that cell switching is superior to both packet and circuit switching.

16) 10-digit national number can support 10^{10} i.e. 10 billion numbers.

17) The 12-digit international telephone number given is '911129534336'. Here '91' stands for country code for India. The country code may further be subdivided as zone code '9' and country code within the zone as '1'. Next '11' stands for area code for Delhi. The area code may be further subdivided as region code as '1' and sub area code within the region as '1'. The following digit '2' stands for operator code, which in this case is MTNL. Digits '953' is the exchange code and '4336' is the exchange line number for the subscriber. The combination '29534336' is called the subscriber number.

18) Since the user is moving location, location portability is required. Since the user is a mobile subscriber and there is no mobile coverage available in the new place, service portability is required. It is assumed the operator is the same in the new place. Otherwise, operator portability is also required.

19) The lengths of network addresses in Class A, B, and C IP addresses are 8, 16, 24 bits respectively.

20) Class C address has 8 bits for host address. With 8 bits $2^8 = 256$ hosts can be supported. But the host addresses of all zeros and all ones are reserved for special purposes. Hence, a maximum of 254 hosts can be supported in Class C address.

21) The given binary address is '00011100 10101000 11101001 00001101'. Its decimal equivalent is 28.168.233.13. Hexadecimal equivalent is 1C.A8.E9.0D. This is Class C address as the most significant digit is zero. Seven bits following the first digit gives the network number which in this case is '0011100' and is 28 in decimal.

22) The subnet mask in binary for /22 Classless address suffix '11111111 11111111 11111100 00000000'

23) Classless address suffix is /27. This means 5 bits are available for host address. Leaving out the special reserved patterns of all zeros and all ones, we can have $(2^5 - 2) = 30$ hosts.

24) HTTP is designed for desktop and laptop and other high-end computers like servers. Here, there are no limitations of memory, computing power and screen size. It also assumes the availability of high-speed data links of at least 64 kbps. The emphasis on HTTP design is flexibility and powerful features. On the other hand, wireless protocols have to work with small portable devices like mobile phones. Here, the screen size is small, the available memory is very low and the CPU is not powerful. Data link speeds may be as low as 1.2 kbps or even less. Hence, the emphasis on WAP and LTP is high efficiency with essential minimal features only.

9.17 KEYWORDS

Client-Server	:	A computing and communication model used extensively in Internet
Connectionless	:	A service or protocol that commences information transfer without establishing a connection with the destination
Connection-oriented	:	A service or protocol that establishes a connection between the source and destination before information transfer commences
DHCP	:	Dynamic Host Configuration Protocol
Encapsulation	:	The process of covering a packet with another layer of header with a different format
Error control	:	The process of detecting and correcting errors
FTP	:	File Transfer Protocol
HTTP	:	HyperText Transfer Protocol
ICT	:	Information Communication Technology
ICMP	:	Internet Control Message Protocol
IMAP	:	Internet Message Access Protocol
Interoperability	:	Ability of different applications to interwork with each other using common data
IP	:	Internet Protocol
IPv4	:	IP Version 4 using 32-bit addresses
IPv6	:	IP Version 6 using 128-bit addresses
LTP	:	Lightweight Transport Protocol
NEIS	:	Networked Electronic Information Society
Open Protocols	:	Protocols that follow industry standards and are capable of running on a variety of platforms
POP3	:	Post Office Protocol Version 3
Protocol	:	A set of rules and procedures for information exchange between computers and applications

SMTP	:	Simple Mail Transfer Protocol
SNMP	:	Simple Network Management Protocol
TCP	:	Transmission Control Protocol
Telnet	:	Remote Login Protocol
UDP	:	User Datagram Protocol
WAP	:	Wireless Application Protocol.

9.18 REFERENCES AND FURTHER READING

Black, U. *Computer Networks: Protocols, Standards and Interfaces*. 2nd Edition. New Delhi: Prentice Hall of India, 1999. Print

Homer, Douglas E. *The Internet*. New Delhi: Prentice Hall of India, 2000. Print

Homer, Douglas E. *Internetworking with TCP/IP*, Volume I. 3rd Edition. New Delhi: Prentice Hall of India, 2001. Print

Lin, Yi-Bing. *Wireless and Mobile Network Architectures*. Singapore: John Wiley & Sons (Asia), 2001. Print

Mansfield, Kenneth C and Antonakos, James L. *An Introduction to Computer Networking*. New Delhi: Prentice Hall of India, 2002. Print

Panko, R. R. *Business Data Networks and Telecommunications*. 4th Ed. New Delhi: Prentice Hall of India, 2002. Print

Stallings, W. *Data and Computer Communications*. 5th Ed. New Delhi: Prentice Hall of India, 2000. Print

Tanenbaum, A. S. *Computer Networks*. 4th Ed. New Delhi: Prentice Hall of India, 2002. Print

Viswanathan, Thiagarajan. *Telecommunications Switching Systems and Networks*,. New Delhi: Prentice Hall of India, 2010. Print

UNIT 10 PROTOCOL ARCHITECTURE

Structure

10.0 Objectives

10.1 Introduction

10.2 Protocol Architecture and Protocol Stack

10.3 Layered Architecture

10.4 Principles of Layering

10.5 ISO-OSI Reference Model

10.6 Internet Protocol Architecture: TCP/IP Architecture

10.7 Bluetooth Protocol Stack

10.8 ISDN Reference Model

10.9 ATM Protocol Stack

10.10 SONET Hierarchy

10.11 Mobile Network Protocol Architecture

10.12 Summary

10.13 Answers to Self-check Exercises

10.14 Keywords

10.15 References and Further Reading

10.0 OBJECTIVES

After going through this Unit, you will be able to understand and appreciate:

- What are protocol architectures and protocol stacks;
- What layered communication is;
- What are open systems;
- Open System Interconnection (OSI) reference model;
- Components of Internet protocol architecture;
- How Internet protocols are layered;
- Bluetooth technology;
- Bluetooth protocol stack;
- Integrated Services Digital Network (ISDN);
- Broadband ISDN (BISDN);
- Protocol model of BISDN;
- ATM protocol stack;

- Synchronous Optical Network (SONET) bit rates hierarchy;
- How mobile networks function; and
- Mobile network protocol architecture.

10.1 INTRODUCTION

Computer networks are complex systems. You already may have gathered an idea about the complexities involved while reading Units 8 and 9. Interconnecting computers with wires or wireless links is only a small part of the task of establishing a computer network. Even here, one has a variety of options. Different types of wires and cables, wireless systems of different speeds and a host of topologies to choose from are some of the considerations at this stage. Then, there are innumerable protocols to work with. Routers, hubs, repeaters and switches and many other devices need to be appropriately chosen and used. There are security considerations and finally applications to run. No doubt, we are dealing with a very complex system while designing and establishing a computer network.

10.2 PROTOCOL ARCHITECTURE AND PROTOCOL STACK

The subject matter of this unit is protocol architecture. Much as the way general architecture helps us in planning, designing and building complex structures, protocol architecture helps us in planning and designing complex network solutions.

Design and construction of any complex system follow three fundamental principles that are pursued in three different stages:

- Top-Down design
- Stepwise refinement
- Bottom-Up build

Top-Down design is the most widely used approach in the design of any complex system. The first step in this approach is the conceptual design. At this stage, one needs to take a holistic view of the complete system including its scalability and future growth. One is not concerned with individual parts or components at this stage. It is like working out an architectural or artistic overview of a major office or housing complex. One makes an attempt to visualise the system in its entirety identifying the major subsystems. The interconnection and interaction of these major subsystems are broadly spelt out. Many alternative conceptual designs are usually worked out, merits and demerits of each design are discussed and debated and one design is chosen finally.

The next stage in planning, design and execution of a major project is stepwise refinement. The chosen conceptual design is now refined or exploded to review further details. The major subsystems are broken down into smaller subsystems. The smaller subsystems are further refined to identify the components that go to make the subsystems. At every step, details of interfaces and interconnection among the subsystems are meticulously planned and precisely spelt out. The stepwise refinement process is thus continued until full details of the design are worked out.

The third and the last stage is the execution of the project, i.e. building the complex system. This is done following the bottom-up principle. As you know, actual building of a structure follows a brick-by-brick approach. Bottom-Up build signifies this aspect.

The proper components are selected first. Sometimes, even the components need to be designed and tested separately. Using the components selected or designed, smallest subsystems are assembled and tested. Many small subsystems are then interfaced to form larger subsystems and these are tested. Finally, the larger subsystems are interconnected to make up the full system. The full system then goes through trial runs before it is declared operational.

The three stages in realising a complex system as listed above are not watertight compartments. There is always certain amount of back and forth movement between the stages. For example, during stepwise refinement some constraint may come to light that might call for minor changes in the conceptual design. Similarly, during the construction process some incompatibility may come to light that might call for some changes in the details worked out during stepwise refinement process. The objective in the overall exercise must be to minimise such back and forth movements.

As you may know, architecture is the art or science of planning and building complex structures. It deals with the manner in which the elements of the complex structures are arranged and organised. Similarly, protocol architecture is the way of planning and building complex network systems. It deals with the manner in which the different protocols are arranged, interlinked and made to interact to achieve complex network functions.

Protocol architecture is defined as a comprehensive organised collection of a set of protocols at different levels with well-defined input/output interfaces. The protocols are designed to function in a co-ordinated manner to perform a variety of network functions to support different network applications. Levels define a hierarchical structure. Protocols in adjacent levels interact. The architecture is a comprehensive view of the entire functionality available. The levels at which different protocols exist are also defined.

Protocol stack is a chosen subset of protocols from the architecture to perform specific functions to support a specific application. While protocol architecture is the general structure reflecting the capability for performing a variety of functions, protocol stack is a specific structure to perform certain specified functions. Many authors use the terms *protocol architecture* and *protocol stack* interchangeably. We, however, make a distinction between the two terms in this course material as defined above.

Designing and supporting a network application follows the three-step process outlined above. With the aid of the protocol architecture and other subsystems required, a conceptual design for an application is finalised. At this stage, various protocol options are evaluated and interfaces to application-specific subsystems are considered. The conceptual design is then refined to arrive at the protocol stack that is required to support the application. The application is then implemented using bottom-up build approach. First, some gross and simple functions are tested and then detailed functions are integrated into the application.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

1) Differentiate between protocol architecture protocol stacks.

....

....

....

....

10.3 LAYERED ARCHITECTURE

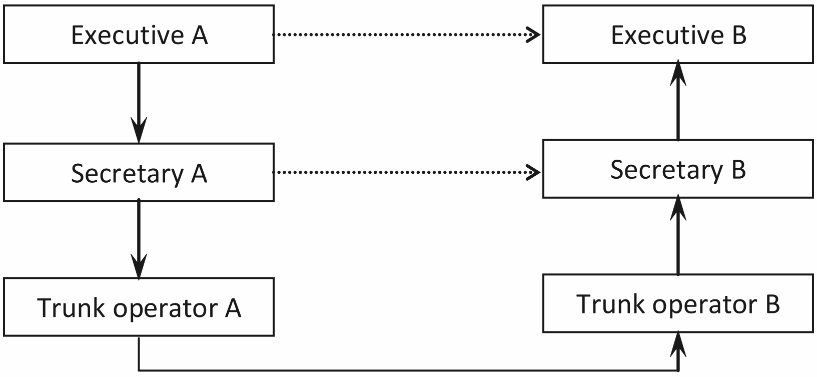


Fig. 10.1: Three-layer structure for trunk call connection

Layering is a natural choice for communication architectures. This is well illustrated by an example.

Consider the activities that are involved when executives A and B of two companies in different cities want to converse over a trunk telephone connection. For the sake of illustration, let us assume that there is no subscriber truck dialling (STD) facility between the two cities and trunk operators put through the trunk calls. Let executive A be the calling party. As a first step, he requests his secretary to connect him to executive B. His secretary in turn calls up the trunk operator A and communicates the calling number, called number, nature of the call, the name of the particular person called etc. Then, the trunk operator A calls up the trunk operator B in the other city and communicates the details. Remote trunk operator B now calls up the secretary of executive B, who in turn confirms with executive B that he would like to receive the call and requests the operator to put through the call. This process is depicted in Fig. 10.1.

A few interesting observations are in order:

1) A three-layer structure is used in this communication process.

2) The conversation between two adjacent layers is strictly business like.

3) There is generally a little private and informal conversation between the two trunk operators and between the two secretaries on account of their familiarity with each other. In other words, persons at the same level or layer exchange information in their own private way.

4) A layer obtains services from its immediate lower layer and provides services to its immediate upper layer. In this sense, a layer acts both as a user as well as a service provider.

5) A layer can communicate only with its adjacent layers.

6) There are fairly well defined functions to be performed by each layer.

7) It is immaterial as to how the functions of each layer are implemented. For example, the secretary may ask his assistant to book the call and as far as the executive is concerned, it is immaterial who books the call.

10.4 PRINCIPLES OF LAYERING

In fact, the above observations regarding a simple telephone conversation are stated as some of the important layering principles in arriving at a standard layered architecture known as *open system interconnection* (OSI) reference model or architecture evolved by International Standards Organisation (ISO). We state the ISO-OSI principles in the following section and discuss the ISO-OSI architecture in Section 10.4.

Principles of layering were first enunciated in the context of evolution of ISO-OSI reference model. Most important principles of layering are listed briefly in the following:

1) Layering is the most natural choice for data or voice communication network architecture.
2) Layers may be created to handle functions that are manifestly different in the process performed or technology involved.
3) Similar functions are to be placed in the same layer.
4) Boundaries are to be created at points such that the number of interactions across the boundaries are minimised.
5) Functions in a layer are to be localised and made autonomous to the extent possible so that the layer may be redesigned without affecting the interfaces with adjacent layers.
6) Redesign layers to improve performance by taking advantage of new advances in hardware and software technology.
7) A layer may be divided into sub-layers if the local function can be subdivided into independent modules
8) Layers are numbered bottom up. The bottom most layer is numbered 1.
9) A layer offers services to the upper layer immediately above.
10) A layer takes services from the lower layer immediately below.
11) Entities in the same layer but not in the same computer system are called **peer entities**.
12) Peer entities communicate using what are known as **peer protocols**.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

2) Particular protocol architecture has 10 layers. To which layer, Layer 5 provides services? From which layer, Layer 8 obtains services?
3) Does Layer 10 provide services? If so, to whom?
4) Does Layer 1 take services? If so, from whom?

....

....

....

....

10.5 ISO-OSI REFERENCE MODEL

As you are aware, data communication among computers involves a number of functions such as physical transmission of bits, error control, and routing and session establishment. In order to implement these functions efficiently, vendors of computer systems evolved their own architectures. Examples of vendor specific architectures are System Network Architecture (SNA) of IBM and Digital Network Architecture (DNA) of Digital Equipment Corporation (DEC). Such architectures permit interconnection of computers from the same vendor but not from different vendors. Systems or networks, which are not open to other vendor systems for networking, are known as 'closed' systems or networks. In order that heterogeneous computer systems from different vendors may be interconnected as a network, an architecture that is used as standard by all the vendors is required. The heterogeneity covers the following aspects:

- Systems of different vendors
- Systems under different management
- Systems of different complexities
- Systems of different technologies

ARPANET, the network project supported by Advanced Research Projects Agency of the Department of Defence, United States, is one of the pioneering efforts in connecting heterogeneous systems. The efforts put in and the experiences gained in the project have significantly contributed to the emergence of a set of world standards for computer communication. These standards, now well known as ISO-Open System Interconnection (ISO-OSI) standards, are widely accepted. The standards are based on a reference architecture which is described in the ISO standard IS 7498. International telecommunication Union (ITU) has adopted this standard under its own number X.200. The ISO-OSI architecture is considered 'open', as any vendor's system conforming to these standards is capable of organising information transfer with any other vendor's system that also conforms to the same standards.

OSI reference model proposes a general layered concept, with provision for adding or deleting layers as demanded by factors like service complexity, technology options, etc. ISO has recommended and standardised a 7-layer architecture shown in Fig. 10.2 taking into account various functions involved in data communication. On top of the seventh layer is the user who runs applications on the network. Hardware and software modules that implement the different functions of a layer are called **entities**. As mentioned earlier, the corresponding entities in the same layer but in different systems are called **peer entities**. The peer entities communicate using **peer protocols**.

Figure 10.2 shows two end systems that communicate with each other via two intermediate nodes. All the seven layers are active in the end systems. Only the first three layer functions come into action in the intermediate nodes. Entities in the first three layers always communicate with peer entities in the adjacent systems. The communication proceeds on a link-by-link basis from source to destination. Hence, layers 1-3 are called **link-to-link** layers. In contrast, the communication in layers 4-7 occurs between peer entities in the end systems. Hence, layers 4-7 are called **end-to-end** layers. The physical layer is concerned with transmission of bit streams either in synchronous or asynchronous mode. The data link layer handles errors and organises reliable transport of layer-2 data on a link-by-link basis. The network layer is concerned with the processing of destination addresses, routing of data units and internetworking. Routing algorithms are executed in this layer. Since the intermediate nodes perform the functions of the first three layers, they are also referred to as layer-3 switches. Layer-3 switches perform the routing function as well. The routing functions are implemented in hardware and hence layer-3 switches are must faster than conventional routers. They are also less expensive. As a result, layer-3 switches are replacing many routers particularly in campuses.

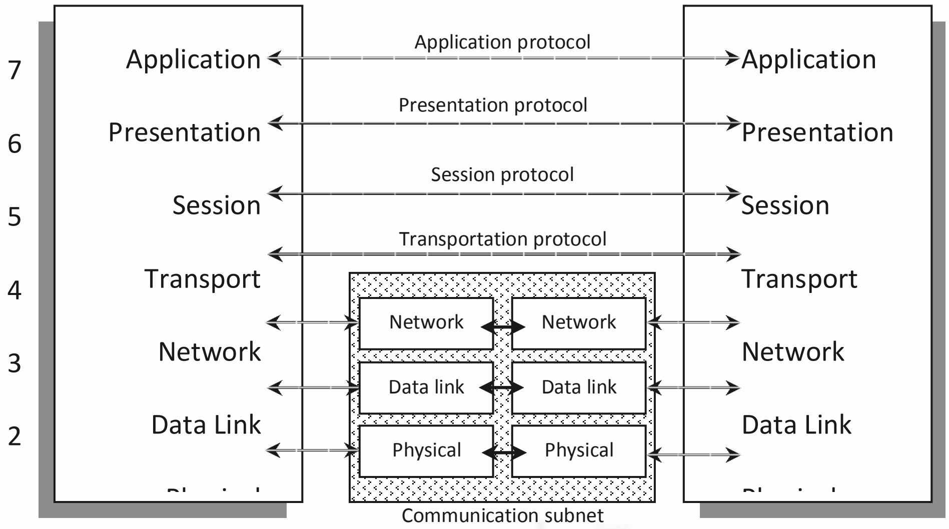


Fig. 10.2: ISO-OSI Reference Model

The transport layer is the first end-to-end layer and is concerned with reliable transport of full user messages between the two end systems. It is an interface layer between the user applications and the underlying network. This layer performs the function of splitting or segmenting the user messages into data units of appropriate size as required by the network layer. At the receiving end, this layer assembles the data units in proper order to reconstruct the full user messages before delivering the same to the upper layer. To ensure reliable transport of user messages, this layer retransmits data units that are lost in transmission or received with errors at the destination. The sessions layer permits two users or a user and an application to establish a session for purposes such as chat, interactive computing or information retrieval. The presentation layer deals with the syntax and semantics of information exchanged. It takes care of differences in data representation in the two end systems that may belong to two different vendors. The top layer is the application layer that enables a user to run different applications like electronic mail, accessing a web page etc. A number of different protocols have been developed that run at different layers of the OSI model.

The work of ISO on open systems followed ARPANET efforts and was based on ideas from ARPANET and the industry. ARPANET being a defence project, its results were considered confidential initially. Hence, the networking community looked up to ISO for open system standards. However, the elaborate procedure of ISO standardisation took time to finalise the architecture. In the meantime, the U.S Department of Defence decided to adopt and make public the ARPANET standards. Networking community started using ARPANET technology and soon it became the *de facto* standard in the public domain. ARPANET technology was adopted by Internet and is used widely today. We discuss Internet architecture in the next section. Although ISO-OSI architecture arrived late, there were many aspects in it that were comprehensive and structured. ARPANET protocols were refined later based on OSI recommendations. ISO-OSI is treated as a **reference model** even today and the new protocols and architectures are evaluated with respect to OSI recommendations. All new major developments in communications area come out with their reference model, protocol architecture or protocol stack. We discuss the important ones in Sections 10.7 - 10.11.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

5) How many layers are there in the end systems and in the intermediate nodes of OSI reference model?

6) Distinguish between link-to-link layers and end-to-end layers in OSI reference model.

....

....

....

....

10.6 INTERNET PROTOCOL ARCHITECTURE: TCP/IP ARCHITECTURE

Internet protocol architecture is popularly known as TCP/IP architecture. As you are aware, IP and TCP are two fundamental protocols used widely in the Internet. Since Internet grew in the initial stages based on these two protocols, its architecture came to be known as TCP/IP architecture. The architecture is presented in Fig. 10.3. The number layers in Internet protocol architecture is a debated topic. Some authors claim that there are only four layers and some talk about five layers. While there is complete agreement about the top three layers, many authors club layer 1 and 2 into one layer and call it by different names. The reason for this is that Internet standardisation process has mainly concentrated on the top three layers and allowed the bottom two layers to follow standards of different technologies. We present and consider 5-layer Internet protocol architecture in this course. In Fig. 10.3, we have shown the nomenclature for different layers along with some protocols in each layer. The protocol repertoire is rather large for each layer except for layer 4. As mentioned in Unit 8, the complete set of protocols run into several hundreds. We now briefly discuss each of the layers of the Internet protocol architecture in the following paragraphs.

Layer 5	Application	FTP, Telnet, SMTP, HTTP
Layer 4	Transmission Control	TCP, UDP
Layer 3	Internet	IP, ICMP, IGP, EGP, BGP
Layer 2	Network Access	Ethernet, ARP, 802.11, 802.16
Layer 1	Physical	X.21, 802.3, 802.5, SONET

Fig. 10.3: Internet Protocol Architecture

The physical layer deals with reliable movement of bits and bytes over point-to-point communication links. The technology and the techniques used depend on the network in use. X.21 is the protocol used in wide area telecommunication data networks. 802.3

and802.5 define the techniques for Ethernet and token ring respectively. SONET is discussed in Section 10.10. There are other protocols that are used for radio networks.

The network access layer deals with media access control protocols. You are familiar with Ethernet and Address Resolution Protocol. 802.11 and 802.16 deal with media access for wireless LAN and broadband wireless LAN respectively.

Internet layer runs IP and other routing protocols like Interior Gateway Protocol (IGP), Exterior Gateway Protocol (EGP) and Border Gateway Protocol (BGP). IGP deals with routing in local systems. EGP and BGP deal with routing between two autonomous computers connected to the Internet. These protocols are used by routers.

Transmission control layer has the most widely used protocols that you are already familiar with. You are also familiar with the protocols in the application layer.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

7) Name any two protocols that form part of Layers 2 – 5 of the Internet protocol architecture.

....

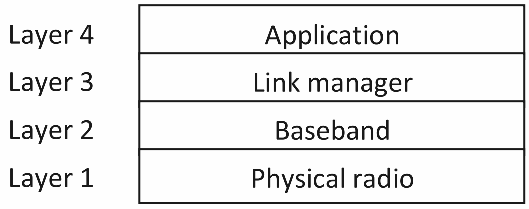
....

....

....

10.7 BLUETOOTH PROTOCOL STACK

As you are aware, Bluetooth is a short-range wireless convergent technology. It is a low power radio technology covering a small range of distances up to 10 metres. The purpose of the technology is to make bluetooth enabled devices that are in the vicinity of a master device to communicate in a wireless mode. The technology is very simple to use. All that one needs to do is to bring a bluetooth device close to bluetooth-enabled computer and a communication can start. There is no cable or modem or driver to be installed. The simplicity of its use is very attractive to users. Bluetooth functions in master-slave configuration. Usually a bluetooth-enabled computer acts as a master. Slave devices are dumb, basically doing whatever the master tells them to do. For example, a mobile phone may connect to a laptop computer and send and receive electronic mail if the two devices have bluetooth interface incorporated in them. Bluetooth communication calls for no additional infrastructure like mobile communication that calls for a radio network to be in place. Bluetooth devices can communicate directly without any network support.



Layer 4	Application
Layer 3	Link manager
Layer 2	Baseband
Layer 1	Physical radio

Fig. 10.4: Bluetooth Protocol Stack

Bluetooth protocol stack is shown in Fig. 10.4. Being a very different technology with very different objective, Bluetooth evolved its own protocol stack quite independent of OSI or Internet protocol architecture. Bluetooth stack has four layers. The physical layer deals with radio transmission and modulation. Its main objective has been to arrive at low-cost device so that the Bluetooth technology would have wide acceptance.

10.8 ISDN REFERENCE MODEL

The baseband layer deals with media access and is somewhat analogous to MAC protocols. It specifies how master controls slaves by defining time slots and allocating them to the slave dumb devices. It also decides the order in which the dumb devices communicate. The link manager establishes logical channels, authenticate devices, perform power management and ensure quality of service.

Integrated Services Digital Network (ISDN) has been perhaps the most important development to emerge in the field of telecommunications in the 1980s, and it will probably continue to dominate the developments in decades to come. Unlike many other developments, ISDN is a well-conceived and planned area of development in the field of telecommunications. Based on a study with the objective of exploring the use of digital technology, International Telecommunication Union (ITU) adopted and issued a definition of ISDN in 1972 as:

Integrated Services Digital Network: An integrated digital network in which the same digital switches and digital paths are used to establish different services, for example, telephony and data.

This definition was further refined and in 1984, a comprehensive and generic definition of ISDN was adopted as:

> An ISDN is a network, in general evolving from telephony IDN, which provides end-to-end digital connectivity to support a wide range of services, including voice and non-voice services, to which users have access by a limited set of standard multipurpose user-network interfaces.

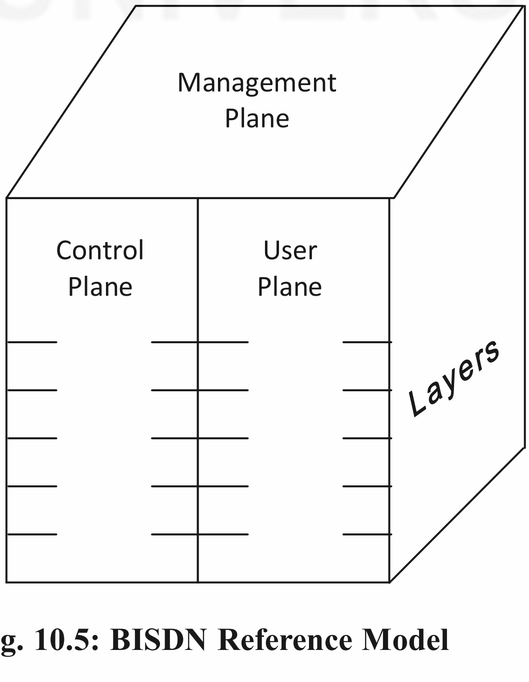


Fig. 10.5: BISDN Reference Model

In the above definition, IDN stands for Integrated Digital Network. Note that IDN did not envisage service integration. ISDN proposed a basic interface of what is known as 2B + D to users. 2B means two digital voice channels and D stands for data channel. It is used for signalling and data transfer. ISDN was designed to work up to speeds of 2 Mbps. Soon the need for higher speeds were felt. Broadband ISDN (BISDN) is defined as a network that caters to speeds higher than 2048 kbps, i.e. 2 Mbps.

Although started off with speeds just higher than 2 Mbps, after many years of deliberations, the basic speed of BISDN has been finalised as 155 Mbps, i.e. about 75 times the maximum speed of ISDN. This is because of the ambitious aim of BISDN to offer studio quality video and imaging services that demand very high bandwidth. The idea is to be able to distribute a wide variety of cultural, entertainment and educational materials to home and offices virtually on demand. The maxim is that "You ask for it and you get it". All these meant a quantum jump in technology, signalling and control, management and user services. This is reflected in the BISDN reference model depicted in Fig. 10.5. The model is designed along the lines of OSI reference model. Three distinct major components are visualised: User services, signalling and control, and network management. Each one of these components merits its own protocol architecture. Hence the reference model is a broad outline of these components having three different planes. Each plane has up to seven layers. Since a wide spectrum of user services were envisaged ranging from traditional voice service to video on demand service, a set of protocol layers is envisaged for the user plane. For example, conventional telephony is accessible in layer 3 and needs no protocol at higher layers. On the other hand, video services that call for maximum of bandwidth resources require protocols for presentation and session establishment. Protocols for these are at layers 6 and 5 respectively. Depending on the service accessed by user, different sets of protocols in different layers are invoked.

For signalling and control a protocol architecture called Signalling System 7 (SS7) is evolved. A set of telecommunications network management protocols are designed for management plane.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

8) Identify the important features of ISDN from 1984 ITU definition.

....

....

....

....

10.9 ATM PROTOCOL STACK

With the conception of BISDN at 155 Mbps came ATM. Cell switching and asynchronous transfer mode are chosen as the basic transport mechanism for BISDN. This development resulted in the definition of ATM protocol architecture. The ATM protocol architecture is depicted in Fig. 10.6. The architecture has three main layers. You may recall that OSI layering principles recommend formation of sub-layers if distinct functions are involved in a layer. This principle is used in forming ATM reference model and sub-layers are defined in Layer 1 and Layer 3. Both of them have two sub-layers each.

Layer	Sub-Layer	Functions
ATM Adaptation Layer	Convergence sub-layer	Convergence
	Segmentation and reassembly sub-layer	Segmentation and reassembly
ATM Layer	NIL	Flow control Cell header formation and analysis VC links management
Physical Layer	Transmission convergence sub-layer	Error control Cell management
	Physical medium	Bit timing

Fig. 10.6: ATM Protocol Architecture

10.10 SONET HIERARCHY

At the lowest level, bit transmission at rates of 155 Mbps and above take place. In addition, physical layer performs header error control functions and cell management that includes multiplexing and demultiplexing of cells. The second layer, ATM layer deals with flow control, cell header formation and management of virtual circuit links. ATM adaptation layer converges a variety of real time and non-real time services of different speeds to ATM standards. In addition it deals with segmentation and reassembly.

Synchronous Optical Network (SONET) is the optical transmission network for ATM. Coaxial cables are also used for transmission. SONET uses light as carrier often referred to as Optical Carrier (OC). Coaxial cables use electrical signals. SONET constitutes one of the underlying transport networks for ATM. SONET does not have a protocol hierarchy or architecture. It has what is called transmission hierarchy defining speeds at which segments of the network may carry information. The SONET/SDH hierarchy is shown in Table 10.1. Transmission in SONET uses synchronous technology. Hence, its transmission hierarchy is also called as Synchronous Digital Hierarchy (SDH). ITU uses the nomenclature SDH. The minimum speed of ATM is defined to be 155.52 Mbps corresponding to STM-1 of SDH. STM stands for Synchronous Transmission Multiplex.

Table 10.1: SONET-OC/ITU-SDH Operating Speeds

SONET-OC	ITU-SDH	Speeds (Mbps)
OC-1	-	51.84
OC-3	STM-1	155.52
OC-9	STM-3	466.56
OC-12	STM-4	622.08
OC-18	STM-6	933.12
OC-24	STM-8	1244.16
OC-36	STM-12	1866.24
OC-48	STM-16	2488.32

In transmission, there is a basic data rate that gets multiplexed for higher speeds. Thus, OC-3 means three OC-1 channels are multiplexed. But for ATM the basic speed is 155.52 Mbps defined as STM-1. ATM starts with 155.52 Mbps and its higher speeds are multiple of the basic rate. STM-4 implies that four 155.52 Mbps streams are multiplexed to obtain a rate of 622.08 Mbps. This corresponds to 12 SONET basic rate channels. ITU has also defined a similar hierarchy for coaxial cable based hierarchy knows Synchronous Transport Signal (STS). STS-1 starts at the same level as OC-1 at 51.84 Mbps.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

9) What is the basic bit rate in ATM? What is its nomenclature in SDH? Which level of SONET corresponds to this rate?

10) What are the bit-rates corresponding to STM-4 and OC-16?

....

....

....

....

10.11 MOBILE NETWORK PROTOCOL ARCHITECTURE

As you are aware, mobile communication is major development that has taken place in the last two decades. It is a complex communication system. It offers a wide variety of services: voice, text and multimedia messaging, and a host of value added services. It is important for you to understand the protocol architecture of this system.

First let us understand how the system functions. A simple network configuration diagram is shown in Fig. 10.7. Let us digress a little from the main technical discussion to focus on a societal issue. Do you know that your mobile phone is continuously transmitting electromagnetic signals? Have you heard of advice not to keep your mobile phone close to your heart? This is because it is suspected that continuous transmission of electromagnetic signals near the heart may be harmful to it. There is no health study to prove this conclusively. If you do follow the advice, it is only a precautionary measure. On the other hand, radiation from mobile towers seems to be reaching levels that are harmful to health. Now back to technical discussion.

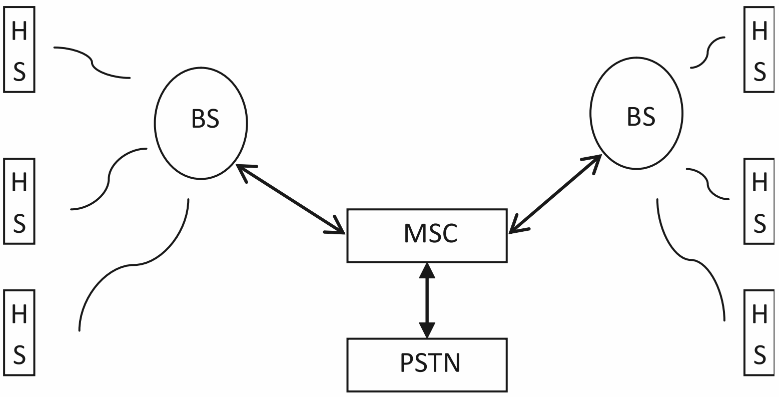


Fig. 10.7: Simple mobile network configuration

Continuous radiation from user handsets (HS) keeps them in touch with the nearest base station (BS). The base stations have two major subsystems. One, the tower and the antennas mounted on them. Collectively, they are called transceiver subsystem. The other subsystem is the control electronics housed at the basement of the tower, generally referred to as the base station control subsystem. The radio link between the base station and the handsets is modelled along the ISDN interface with two voice channels and one data channel. It is through the data channel that a handset is in continuous contact with the base station. It is also the data channel that is used for signalling and data services like SMS. You would have noticed that your mobile has facility to receive two phone calls simultaneously.

The radio coverage area of a base station is called a cell. The nomenclature 'cell phone' comes from this definition of cell. The cells are arranged in the form of a beehive with hexagonal area coverage. Two adjacent cells have different frequencies of operation. This is necessary to distinguish between the radio coverage of two adjacent base stations and to avoid interference between signals emanating from adjacent base stations. When a cell phone moves from one cell to another, a handoff procedure is followed to ensure that the transition is smooth without any break in the conversation. You may note that the frequency of operation of the handset changes when it moves from one cell to another. The handsets are designed to operate over the complete range of frequencies used by the base stations. This is what allows your handset function even when you are roaming.

The handsets are remotely monitored. Status information about handsets is stored in a database called mobility database (MDB). When you switch off a handset, it sends a signalling message to the BS, which in turn passes on the information to MDB via Mobile Switching Centre (MSC).

All base stations are connected via landline or dedicated microwave link to MSC. It is through MSC a connection between two mobile phones is established. MSC searches the MDB to ascertain the location and status of the called mobile and then establishes the connection via the concerned BS. MSC is also connected to the landline network, Public Switched Telephone Network (PSTN) for providing connectivity between mobiles and landline phones.

Having seen how a mobile network functions, we may now focus on the protocol architecture of mobile networks. The architecture is shown in Fig. 10.8. There are four layers in the architecture. The physical layer deals with four types of transmissions: radio transmission between BS and HS, microwave or landline transmission between BS and MSC and landline transmission between MSC and PSTN. Similarly, the link management layer also deals with four types of links. The third layer, network layer handles routing. Routing is done in four contexts in mobile networks:

- Mobile to Mobile routing with the same operator
- Mobile to Mobile routing with different operators
- Mobile to Landline routing with the same operator
- Mobile to Landline routing with different operators

Layer 4	Application	SMS, MMS, Music etc.
Layer 3	Network	Routing
Layer 2	Link manager	Radio, Microwave, Landline
Layer 1	Physical	Radio, Microwave, Landline

Fig. 10.8: Mobile Protocol Architecture

Application layer deals with applications like text and multi media messaging and value added services like music on demand, scores of cricket matches etc.

As you may be aware, mobile communications have undergone three generations of development: 1G, 2G and 3G. The fourth generation has been announced and is coming up soon. The above protocol architecture is applicable to 2G systems. 3G and 4G systems are broadband systems offering a wide variety of user services and may be expected to have architecture along the lines of BISDN.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

11) What are the different communication links that the link manager of mobile networks needs to deal with?

12) What are the different contexts of routing that the network layer needs to deal with?

....

....

....

....

10.12 SUMMARY

This unit is concerned with protocol architectures. Protocol architectures are layered structures. Layering is a natural choice for any communication process. International Standards Organisation (ISO) has evolved 7-layer reference architecture for data communication. ISO lays down well-defined principles for layering. The layers are numbered from the bottom starting with 1. A layer hides all the layers below it from the layer just above it. Layering principles promote autonomous functions in each layer and also the formation of sub-layers for independent local functions. Most protocol architectures are built around OSI model. In any architecture, each layer houses many protocols that are useful in a variety of contexts. An application may use none or only one or at best two protocols from each layer to run the application. The selection of protocols for a specific application and the associated layered structure is called a protocol stack.

Internet protocol architecture is popularly known as TCP/IP architecture. TCP/IP architecture has adopted 4/5-layer structure. In this unit, we have considered 5-layer architecture. Bluetooth, which is a specific application, has four layers in its protocol stack. Broadband integrated services digital network (BISDN) is seen as the future telecommunication infrastructure. Considering very high speeds and plethora of user services in future networks, BISDN has proposed a 3-dimensional reference model with protocol stacks for user, network control and network management. Asynchronous transfer mode (ATM) network is considered to be the basic infrastructure for BISDN. In view of this, ATM protocol stack is presented. ATM has a 3-layer protocol stack with sub-layers in two of the layers. The optical network SONET provides the underlying bit transmission infrastructure for ATM. The transmission hierarchy associated with SONET is presented. Finally, the unit concludes with a discussion on mobile network protocol architecture.

10.13 ANSWERS TO SELF-CHECK EXERCISES

1) Protocol architecture is defined as a comprehensive organised collection of a set of protocols at different levels with well-defined input/output interfaces. The protocols are designed to function in a co-ordinated manner to perform a variety of network functions to support different network applications. Levels define a hierarchical structure. Protocols in adjacent levels interact. The architecture is a comprehensive view of the entire functionality available. The levels at which different protocols exist are also defined.

   Protocol stack is a chosen subset of protocols from the architecture to perform specific functions to support a specific application. While protocol architecture is the general structure reflecting the capability for performing a variety of functions, protocol stack is a specific structure to perform certain specified functions.

2) Layer 5 provides services to Layer 6. Layer 8 obtains services from Layer 7.

3) Layer 10 being the top layer provides no service to any other layer. But it provides services to the user. User interfaces with the top layer of any protocol architecture.

4) Layer 1 being the bottom most layer it takes service from no one.

5) There are 7 layers in the end systems and 3 layers in the intermediate nodes.

6) The first three layers (Layers 1 – 3) of OSI reference model are called link-to-link layers. Lin-to-link layers exist in all the end systems as well as in the intermediate nodes. Entities in the end-to-end layers always communicate with peer entities in the adjacent systems. The communication proceeds on a link-by-link basis from source to destination.

   The top four layers (Layers 4 – 7) of OSI reference model are called end-to-end layers. End-to-end layers are present only in the end systems and not in the intermediate nodes. The communication in end-to-end layers occurs between peer entities in the end systems.

7) Protocols of the Internet protocol architecture:

   Layer 2: Token ring, Ethernet, 802.11, 802.16 (any two)

   Layer 3: IGP, EGP, BGP, IP (any two)

   Layer 4: TCP, UDP

   Layer 5: FTP, Telnet, SMTP, MIME, SNMP (any two)

8) The important features of ISDN from 1984 ITU definition are:

   - Provides end-to-end digital connectivity

- Supports a wide range of services, including voice and non-voice services
- Users have access to the services by using a limited set of standard multipurpose user-network interfaces.

9) The basic bit rate in ATM is 155.52 Mbps. Its nomenclature in SDH is STM-1. The level of SONET corresponding to this rate is OC-3.

10) The bit-rates corresponding to STM-4 and OC-16 are 622.08 Mbps and 829.44 Mbps respectively.

11) The different communication links that the link manager of mobile networks needs to deal with are:

    - Radio link between mobile handset and the base station
    - Microwave link between the base station and the MSC.
    - Landline link between the base station and MSC
    - Landline link between MSC and landline network PSTN.

12) The different contexts of routing that the network layer of the mobile protocol architecture needs to deal with are:

    - Mobile to Mobile routing with the same operator
    - Mobile to Mobile routing with different operators
    - Mobile to Landline routing with the same operator
    - Mobile to Landline routing with different operators

10.14 KEYWORDS

2B + D	:	User interface in ISDN with 2 voice channels and one data channel.
ATM	:	Asynchronous Transfer Mode. Cell transport network used in BISDN.
BGP	:	Border Gateway Protocol. Used for routing information between autonomous systems connected to the Internet.
BISDN	:	Broadband ISDN. Conceived to be the future telecommunication infrastructure.
Bit rates hierarchy	:	Defines a basic rate and a set of multiplexed rates for data transmission.
Bluetooth	:	Short-range wireless technology.
Cell (mobile network)	:	Radio coverage area by a base station.
EGP	:	Exterior Gateway Protocol. Used for routing information between autonomous systems connected to the Internet.
IDN	:	Integrated Digital Network without service integration.

IGP : Interior Gateway Protocol. Used for routing information between systems connected to a local network.

ISDN : Integrated Services Digital Network that supports service integration.

ISO : International Standards Organisation.

Layered Architecture : A comprehensive set of protocols arranged in layers.

MDB : Mobility database. Used for holding status and location information about mobile handsets.

MSC : Mobile Switching Centre that establishes connection between calling and called numbers in mobile networks.

OC : Optical Carrier. Unit of data rate in SONET.

Open systems : Systems that use international standards so that the can interconnect with other heterogeneous systems

OSI : Open System Interconnection. A reference protocol architecture model proposed by ISO.

Protocol Architecture : A comprehensive organised collection of a set of protocols at different levels with well-defined input/output interfaces.

Protocol stack : A chosen subset of protocols to perform specific functions to support a specific application.

PSTN : Public Switched Telephone Network. Commonly used landline telephone network.

Reference Model : Protocol architecture model that is used as a reference to compare other architectures.

SDH : Synchronous Digital Hierarchy. Defines digital data rate hierarchy in optical transmission.

SONET : Synchronous Optical Network. One of the bit transport mechanism used in ATM.

SS7 : Signalling System 7. Signalling protocol stack used in telecommunication networks

STS : Synchronous Transport Signal. A hierarchy of data rates supported using cables.

10.15 REFERENCES AND FURTHER READING

Bryce, James Y. *Using ISDN*. 2nd Edition. New Delhi: Prentice Hall of India, 1998. Print

Homer, Douglas E. *Internetworking with TCP/IP*, Volume I. 3rd Edition. New Delhi: Prentice Hall of India, 2001. Print

Lin, Yi-Bing. *Wireless and Mobile Network Architectures*. Singapore: John Wiley & Sons (Asia), 2001. Print

Mansfield, Kenneth C and Antonakos, James L. *An Introduction to Computer Networking*. New Delhi: Prentice Hall of India, 2002. Print

Stallings, William. *ISDN: An Introduction*. New York: Macmillan Publishing Company, 1989. Print

Stallings, William. *ISDN and Broadband ISDN with Frame Relay and ATM*. 4th Ed. Singapore: Pearson Education Asia, 2001. Print

Tanenbaum, A. S. *Computer Networks*. 4th Edition. New Delhi: Prentice Hall of India, 2002. Print

Verma, Pramod K. *ISDN Systems: Architecture, Technology, and Applications*. New Jersey: Prentice-Hall Inc, 1990. Print

Viswanathan, Thiagarajan. *Telecommunications Switching Systems and Networks*. New Delhi: Prentice Hall of India, 2010. Print

UNIT 11 NETWORK APPLICATIONS AND MANAGEMENT

Structure

11.0 Objectives

11.1 Introduction

11.2 Service and Application Types

11.3 Electronic Text Messaging

11.4 Multimedia Messaging

11.5 Electronic Mail

11.6 Interactive Television (ITV)

11.7 Interactive Music (IM)

11.8 Application Delivery

11.9 Performance Issues

11.10 Why Network Management?

11.11 Simple Network Management Protocol (SNMP)

11.12 Summary

11.13 Answers to Self-check Exercises

11.14 Keywords

11.15 References and Further Reading

11.0 OBJECTIVES

After going through this Unit, you will be able to understand and appreciate:

- The need for global and national information infrastructures: GII, NII;
- The difference between network services and user applications;
- Fundamental forms of digital information;
- Interactive and distributive applications;
- Differences between broadcast, multicast and unicast transmissions;
- How text messaging (SMS) is implemented in mobile networks;
- The evolution of SMS language;
- How multimedia messages are sent in mobile networks;
- E-mail system and its features;
- What interactive television is;
- Interactive Music (IM) or Music-on-demand application;

- How real time service delivery is achieved;
- Performance issues in real time service delivery; and
- Network management and SNMP.

11.1 INTRODUCTION

As you know, the world is evolving towards a *Networked Electronic Information Society* (NEIS). Networked society means one in which a large proportion of the world population is interconnected or networked by some form of telecommunication system and the people carry out their day-to-day activities using the network predominantly. Day-to-day activities may involve tasks such as banking, ticket booking for travel or entertainment programmes, product ordering, financial transactions, exchange of mails, retrieving of information from a database, downloading of music files, simple telephone conversation etc. Electronic information is central to all these tasks.

Today's network infrastructure of NEIS is Internet. Internet is only a minuscule of a network that is envisioned for NEIS. Today's Internet services are predominantly text and data oriented with only sprinkles of graphics, still pictures and slow motion video. Only about one-sixth of the world population is connected to the Internet. Even with this level of service and connectivity, Internet is having serious problems of address space and bandwidth capacity. Experience shows that Internet is slow for many network applications and the quality of services is far from acceptable level in many cases. Internet is designed for data transport, and real time services like voice and video transmissions have serious quality problems. Internet is predominantly built over voice grade telecommunication infrastructure, its protocols have heavy overheads, and there are too many ad hoc solutions for problems encountered during operations. All these compound to almost insurmountable difficulties in bringing up Internet to any meaningful level of performance for NEIS.

The vision of NEIS calls for transportation of high quality audio including high fidelity music and high quality motion video apart from high-resolution graphics. With the present level of development and trend, support of such services on the present Internet is almost impossible. The key to the evolution of NEIS lies in building **Global Information Infrastructure (GII)** that would have adequate capacity and efficiency to support full-scale services envisioned for NEIS. What are these full-scale services and applications that would run on GII? How would such a massive infrastructure be managed? These are the subject matter of this unit. The way to GII is via **National Information Infrastructure (NII)** that needs to be set up by the each of the nations. In India, Telecommunications Regulatory Authority of India (TRAI) has proposed the setting up of a broadband fibre optic network spanning the entire country in the timeframe of 2011 – 2015 at an estimated cost of Rs. 30,000 crores.

Before we proceed further, two pertinent points are to be noted. First we need to make a distinction between services and applications. Services are provided by networks whereas applications are designed and executed by the users. Often, these terms are used interchangeably. But that is not correct. We illustrate the distinction by an example. Facilities offered by an e-mail server fall under network services whereas actual despatch of e-mail falls under user applications. A user requires network services to be able to run his/her applications.

Second is about the fundamental nature and forms of information. Advances in computer and communication technologies have brought about the representation, recording and communication of information in electronic form. Electronic information may be in analog or digital in representation. However, the terminology 'digital information' implies that the representation is entirely in digital form. At present, there is a perceptible trend towards the use of digital technology in both communication and computer fields. Everything electronic is moving towards digital technology. One may say that there is a digital revolution that is currently sweeping the world. As a result, electronic information is also going digital. These days, even sound and video are being recorded using digital technology. Many of you may be aware that many cinema theatres have modernised their projection system and use digital (Dolby) sound systems. Digitally recorded audio and video CDs are in common use today.

11.2 SERVICE AND APPLICATION TYPES

Finally, human beings are the ones who perceive information. Of the five senses, we use only vision and hearing for perceiving information. In vision, we make a distinction between language oriented text information and pictorial information. Accordingly, we have text and video as components of electronic information. One may now say that there are three fundamental components of electronic information: audio, text and video. All the three components are referred to as data, e.g. text data and video data. These three components together constitute **multimedia** information. We now move on to studying the different types of services and applications in the context of NEIS.

GII would bring digital data sockets to home that would support speeds of the order of 155 Mbps. Compare this with the present day so called broadband connection at 512 kbps. GII socket speeds would be about 450 times faster than the maximum speed that is available today. This is because of the ambitious aim of NEIS to offer studio quality video and imaging services that demand very high bandwidth. Imagine a communication facility at home that allows the members of the household to view a movie of their choice at any time of the day and for any chosen duration without any interrupting advertisements. The movie may be viewed partly, a bookmark created and resumed sometime later. Such a service, known as **video on demand** is one of most exacting services to be offered in NEIS. It is for this purpose that such high speeds are required in NEIS. In fact, the proposed precise socket speed is 155.84 Mbps corresponding to OC-3 standard of SONET or STM-1 of Synchronous Digital Hierarchy (SDH). You studied about SONET and SDH and the speed hierarchies in Unit 10.

Interestingly, all the network services or user applications that one can ever visualise for NEIS or otherwise can be placed under two broad categories:

- Interactive services or applications
- Distributive services or applications

A user needs interactive services to be able to run interactive applications and distributive services to run distributive applications. Interactive services are those in which there is two-way exchange of information. Such exchange of information may occur between two end users or a user and an application running on a server or between two application programs. Conferencing application where more than two end users participate is also interactive. Similarly, more than two application programs may interact with each other. Distributive services are those in which the information transfer is primarily one way, i.e. from a server to a user. Both the interactive and distributive services can be further classified as shown in Fig. 11.1. Interactive services include conversational services, messaging services and retrieval services. Distributive services include broadcast, multicast and unicast services as well as cyclic services.

Conversational services provide real time circuits (comprising two channels) for full-duplex communication among the conversing parties. The capacity and characteristics of the circuits are based on the requirements or demands of the user. Audio conferencing, video conferencing and collaborative research meetings are some example applications that run using conversational services. Please note that all these applications are interactive.

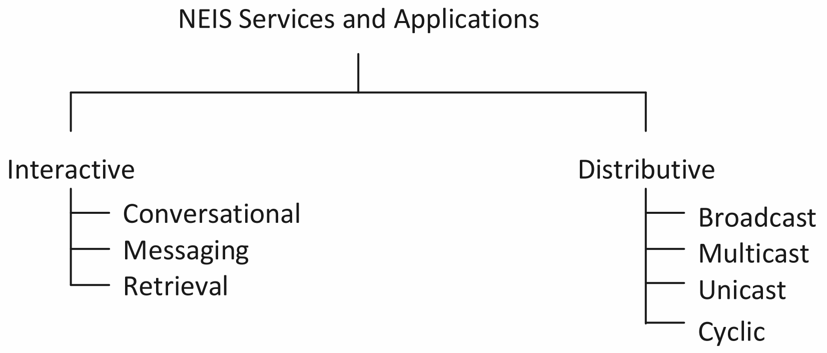


Fig.11.1: Categories of Services and Applications in NEIS

Messaging services are non-real time. They offer interpersonal communication among users using store-and-forward technique, mailbox or message handling functions. Since these services are not real time, they place less demand on the network resources. Typical messaging services include electronic text messaging, online chat, multimedia messaging, electronic mail, audio mail and video mail. As you know, electronic mail replaces the mailing of a letter. Similarly, audio and video mails replace the mailing of audiocassettes or videocassettes. Video mail is expected to become one of the most prevalent forms of messaging in NEIS. This is particularly so in the context of quality digital video cameras and monitors becoming available at affordable prices.

As students of Library and Information Science, you are familiar with retrieval services. Using retrieval services, user is able to access information stored in web sites or information centres that are, in general, available for public use. The service is interactive because the information is made available on user demand with specific queries. Based on the user query, a search process is initiated and the requested information is retrieved. Based on the output, the user can modify his/her query and obtain more specific information suiting his/her needs. The user thus has control over the information being provided to him/her.

Distributive services can be of one-time delivery type or cyclical delivery type. One time delivery includes broadcast, multicast and unicast. Broadcast services provide a continuous flow of information that is distributed from a central source to all the authorised receivers/users of the network. Every user has access to this information but has no control over it. Users simply tap into the flow of information.

Multicast is limited broadcast. The information from the central source is not distributed to all users but to a specified group of users only. The users are identified by a **group address**. Many groups may exist and users may be part of one or more groups. They may also be part of no group at all. Multicasting is directed towards groups. The information is delivered to all users of the group.

Unicast is also a form of distributive service. Here the information from the central source is delivered only to one user. Video-on-demand and music-on-demand are example applications that use unicast service. Here, there is a direct real time connection between the central server and the user. Thus, point-to-point connection characterises unicast application.

In addition to the above, there are two other related, but less used, forms of distributive services: **cluster cast** and **any cast**. Both these are also group based user services, but function differently. In cluster cast, the central source distributes the information to the nearest (usually) user who in turn distributes the information to all the other users in the group. There could also be other criterion for selecting the first recipient. In any cast, the information is distributed to only one of the users in the group. The information is not further distributed. The criterion for selecting a user could be based on who is less occupied currently. A typical application that uses any cast is query direction in a call centre. A user query is directed to the call centre executive who is currently free or to the one who becomes free next.

Cyclic information services distribute a set of information entities repeatedly to users in any one of the cast categories. The information is usually updated in every cycle. Example of applications include distribution of day temperatures, say every one minute or stock prices of selected stocks, say every 30 seconds.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

1) A mall has a video surveillance system installed. How would you categorise this application, interactive or distributive? Give reasons.

2) Give an example of text messaging system that is used commonly these days.

3) Differentiate between broadcast and multicast.

4) Differentiate between multicast and cluster cast.

5) Differentiate between unicast and any cast.

....

....

....

....

11.3 ELECTRONIC TEXT MESSAGING

Electronic text messaging primarily refers to Small Messaging Service (SMS) offered by mobile communication systems or Instant Messaging Service (IMS) offered by online chat rooms on the Internet. In the 1990s electronic mail used to be referred to as text messaging. But this is not appropriate in the present context for two reasons. One, text messaging implies short and quick communication in the present day context. Second, today's electronic mail is capable of carrying multimedia information. Text messaging is like conventional telegram. Electronic mail is more like conventional letter. Facilities in text messaging are limited and restrictive. Facilities offered by e-mail servers are fairly extensive. We study SMS in this section. E-mail service is covered in Section 11.5.

Most of you are familiar with SMS. Some of you might have used IMS and done online chats as well. Both SMS and IMS are part of messaging services and are interactive. In

SMS, the interaction may or may not be in real time. In IMS, it usually happens in real time. However, IMS generally supports a feature that allows a user to leave a text message for the other party even though he/she is not online.

SMS is a store-and-forward packet transfer service. It uses one of the control channels of the mobile network to send and receive short messages. The speed of the control channel used for SMS is fairly low. It generally operates at 110 bits per second (bps) or about 11 characters per second. Because of the limited screen, memory and processor capacity of the mobile handset, the length of short messages (SM) is limited to 140 bytes or characters. With 11 characters per second speed, it may take about 15 seconds for a message to be transferred. Longer messages are sent by concatenating multiple 140-byte messages. A continuity flag is used in the message header for this purpose. There is a SMS service centre (SMS-SC) in each mobile network that receives, stores and forwards the short messages.

Two types of short message services are supported on mobile networks:

- Cell broadcast SMS
- Point-to-point SMS

In cell broadcast SMS, a base station transmits short messages to all the mobiles in its coverage area. It is through this service that you receive a large number of advertisements on your mobile. If multiple base stations are chosen for transmitting the same message, then the number of mobiles that would receive the message goes up tremendously. This is how advertisements are distributed to a large number of users.

The base stations can be programmed to provide cyclic services or do multicast. In the case of cyclic services, information can be sent to users periodically. As an example, temperature in the coverage area of the base station can be sent to users, say every half hour or so. Using multicast, paid services can be implemented, i.e. information is sent only to users in a specified group, who have paid for the service. Using both cyclic and multicast, paid cyclic services can be implemented. Examples include providing cricket or badminton score updates.

Mobile networks support three categories of point-to-point SMS:

- Message originating from a mobile and terminating on another mobile
- Message originating from mobile and terminating on non-mobile device
- Message originating from non-mobile device and terminating on a mobile

Non-mobile devices include fax machines and personal computers. Receipt and delivery mechanisms differ in each of the above categories. Operating procedures also differ. Mobile networks use three special subsystems to implement SMS:

- SM service centre (SM-SC)
- SMS inter working Mobile Switching Centre (SMS-IWMSC)
- SMS gateway Mobile Switching Centre (SMS-GWMSC)

Mobile originated messages are first delivered to SMS-IWMSC, which in turn, passes the same on to SM-SC. Non-mobile devices receive and deliver messages from/to the SM-SC. SM-SC also connects to the SMS-GWMSC for delivering messages to mobiles. In effect, SMS-IWMSC acts as the input interface for mobile originated messages, SM-SC as input/output interface for non-mobile devices and SMS-GWMSC as the delivery interface for mobile terminating messages. As mentioned earlier, SMS is a store-and-forward service. Short messages cannot be sent directly from the sender to the recipient. They have to pass through the SM-SC.

Mobile-terminating short messages can be targeted to one of three destinations in a mobile:

- User specific
- Mobile equipment (ME) specific
- Subscriber Identity Module (SIM) specific

User messages are displayed to the users. ME-specific message may be used for activating or deactivating a function on the ME remotely. The SIM card processes SIM-specific messages. Using SIM-specific messages, special functions can be triggered by the network operator.

Limitation of 140-character messages in SMS has led to the evolution of what one might call as 'SMS Language' (SMSL). SMSL uses abbreviations to allow maximum use of the limited space as well as to compose messages quickly. Two techniques are used in framing abbreviations:

- Use of similar sounding letters or numbers to replace words or syllables
- Omission of letters from words, especially vowels.

The first technique is phonetic based one. Let us see some examples. The letter 'C' sounds the same as the words 'see' or 'sea'. The letter 'U' sounds the as the word 'you'. So the message "C U later" means "See you later" saves four characters in total length of 13 characters. This is a saving of about 25%. Similarly, numbers like 2, 4, and 8 can be used to substitute words or syllables that sound similar. For example, GR8 may mean 'great' and 'I wa8 4 u" may mean 'I wait for you'. Examples of the second technique are 'pls' for 'please' and 'msg' for 'message'. Another interesting abbreviation is 'ILU' for 'I love you'. Of course 'ILU' has been in use in Hindi movies for decades now.

Another feature of SMS is the use of what are called 'emoticons' that are representations of facial expressions formed using keyboard characters. Examples are :-) for happiness and :-(for unhappiness. Emoticons are quick and amusing way of conveying the emotions of the sender. Draw the above two emoticons on paper by rotating them clockwise by 90° and observe the emotions that they convey. Many SMS abbreviations are finding the status of accepted conventions for most users of the language. Example includes B2B that means 'Business to Business".

To simplify message generation, mobile handsets execute programs that use *predictive text input* algorithms. These algorithms predict the full word being typed based on the first 3 or 4 characters input. Some sophisticated algorithms even predict the next possible word.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

6) Distinguish between cell broadcast SMS and point-to-point SMS.

7) A car garage wants to send promotional offers to its registered customers on their mobile. Suggest a suitable SMS solution for the same.

8) What is the role of SM-SC?

9) A mobile operator has the capability to remotely activate WAP on mobiles. Which type of SMS is required in this case?

10) Recast the following sentence in SMS language 'Happy to see you tomorrow'

....

....

....

....

....

....

....

....

11.4 MULTIMEDIA MESSAGING

As you know, multimedia means a combination of video, sound and text. Multimedia messaging (MMS) usually means multimedia information being exchanged on mobile networks. In principle, transmission of multimedia information demands much higher bandwidth than text transmission (SMS). Maximum bandwidth is required when multimedia information comprises motion video with associated sound and overlaid text. MMS currently supported widely in mobile networks comprises only still picture. Any text in it is also a scanned picture. Such MMS requires relatively low bandwidth. But even this bandwidth is large when compared to the bandwidth required for SMS. As mentioned earlier, SMS is sent using control channels of mobile networks. Control channels have very limited bandwidth. MMS cannot be sent through these channels. Hence, MMS is sent via traffic channels of mobile networks. Traffic channels that carry voice have larger bandwidth and support higher data rates.

2G (second generation) or 2.5G mobile networks are still prevalent in most parts of the world. These networks do not have adequate bandwidth to transmit motion video etc. This is the reason why today's MMS is limited to still pictures. 3G mobile networks have larger bandwidth and hence can support the transmission of motion video. This is how these networks offer video calling and TV reception.

Digital image or still picture is stored and transmitted using a microscopic process. The picture is formed as a matrix of dots called *pixels* or *pels.* The word pixel or pel is a short form for picture element. The density of dots could vary from 75 dots per inch (dpi) to 2400 dpi both horizontally and vertically. The horizontal and vertical dot densities together are called the *resolution of the picture*. Larger the resolution, the better is the clarity of the picture. Larger the resolution, the higher is the bandwidth required for transmission. In 2G mobile systems, the picture resolution varies from 75 × 75 dpi to 300 × 300 dpi. The clarity is low at these resolutions. The pixels may be in colour or in black and white (B&W). Accordingly, the still picture is in colour or in B&W. In B&W, the pixels represent grey levels leading to the appearance of different shades. The pixel values are stored as binary numbers. Usually, grey values in B&W pictures are stored using 4-bit numbers. This allows 16 grey levels to be represented from black to white. Colour values require much larger binary string. Colour pixels use 8, 12, 16, or 24-bit representation. As a result, colour pictures require larger bandwidth for transmission.

There are three commonly used formats for storing and transferring digital images:

- Tagged Image File Format (TIFF)
- Graphics Image Format (GIF)
- Joint Picture Expert Group (JPEG) format.

TIFF has been developed as the common format for image scanners and DTP software. GIF has been developed for use on the Internet. GIF uses 8-bit representation for the pixels and hence can represent only 256 colours or grey levels. In this sense, it has limited resolution but the file sizes are small and can be transported easily across Internet. JPEG format is an image coding standard that has been optimised for continuous tone products such as photographs. It supports 16 million colours. In MMS, GIF format is used widely.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

11) A 1" × 2" picture is stored in GIF using 75 × 75 dpi resolution. How many bytes are required to store this image?

....

....

....

....

11.5 ELECTRONIC MAIL

Electronic mail, popularly known as e-mail, is the most widely used form of communication on the Internet and other computer networks for over two decades now. It has grown exponentially to the point where its volume per day is far in excess of the conventional paper mail.

A comparison of fax and e-mail is illustrative of the power of e-mail. Fax communication is terminal-to-terminal, whereas electronic mail communication is user-to-user. Being terminal-to-terminal communication, fax works on a circuit switched mode. E-mail is a store and forward (S&F) service and uses packet switching. In fax, messages destined to a number of users in the same office are sent to one terminal from where it is distributed by an operator or a messenger. The message is open and can be seen by the operator or the messenger. There is no privacy. On the other hand, electronic mail is delivered to the mailboxes of individuals. Privacy is ensured as the mail is delivered to individual's mailbox, which can be opened only by the intended recipient.

Being a person-to-person communication system, electronic mail turns out to be a cheaper alternative to telephone conversation and eliminates the time spent in establishing phone calls. For a telephone conversation to materialise, both the calling and the called party must be present simultaneously. It is not unusual that Mr X calls Y and Y is not present. Some time later, when Mr Y is in, he returns the call only to find Mr X is not at his desk. In fact, some studies indicate that as much as 70 per cent of the business phone calls during business hours do not succeed in the first attempt due to non-availability of the called party. Electronic mail permits communication between two parties without the parties actually being present simultaneously.

Another important advantage of electronic mail is its ability to reduce the consumption of paper in the office. Internal memos and reports can be exchanged electronically without using paper. Being a computer based messaging system, files prepared using office automation packages like word processor, database manager and spreadsheet package can be easily exchanged as electronic mail. This facility has the potential of improving office efficiency considerably.

E-mail systems have two major subsystems:

- User Agent (UA)
- Message Transfer Agent (MTA)

User agent (UA) usually runs on the user machine and the MTA on a mail server. When e-mail services offered by the Internet Service Providers (ISP), the UA functions are available on the web site of the ISP. This is sometimes referred to as web mail. Alternatively, the user may download the messages on to his/her machine and use a local UA for managing mails. Microsoft's Outlook express and MS outlook are examples of UA that run on user machines.

UA performs functions relating to the preparation, submission, and receipt of messages. It also assists the user in other message functions such as filing, replying, retrieving and forwarding. Message transfer agent (MTA) is concerned with transfer of messages across the network. It obtains messages from the source UA and delivers the same to the destination UA. On receiving a message, the MTA performs either a delivery function or a routing function. If the destination UA is in the same local network as the MTA, then the MTA performs a delivery function; otherwise it performs a routing function. The message is sent to another MTA that is en route to the destination. Another important function of the MTA is the reporting of the status of a message. The status may be 'delivered', 'rejected', 'lost', or 'destination unknown' etc.

Submission of a message by UA to MTA is via a protocol called Simple Mail Transfer Protocol (SMTP). Delivery of messages from MTA to the UA uses one of two protocols: Post Office Protocol version 3 (POP3) or Internet Message Access Protocol (IMAP).

We now list and discuss some of the important facilities offered by the UA of current mail systems:

- Mailbox management
- Address book
- Group mailing or mailing list
- Mail acknowledgement
- Mail encryption
- Digital signature
- File attachment

Present day mail services offer powerful mailbox management features. Standard components of mailboxes include:

- Inbox
- Outbox
- Sent Items
- Deleted Items
- Drafts
- SPAM

All these components can be further subdivided as folders much as the way files are organised in offices. For example, inbox may contain a folder for each contact. Items in these components can be arranged in a variety of ways: date wise: received or sent date, name wise, priority wise, subject wise etc. Such arrangements may be in ascending or descending order.

Outbox contains mails that are ready for despatch but not yet sent. The folder 'sent items' contains the messages that have actually been despatched. The folder 'Deleted items' is similar to a waste paper basket that has not been emptied. This folder can be emptied by selecting it and executing a 'delete' command. SPAM folder contains messages that are identified as SPAM, i.e. unwanted. SPAM mail is discussed in Unit 12 on Network Security.

Every e-mail user needs to have one or more e-mail addresses or ids. The address is of the form *username@mailserver.company.com*. It contains a user name, the mail server in which he/she is registered, the name of the company that owns the e-mail server and provides mail services and the top domain name. If the company is unique like *Yahoo*, then separate mail server name may not exist, e.g. rahul@yahoo.com. It is difficult to remember e-mail addresses in the above form. Address book helps us to overcome this difficulty. The easy to remember names of the users and their e-mail addresses are stored in the address book along with a lot more details like phone and mobile numbers, office and home addresses etc. While sending e-mail, user is selected by name and the e-mail address is automatically filled in by the UA from the address book.

Address books have provision to define groups or mailing lists. A group name is defined and individuals are added to this group. Messages may be sent to a group name. In this case, the message is despatched to all those who are part of the selected group. This feature is useful for distributing circulars or in a coordinated multi member project.

A user may request an acknowledgement for delivering a message. It is like registered mail with acknowledgement due. An acknowledgement may be sent as soon as the recipient opens and sees the mail. This may happen automatically or with the consent of the user.

E-mails can be encrypted for maintaining secrecy. Digital signatures may be affixed to indicate authenticity of the message. Encryption and digital signature are discussed in Unit 12 on Network Security.

E-mail systems allow files to be attached as part of e-mail. There is usually a limit on the number files that may be attached and the size of the individual files. There is usually no restriction on the type of file. As a result, files containing multimedia information can be sent with the mail.

Much as the conventional mail, e-mail also has the concept of **envelope** and **contents**. The envelope carries the destination address and other delivery information like priority. The source address is not carried as part of the envelope but is placed as part of the contents. The envelope also has provision to mark carbon copy (Cc) and blind carbon copy (Bcc) of the mail to others. Blind carbon copy hides from the recipient the fact that copy has been marked to others. The contents are delivered to the user entirely. The contents portion has two parts: **header** and **body**. The header part contains the source address and a subject line. The body contains the sender's message.

Although multimedia files can be attached with e-mail, often need was felt for embedding multimedia information in the body of the message itself. This need has led to the development of a protocol called Multi-purpose Internet Mail Extension (MIME). Using MIME message format, multimedia information such as video images and sound bites can be embedded in the message body.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

12) List any four advantages of electronic mail.

13) What are the functions of UA and MTA?

14) Match the most appropriate in the following:

A.	Carbon copy	F.	Message submission
B.	User names	G.	POP 3
C.	SMTP	H.	Envelope
D.	Mail downloading	I.	Mailbox
E.	Sent items	J.	Address book

15) What is group mailing? How it is done?

16) Where is the sender's address placed in e-mail?

....

....

....

....

....

....

....

....

11.6 INTERACTIVE TELEVISION (ITV)

Interactive television (ITV) is a network application for the user. First let us understand what interactive TV means. Today, television is a broadcast service. Different channels transmit pre-scheduled programmes. The user can select a channel of his/her choice and view the programme that is being telecast currently. The user will have to put up with interrupting and often annoying advertisements. The user has no control over what programme he/she can see and when. ITV offers flexibility and control in viewing programmes. Conventional TV transmission may be analog or digital whereas ITV is digital only. Conventional TV is available via sky, satellite or cable communication whereas ITV is available via network. Hence, pre-requisite for ITV is a broadband network communication facility at user premises capable of handling high quality multimedia, i.e. voice, video and data. This network facility allows two-way communication between ITV service provider system and the user TV system. In ITV, one-way video broadcasting system is turned into a two-way multimedia communication system. We may say that ITV is a television set with a return path to the ITV service provider. The return path makes it interactive. Information flows not only from the ITV service provider to the viewer but also from the viewer to the ITV service provider and/or his computer.

When fully developed, ITV would allow users to view a programme of their choice at any time of the day and for any chosen duration without any interrupting advertisements. The programme may be viewed partly, a bookmark created and resumed sometime later. Thus every individual user gets full flexibility of what to view and when. This is what is fully interactive television. Unlike broadcast service where the viewer can only watch one of the many programmes currently being broadcast, ITV gives the viewer an individual choice of content that is exclusive to that viewer.

ITV is not available commercially as of now and is in the evolution stage. ITV is evolving in three stages:

- Broadcast of TV programmes on the Internet
- Multicast of TV programmes on the Videonet
- Unicast of TV programmes on the Videonet

The first requirement of ITV is to make TV programmes available on the network. True ITV is possible only when NII and GII are in place. In the absence of these, attempts are being made to deliver video programmes on the Internet. Each ITV service provider broadcasts one programme at a time as per previously announced schedule. A user may log on to a video site and watch the programme that is being telecast at that time. This is much like conventional television except that the programme is available via Internet. This arrangement calls for Internet access from your TV set. This is provided by many ISPs today. Strictly speaking, there is no real interaction in this case except for logging in. However, this is the first step towards the evolution of ITV. Video is brought to your TV or computer using a technique called **video streaming**. Internet protocols have been evolved to support video streaming. We will learn more about streaming technique in Section 11.8.

In the second stage, the development concentrates on providing the necessary basic infrastructure for ITV. A video network called **Videonet** is being evolved. Videonet is a high-speed network capable of transmitting and distributing high quality video information. It consists of **video servers**, a **backbone network** and a set of **video switches** or **video distribution servers** located in what are called **distribution centres**. ATM with SONET is a suitable candidate for the backbone network. Video servers have massive storage facility and store a large number of programmes to choose from. Depending upon user requests, they despatch many programmes to the video distribution centres via the backbone network. To start with, video distribution centres will contain video switches. Video distribution servers will replace them in the future. Video switches distribute the programmes locally to the end users in multicast mode. Only one copy is received from the video server and distributed to many users locally.

This is like using group address. All the members of the group receive the same programme at the same time.

At this level of evolution, the user interaction improves in two ways: First, the user has a choice of logging on to one of many video distribution centres and second, he/she has a large number of programmes to choose from. Obviously, this is not yet real interaction as no individual can exercise any control on the programme that is being telecast. However, distribution centres may offer some control for the group as a whole. Group multicast is commercially viable because many people have interest in seeing popular programmes at any point of time.

The third stage of evolution aims at full interaction features. Full interaction implies every individual would have the ability to control and view programmes. This can be achieved only by unicast. The distribution centres will send individual video streams to each user who gets full control and flexibility in viewing. Such a service termed as **video on demand**.

How does a user interact with ITV server? Facility for interaction is another area of evolution in ITV. This is also evolving in three stages:

- Programmes are broadcast via Videonet and the user interaction path is provided via telephone network.

- Programmes are broadcast via Videonet and the user interaction path is provided via set-top box.

- Video on demand programmes unicast via Videonet and the user interaction path is via set-top box.

The set-top box is so called because it is usually placed on top of the TV set. It is more than a TV tuner. It has a computer with a phone, coaxial cable or satellite link to the ITV service provider and the Internet. There is a phone modem or a cable modem or a network card, which connects the set-top box to a public data network. Recently, television sets are being manufactured with built-in set-top features so that a separate box sitting on top of the TV is not required. This approach also points to what is likely to be the future home viewing set-up.

The potential use of ITV is enormous. To name a few, ITV may be used for marketing, advertising, child counselling, public relations, education and even politics. For the user, there is the promise of choice, fun, convenience and empowerment. Sitting at home, one will be able to get literally any product or service delivered at the touch of a button. This is known as *t-commerce*, television commerce. Users may click on advertisements to know more about the product. Viewers may choose camera angles while watching their favourite sports event. This indeed is exciting. Users may pause and resume programmes that they are viewing, provided the programmes are not live and are being delivered from video distribution servers located in video distribution centres. With live programmes that are not available on video servers, the users may record the programmes for later viewing.

ITV does not come free. Users will have to pay for ITV services. Two modes of payments are in vogue. First, a user pays a fixed subscription. Second is the *pay-per-view* model. In the latter, the user pays only for the programme s/he views. The two models are analogous respectively to the post-paid and pre-paid schemes that are used in mobile networks. Obviously, broadcast services are the cheapest and the unicast services are the most expensive. Multicast services fall in between.

There is one serious aspect of ITV, which is now spreading to Internet. All ITV systems have a feature called *click stream analysis.* This feature creates a complete record of the clicks that a user performs on his/ her set-top box. This record is later analysed to build profiles of users. In a positive sense, the purpose is to provide the user information that s/he is interested in a focussed manner. The negative aspect is that the service provider is actually treading into the private life of individuals and may use the information collected to blackmail users.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

17) Discuss ITV features.

18) What are the basic infrastructure components of ITV?

....

....

....

....

....

11.7 INTERACTIVE MUSIC

Interactive Music (IM) is an interactive network application. It is similar to ITV. It is also called **music-on-demand**. The user has flexibility and full control in listening to music on the network like Internet. The control includes functions like *pause, resume, stop, fast forward* and *fast backward*. Pause and resume are similar to creating bookmarks and continuing from thereon later. Music-on-demand services are evolving in two directions:

- Delivering music on the Internet in broadcast mode. This further evolves to **Internet radio**.
- Establishing audio servers and audio distribution switches to deliver audio programmes in unicast mode.

Unlike in the case of ITV, delivery of chosen music on multicast mode is not very attractive from a commercial point of view. It is unlikely that many people would choose the same song or music programme at the same time except when some major musical programmes take place, which are far and few. For example, many may hear a programme of Michael Jackson or a music festival but may not choose the same music on a day-to-day basis. In the case of ITV, popular TV programmes appear almost on daily basis. For example, there are over 500 movies are produced in a year. In addition, there are innumerable other TV programmes. A large number of persons have interest in viewing the same video programme at the same time. Hence, it makes sense to provide multicast for IRV. As in the case of ITV, charging for music on demand can be based on subscription or *pay-per-listen* mode. Pay-per-listen generally has an initial payment as well.

There are some differences between music-on-demand and Internet radio. In music-on-demand, the music bit is selected by the user whereas in Internet radio, the station plays out the programme. There is no user interaction for selection. User also has no control like *pause, resume* etc. Some radio stations play a second channel that is delayed by about 10 minutes to allow the users to take a break. After the break, the user may switch over to the delayed channel and listen to the programme from where s/he left. Another important difference is that music-on-demand is a unicast service whereas Internet radio is a multicast service. Many persons listen to Internet radio at a time. Hence, the same audio stream has to be sent to different destinations.

Audio is brought to your music system or computer using a technique called **audio streaming**. Internet protocols have been evolved to support audio streaming. We will learn more about streaming technique in Section 11.8.

11.8 APPLICATION DELIVERY

There are two fundamental ways in which music or video programmes are delivered on the network:

- Streaming technique
- Use of distribution servers

Both techniques are used for delivering audio and video. Streaming is more suitable for music. Distribution server approach is better suited for delivering video programmes. We discuss these two approaches in the following. We discuss *streaming* in the context of audio delivery and distribution server approach in the context of video delivery.

Streaming Audio

Streaming audio application is designed under client-server architecture. There is a media server on the server side in addition to the usual web site. There is a media player on the client side in addition to the usual web client, i.e. web browser. The media server and the media player interact via the convergent *real time protocol* (RTP). Initially, the user interacts with the server via the HTTP browser protocol for selecting the music bit. Thereafter, the media server and the media player come into picture to play the music in real time. The streaming audio configuration is depicted in Fig. 11.2.

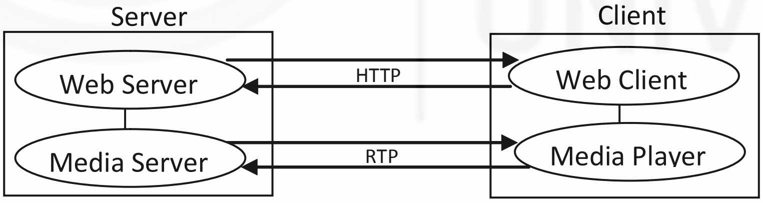


Fig. 11.2: Configuration for music delivery on Internet

In streaming audio, the digitised music is formed as small packets containing about 5 – 8 ms of music. Each digitised sample is 8 bits in size and is generated every 125 ìs. Each sample is not transmitted as it is generated. About 40 - 64 samples are grouped together as a packet and sent. Since packet transmission is on store and forward basis, different packets may take different time to reach the destination. Some packets may even be lost on the way. Delay variation and packet loss cause jitters in the music affecting the quality. Streaming audio takes care of these problems by providing a music buffer at client's end.

At the start of a listening session, the buffer is filled with music samples before the music is played out to the user. The buffer holds about 15-20 seconds of music. The user has to wait only for a short while until the buffer is filled to the required level before he can start listening to music. The music packets continue to arrive from the server while the media player is playing out the music and emptying the buffer contents. This is a streaming operation, i.e. on the one side the buffer is being emptied and on the other it is being filled. Hence the scheme is named as *streaming audio*. The provision of the buffer ensures that continuous music is available even in the presence of variable packet delays. The streaming operation is similar for video distribution as well and usually called **video streaming**.

Video distribution

Video distribution is done via **Videonet**. Videonet is a high-speed network capable of transmitting and distributing high quality video information. It consists of three parts: **video transmission centres (VTC)**, **backbone network** and a set of **video distribution centres (VDC)**.

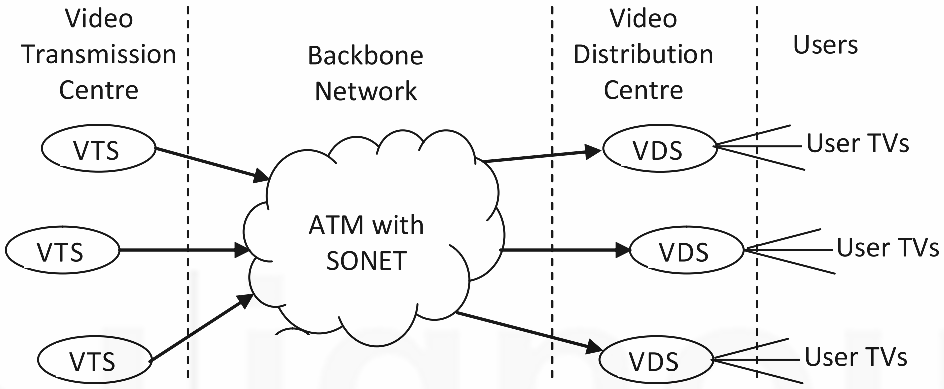


Fig. 11.3: Videonet Configuration

Video transmission centres house **video transmission servers (VTS)** or simply **video servers**. Video distribution centres house **video distribution servers (VDS)**. Video switches are used in place of distribution servers in case of video streaming. The configuration of Videonet is shown in Fig. 11.3 using VDS. Backbone network is broadband network like ATM with SONET. Video transmission servers have massive storage facility and store a large number of programmes to choose from. To get an idea of the quantum of storage required, let us consider storing movies. On an average, a movie requires 4 GB of storage. To store 10,000 movies, we would need storage of 40,000 GB or 40 terabytes (TB). With the present day technology, such capacities can only be obtained by using what are called 'disk farms'. Disk farms are a collection of disk arrays interconnected by a very high-speed network structure offering very large storage capacities.

Depending upon user requests, VTS despatch many programmes to the video distribution centres via the backbone network. Video distribution servers store the programmes locally and distribute the same to the end users in unicast or multicast mode. VDS is similar to VTS except that they have much smaller storage capacity, say 80 GB to store about 20 movies. The fundamental difference between video switches and VDS is that video switches do not have storage facility and stream the video to the users from the transmission servers in real time. VDS store the most recently requested video programmes by the users. Popular programmes remain locally in VDS, since many users view them. If a programme is available in VDS, the access is instantaneous. If it is to be downloaded from VTS, there is considerable delay. When a user makes a request for a programme, s/he is informed of the availability of the same locally or otherwise. VDS acts like cache memory for VTS.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

19) Why is streaming more suitable for music than the use of distribution servers?

20) Where is the buffer located in audio streaming? What purpose does it serve?

21) What are disk farms? Why do we need them?

....

....

....

....

11.9 PERFORMANCE ISSUES

As mentioned earlier, Internet is packet switched network designed for data transport. Packet switching is ideally suited for data transfer. Attempting to deliver real time services like music and video on a packet switched network gives rise to performance issues on account of the following:

- Time taken for packets to travel from source to destination varies from packet to packet.
- Packets may arrive out of sequence.
- Packets may be lost.

Performance issues arriving out of the first two can be taken care of by providing buffers at the receiving end and suitably managing them. The performance issue arising out of the third can be minimised by some special techniques. Let us see the issues and the solution thereof in the following paragraphs.

In packet switched networks, the packets are moved from node to node until the respective destination nodes are reached. At every intermediate node, packets may experience queuing delay. Depending on the queue lengths the wait times for the packets vary. Hence, even with the same route different packets may take different times to reach the destination. In real time services, such delay variation causes jitter. The sound reproduced at the receiving end is not natural and the video flickers. If the packets are delivered out of sequence, the resulting audio or video is distorted. If the packets are lost, the signal is broken. Parts of the audio and video are missed.

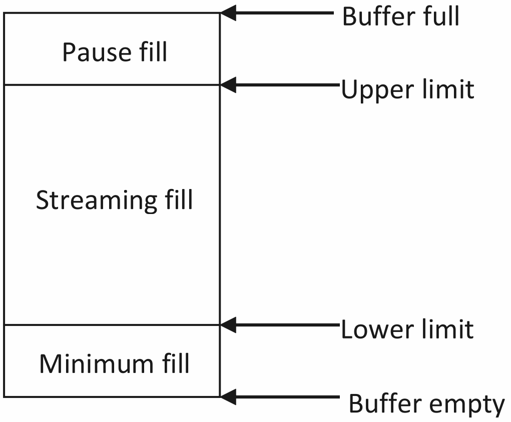


Fig. 11.4: Buffer management for streaming

Let us now see how buffering at the receiving end helps to overcome some of these problems. A typical buffer organisation is shown in Fig. 11.4. Buffer is part of media player at the user end. Refer to Fig. 11.2. When the user makes a request for a music or video bit to be played, the buffer is empty. The media server and media client interact and information streaming is started. The buffer is being filled. When the fill reaches the lower limit level, i.e. the minimum fill portion of the buffer is full; the media player starts playing the information to the user. The minimum fill portion of the buffer holds about 15-20 seconds of information. The user has to wait for this short period before s/he can start listening to music or viewing the video. While the media player is playing out, the media server continues to stream and fill the buffer. When the fill reaches the upper limit level, the media player asks the media server to pause the flow momentarily. 'Pause' is a flow control command. It controls the flow of information. There is a small time gap before the pause command takes effect. Information would continue to stream during this period and is received in the 'pause fill' portion of the buffer. The flow then stops but the media player continues to play the information to the user. The buffer is being emptied. When the information level reaches the lower limit level, the media player issues the flow control command 'Resume' to the media server. The information flow then resumes. There is a small time gap before the resume command takes effect. In the meantime, the media player continues to play from the minimum fill portion of the buffer to the user. On the one side the buffer is being emptied and on the other it is being filled. Delay variation occurring at the input side of the buffer does not affect the play out to the user. Out sequence packets can be sequenced on the input side without disturbing the flow on the output side. Thus user gets continuous and smooth information.

All flow control commands like 'pause' and 'resume' are implemented using a protocol known as *real time control protocol* (RTCP), which is used in conjunction with RTP. There are two other protocols associated with streaming services. They are protocols used for establishing a real time session before streaming operation starts and managing the stream flow once streaming starts. The protocols are *session initiation protocol* (SIP) and *real time streaming protocol* (RTSP) respectively.

Loss of packets results in break in signal to the user. The user will observe breaks in the information. In real time services, lost packets cannot be recovered in time to fill the gap. In order to minimise the effect of such packet loss, some media severs send alternate samples in packets, say odd numbered samples in one packet and even numbered ones in another. The two packets together provide the full information. If one packet is lost, the alternate samples are lost and the effect may not be perceivable by the user. If full packets of sequential samples are lost, the effect will be perceivable to the user.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

22) What purpose does the lower limit in buffer serve?

23) The size of buffer required for video streaming is much larger than that required for audio streaming. Why?

....

....

....

....

11.10 WHY NETWORK MANAGEMENT?

The complexity of networks is ever increasing. More and more users are converting their single device end equipment into multiple workstation local area networks (LAN). These LANs connect with each other via wide area networks (WAN) resulting in the well-known Internet infrastructure. At the applications level, the Internet infrastructure has given birth to what are known as **Intranets** and **Extranets**. An Intranet is a secure web-based private network that supports the business requirements of a corporate body across different geographical locations. An Extranet is an extension of the Intranet that connects a company's Intranet to the networks of its business partners and selected customers and suppliers. Both Intranets and Extranets operate on the same principles as Internet except that they are more secure closed networks amongst an identified set of users. In contrast, Internet is a public network.

Clearly, the management of such a complex network is not possible by manual means and calls for a set of automated and well-defined management tools and applications. To ensure interoperability of network management systems the international bodies like ISO, ITU and Internet Society have undertaken extensive studies and standardisation in the area of network management.

Historically, ISO is the first international body to give attention to network management aspects. As early as 1979, ISO initiated standardisation activities for network management as part of its standardisation process for ISO-OSI reference model. Its initial concern was to deal with management aspects of data networks. Later the work generalised to cover all types of telecommunication networks. ISO's network management standards are broadly covered under the subject heading **Common Management Information Protocol** (CMIP). ITU started working on management aspects in the 1980s. Its recommendations are covered under the subject **Telecommunications Management Network** (TMN).

The development of network management products based on ISO/ITU standards took time due to the efforts directed towards a comprehensive definition of features in these standards. In the meantime, the unprecedented growth in Internet led to a pressing and urgent need for network management tools for Internet. The Internet Society through its Internet Architecture Board (IAB) and its subordinate group Internet Engineering Task Force (IETF) decided to introduce a simplified version of network management derived from ISO/ITU proposals. In 1989, a standard known as **Simple Network Management protocol** (SNMP) was adopted for TCP/IP based Internets. The implementation of SNMP turned out to be a remarkable success demonstrating that the much needed network management tools actually solved real-life problems. The diffusion of SNMP, soon after its introduction, far exceeded the general expectations resulting in the definition of SNMP version 2 in 1993 and SNMP version 3 in 1998.

11.11 SIMPLE NETWORK MANAGEMENT PROTOCOL (SNMP)

CMIP, TMN and SNMP concentrated mainly on WAN management tools. In the early 1990s, it was realised that local networks (LAN) too require considerable network management support for a variety of purposes that we elaborate in the next section. For meeting such needs, a supplement to SNMP known as Remote Monitoring of Network (RMON) was issued in 1993 and an upgraded version RMON2 was issued in 1995.

Network management functions are placed under five broad areas as defined by ISO and adopted by ITU and Internet Society. These areas are:

- Fault Management
- Configuration and Name Management
- Accounting Management
- Performance Management
- Security Management

The order in which these areas are listed above leads to an easy to remember acronym formed by taking the first letter of each area: F-CAPS. Simple Network Management Protocol (SNMP) and Remote Monitoring (RMON) together are designed to support the above management functions. SNMP has three major components:

- The protocol itself
- Structure of Management Information (SMI)
- Management Information Base (MIB)

The protocol defines the format of SNMP messages and the rules on how the messages are exchanged. SMI specifies rules to name and define individual objects that need to be managed. It is also used to define the type of management information to be collected. MIB, as the name implies, is a database of all information objects with their attributes or variables and their values. MIB is maintained by each device and contains information about the managed objects in that device. Any device like router, bridge or switch is usually called ‘Managed Device’ if it implements SNMP.

SNMP is envisaged for use by environments defined by ISO, ITU and IAB (Internet). SMI allows definition of data objects for each of these agencies. For Internet SMI defines eight categories under which information can be maintained in MIB:

System: Information about the hardware, operating system and operations of the device. Example variable: System up time.

Interfaces: Information about individual network interfaces in a device. Example variable: Number of interfaces.

Address Translation: Information about address mappings: Example variable: IP address and the corresponding NIC address.

Internet Protocol: Statistics about IP. Example variable: No of datagrams received.

Internet Control Message Protocol: Statistics about ICMP. Example variable: Number of ICMP messages received.

Transmission Control Protocol: Statistics about TCP. Example variable: Number of TCP segments received.

User Datagram Protocol: Statistics about UDP. Example variable: Number of UDP datagrams received.

Exterior Gateway Protocol: Statistics about EGP. Example variable: Number of EGP messages received.

SMI and MIB are independent of the network management protocol used. Following SMI rules, vendors can define their own MIB variables in any of the categories. This is very useful for testing new products or enhancements to existing products in real network environment.

So far we have discussed about how management information is defined and organised. This information needs to be accessed and acted upon by the Network Management System (NMS). Access to management information is provided by the protocol. SNMP supports five commands for this purpose:

Get request: Fetch the value of a specified variable. This command is usually issued by NMS.

Get-next request: Fetch the value of the next variable. Name of the variable is not specified.

Get response: Reply by a device to a fetch command from NMS. The requested value is sent in this reply.

Set request: Issued by NMS to store a value in a specified variable.

Trap: Specify an event, on the occurrence of which the device sends a response to the NMS.

As is seen from the foregoing discussions, SNMP provides a means to specify and organise management information variables and to access the same by a set of commands.

As mentioned earlier, RMON was defined to support network management functions in LANs. They include the following:

- Monitor traffic type and network usage
- Identify network problems quickly
- Plan for future growth of the network
- Protocol decoding
- Protocol usage

RMON2 allows collection of statistics regarding protocol usage in a variety of ways such as:

- Network segment wise
- Network address wise
- Traffic wise between a pair of network addresses
- Application wise

Protocol analyser and LAN explorer are two devices that are used extensively for monitoring LAN segments under SNMP and RMON environment.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

24) What are the main components of SNMP?

25) What is RMON? What is its main purpose?

....

....

....

....

11.12 SUMMARY

This unit covered main user applications that run on Internet and network management. The applications discussed include non-real time ones like text and multimedia messaging and real time ones like interactive television and music on demand. To start with, a distinction is made between application layer network services and user applications. A user requires network services to be able to run his/her applications. User applications fall under two broad categories: interactive and distributive applications. Interactive applications may be conversation, messaging or retrieval oriented. Distributive services may be broadcast, multicast or unicast. Cyclic services are also distributive. In addition, two other forms of distributive services viz. cluster cast and any cast are also discussed.

Under electronic text messaging, the unit covers Small Messaging Service (SMS) in detail. SMS is of two types: Cell broadcast SMS and point-to-point SMS. In cell broadcast SMS, a base station transmits short messages to all the mobiles in its coverage area. Point-to-point SMS is between two mobile users or one mobile user and a landline user. Every mobile network has Small Message Service Centre (SM-SC) through which all SMS messages pass. SM-SC serves as a clearinghouse for short messages. Mobile-terminating short messages can be user specific, mobile equipment specific or SIM specific.

The unit then covers Multimedia Messaging Service (MMS). Transmission of multimedia information demands much higher bandwidth. MMS in 2G mobile networks supports only still picture transmission. 3G mobile networks have larger bandwidth and hence can support the transmission of motion video. Digital image is formed as a matrix of dots called *pixels.* The horizontal and vertical dot densities together are called the *resolution of the image*. Larger the resolution, the better is the clarity of the picture and the bandwidth required is higher. Three formats are commonly used for storing and transferring digital images: TIFF, GIF and JPEG.

Electronic mail application is then discussed. E-mail systems have two major subsystems: User Agent (UA) and Message Transfer Agent (MTA). UA performs functions relating to the preparation, submission, and receipt of messages. It also assists the user in other message functions such as filing, replying, retrieving and forwarding. Message transfer agent (MTA) is concerned with transfer of messages across the network. Some of the important facilities offered by the UA include mailbox management, address book, group mailing or mailing list, mail acknowledgement and mail encryption. Mailbox management covers management of inbox, outbox, sent items etc.

The discussions in the unit then turn to real time applications. Both Interactive Television (ITV) and Interactive Music (IM) offer flexibility and control in viewing and listening to programmes. ITV is also known as Video-on-demand and IM as music-on-demand. Both are two-way multimedia communication systems. The return path makes the systems interactive. Information flows not only from the service provider to the viewer but also from the viewer to the service provider. The systems allow users to view or listen to a programme of their choice at any time of the day and for any chosen duration. The programme may be viewed/heard partly, a bookmark created and resumed sometime later. Every individual user gets full flexibility of what to view/hear and when. Pay-per view or pay-per-listen are charging schemes in ITV and IM applications respectively.

Delivery of real time programmes on packet switched networks is discussed next. Streaming audio or streaming video technique is used on packet networks. Distribution servers are used on Videonet, which is a high-speed network capable of transmitting and distributing high quality video information. Performance issues relating to application delivery are then discussed. In packet networks, delay in packet delivery, out of sequence reception and packet loss give rise to performance issues.

Finally, network management aspects and the associated protocol for Internet are discussed. The ever-increasing complexity of networks calls for a set of automated and well-defined management tools and applications. ISO, ITU and Internet Society have undertaken extensive studies and standardisation in the area of network management. The respective standards are Common Management

11.13 ANSWERS TO SELF-CHECK EXERCISES

Information Protocol (CMIP) from ISO, Telecommunications Management Network (TMN) from ITU and Simple Network Management protocol (SNMP) from Internet society. Local networks (LAN) too require considerable network management support. A supplement to SNMP known as Remote Monitoring of Network (RMON) was issued for this purpose. SNMP has three major components: the protocol, Structure of Management Information (SMI) and Management Information Base (MIB).

1) Video surveillance system is an interactive application. The reason is that there are human beings monitoring the video images generated by the cameras. They can control the system. They can choose which camera to monitor or even pan the camera. Hence it is interactive.

2) An example of commonly used text messaging system is SMS in mobiles.

3) Multicast is a form of limited broadcast. In broadcast, information is distributed to all users, whereas in multicast, it is distributed to a group of users who form a subset of all users.

4) In multicast, the central source distributes information to all users in a group directly. In cluster cast, the central source distributes information to any one user in the group who in turn distributes the information to other users in the group.

5) In unicast, the central source distributes information to a specific user using point-to-point connection. In any cast, information is sent to any one of the users in a specified group.

6) In cell broadcast SMS, a base station transmits short messages to all the mobiles in its coverage area. Point-to-point SMS is between two users.

7) Cell multicast may be used as the solution as explained in the following. Since the promotional material is to go only to registered users, general broadcast cannot be used. We need to use group address for the registered users and multicast the messages to them. Registered customers may be spread out in different geographical areas. Therefore, all the concerned cells in different geographical areas must be used in multicast mode to achieve the functionality desired.

8) SMS-SC receives, stores and forwards the short messages. It acts as the input/output interface for non-mobile devices.

9) Wireless access protocol (WAP) is a mobile functionality. Hence, we require ME-specific (mobile equipment specific) SMS. In some cases, a SIM-specific message may also be required as the SIM card might perform some WAP related functions.

10) SMS language may be as “J 2 C U 2morow”

11) The picture size is 1" × 2". The resolution is 75 × 75 dpi. Therefore, the total number of pixels is (1 × 75) × (2 × 75) = 11,250. The picture is stored in GIF format. GIF uses 8 bits, i.e. one byte to represent pixel values. Therefore, the number of bytes required to store this image are 11,250 bytes.

12) Advantages of e-mail are listed below. Any four from the following list may be given as answer:

- Electronic mail communication is user-to-user and not terminal-to-terminal communication as in the case of fax.
- The mail is delivered to individual's mailbox and can be opened only by the intended recipient. Therefore, privacy is ensured.
- E-mail service uses store and forward (S&F) and packet switching form of communication. Hence, it uses network resources more effectively. It is cheaper than circuit switched services like fax.
- Local distribution of faxes call for a person whereas the distribution is automatic in e-mail.
- Electronic mail is a cheaper alternative to telephone conversation and eliminates the time spent in establishing phone calls.
- Electronic mail permits personal communication between two parties without the parties actually being present simultaneously on the network.
- Electronic mail reduces the consumption of paper in offices. Internal memos and reports can be exchanged electronically without using paper.
- Being a computer based messaging system, files prepared using office automation packages like word processor, database manager and spreadsheet package can be easily exchanged with electronic mail. This facility has the potential of improving office efficiency considerably.

13) User Agent (UA) performs functions relating to the preparation, submission, and receipt of messages. It also assists the user in other message functions such as filing, replying, retrieving and forwarding.

Message transfer agent (MTA) is concerned with transfer of messages across the network. It obtains messages from the source UA and delivers the same to the destination UA. On accepting a message, the MTA performs either a delivery function or a routing function. If the destination UA is in the same system as the MTA or is attached to the MTA directly, then the MTA performs a delivery function; otherwise it performs a routing function. Another important function of the MTA is the reporting of the status of a message such as 'delivered', 'rejected', 'lost', or 'destination unknown'.

14) Most appropriate matching are:

10.0 A – H

10.1 B – J

10.2 C – F

10.3 D – G

10.4 E – I

15) Electronic mail subscribers can be placed in groups. Mail can be sent to groups and it gets automatically delivered to all the members of the group. This is called group mailing. A group is also referred to as mailing list.

    Address books have provision to define groups or mailing lists. A group name is selected and individuals are added to this group. When a message is sent to a group it is delivered to all those who are part of the selected group.

16) The sender's address is placed in the header part of the contents portion.

17) ITV offers flexibility and control in viewing programmes. ITV is digital. ITV is available via a broadband network that allows two-way communication between ITV service provider system and the user TV system. ITV allows users to view a programme of their choice at any time of the day and for any chosen duration without any interrupting advertisements. The programme may be viewed partly, a bookmark created and resumed sometime later. Thus every individual user gets full flexibility of what to view and when. Unlike broadcast service where the viewer can only watch one of the many programmes currently being broadcast, ITV gives the viewer an individual choice of content that is exclusive to that viewer.

18) The basic infrastructure of ITV is Videonet. The infrastructural components of Videonet are video servers, a backbone network and a set of video switches or video distribution servers located in what are called video distribution centres.

19) In streaming, information is directly transferred to the user from the server. Distribution servers store information locally and distribute the same to the users. This is efficient if many people are likely to choose the same programme over a short period of time. It is unlikely that many people would choose the same music on a day-to-day basis whereas popular video programmes are viewed by a large number of people. Hence storing music locally is unnecessary overhead and not commercially remunerative. Therefore, streaming is better for music.

20) In audio streaming, the buffer is located at user end. The provision of the buffer ensures that continuous and smooth music is available to the user even in the presence variable packet delays.

21) Disk farms are a collection of disk arrays interconnected by a very high-speed network structure. We need them to build storage capacities in terabytes range for storing large quantity of information, particularly video.

22) Lower limit level of the buffer has two purposes. First is at the starting time when the buffer is empty. When the information fill reaches the lower limit level i.e. the minimum fill portion of the buffer is full, the media player starts playing the information to the user. Second is when the information is being played out and the server is in the pause state. In this state, the buffer is being emptied. When the information fill depletes to the lower limit level, the media player issues the flow control command 'Resume' to the media server. The information flow then resumes. There is a small time gap before the resume command takes effect. In the meantime, the media player continues to play from the minimum fill portion of the buffer to the user.

23) Video occupies larger bandwidth. The rate of the digitised video is much larger than that of the rate for audio. More bits are received per second. To store about 10 seconds of video, we require more storage space than that is required for 10 seconds of audio. Hence, the size of buffer required for video streaming is much larger than that required for audio streaming.

24) The main components of SNMP are the protocol, structure of management information (SMI) and management information base (MIB).

25) RMON stands for Remote Monitoring of Network. It is the tool for monitoring management information parameters in LANs.

11.14 KEYWORDS

Any cast : A form of information distribution: To Any one of the entities in a group.

Broadcast : A form of information distribution: To all entities in a network.

Cell broadcast SMS : A form of short message (SM) distribution: To all the mobiles in cell's coverage area.

Cluster cast : A form of information distribution: To the nearest (usually) entity which in turn distributes the information to all other entities in a group.

CMIP : Common Management Information Protocol evolved by ISO.

GII : Global Information Infrastructure.

Group mailing : A feature for sending e-mail to all users in a group.

Interactive Music : An interactive network application where the user has flexibility and full control in listening to music.

Interactive Television : An interactive network application where the user has flexibility and full control in viewing video programmes.

Media player : Software program at the user end capable of receiving and playing multimedia programmes to the user.

Media server : Software program at the server end capable of sending multimedia programmes to the media player.

MMS : Multimedia Messaging Service.

MTA : Message Transfer Agent: A component of e-mail systems.

Multicast : A form of information distribution: To all entities in a group.

Multimedia : Information containing audio, text and/ or video.

Music-on-demand : Another name for Interactive Music.

NEIS : Networked Electronic Information Society.

NII : National Information Infrastructure.

Pay-per-listen	: A form of payment in Interactive Music.
Pay-per-view	: A form of payment in Interactive Television.
Point-to-point SMS	: Short Message exchanged between two users via SM-SC.
RMON	: Remote Monitoring of Network
RTP	: Real Time Protocol
SMI	: Structure of Management Information: A component of SNMP.
SM-SC	: Small Message Service Centre.
SNMP	: Simple Network Management Protocol.
Streaming	: Technique used for sending audio or video in real time.
TMN	: Telecommunications Management Network: A network management standard evolved by ITU.
UA	: User Agent: A component of e-mail systems.
Unicast	: A form of information distribution: To only one user.
Videonet	: A high-speed network capable of transmitting and distributing high quality video information.

11.15 REFERENCES AND FURTHER READING

Homer, Douglas E. *Internetworking with TCP/IP*, Volume I. 3rd Edition. New Delhi: Prentice Hall of India, 2001. Print

Lin, Yi-Bing. *Wireless and Mobile Network Architectures*. Singapore: John Wiley & Sons (Asia), 2001. Print

Mansfield, Kenneth C and Antonakos, James L. *An Introduction to Computer Networking*. New Delhi: Prentice Hall of India, 2002. Print

Simoneau, Paul. *SNMP: Network Management*. New Delhi: Tata McGraw-Hill, 1999. Print

Stallings, William. *ISDN: An Introduction*. New York: Macmillan Publishing Company, 1989. Print

Stallings, William. *ISDN and Broadband ISDN with Frame Relay and ATM*. 4th Ed. Singapore: Pearson Education Asia, 2001. Print

Tanenbaum, A. S. *Computer Networks*. 4th Edition. New Delhi: Prentice Hall of India, 2002. Print

Verma, Pramod K. *ISDN Systems: Architecture, Technology, and Applications*. New Jersey: Prentice-Hall Inc, 1990. Print

Viswanathan, Thiagarajan. *Telecommunications Switching Systems and Networks*. New Delhi: Prentice Hall of India, 2010. Print

UNIT 12 NETWORK SECURITY

Structure

12.0 Objectives

12.1 Introduction

12.2 Why Information Security?

12.3 Types of Attacks

12.4 AAA Security

12.5 Firewalls and Proxy Servers

12.6 Web Security

12.7 Malicious Software

12.8 Viruses

12.9 Spyware, Spam, Phishing and Cookies

12.10 Encryption

12.11 Digital Signature

12.12 E-mail Security

12.13 Summary

12.14 Answers to Self-check Exercises

12.15 Keywords

12.16 References and Further Reading

12.0 OBJECTIVES

After going through this Unit, you will be able to understand and appreciate:

- Different aspects of and importance of information security;
- Internal and external threats to organisations;
- Computer system security and security over networks;
- Information properties prone to attacks;
- Information security attackers and their attacks;
- Authentication, authorisation and audit features;
- Password design and password policies;
- Password authentication process;
- Firewall features and the use of proxy servers;
- How secure transactions are done on web sites;
- Internet security features such as anti-virus, anti-spyware etc.;

- What are spam mails, phishing and cookies;
- Encryption and digital signature techniques;
- Why digital certificates are required; and
- How to send secure e-mail.

12.1 INTRODUCTION

In the pre-computer era, information in an organisation used to be secured physical and administrative measures. For example, a conventional library has manual checks at the entry/exit point. In other office areas, secure cabinets and restricted physical access were used for information security. With the advent of computerisation and networking, information is being stored and transmitted electronically. As a result, information security has shifted largely to electronic domain. Some aspects like access to server room are still controlled by physical means. In this unit, we are concerned only with securing electronic information and not with the physical aspects of security. Electronic locks are a part of physical security and are not discussed here. We are purely concerned with securing information that is either stored or transmitted electronically.

The electronic information security measures fall under two different classes: **computer security** and **network security**. The measures used for securing information stored inside a computer fall under the class computer security. The measures used while transmitting information over a network fall under the class network security. The business nature of an organisation and its interface with the external world would determine the extent of measures required in each class. In this unit, we shall be studying about both the classes of information security.

Threat to information resources may come from within an organisation or from external sources. Accordingly, we have **internal threats** and **external threats.** For example, internally, an employee of the organisation may try to tamper with the financial data to gain pecuniary benefits. Externally, a competitor organisation may try to steal the strategic plans of a company. Security measures are required in both cases. The extent of measures is usually based on the threat perception in each case. Threat is evaluated by studying the nature of the business of an organisation. The external threat is high in the case of organisations that provide public information services as in the case of a library or are involved in sensitive or controversial activities. On the other hand, an organisation that has an Internet connection for the benefit of its own employees may not face serious external threats. In this unit, we shall be studying about security measures to deal with both internal and external threats.

12.2 WHY INFORMATION SECURITY?

The answer to the question 'Why information security?' is as simple or complex as the answer to the question 'Why security in the normal life of society?' It is because there are some persons in the society who attempt to derive benefits from required to secure attempts to destroy others' resources in a surreptitious (covert) manner. Such attempts are to be prevented and thwarted. Security is meant to do this. Information is a valuable commodity and some persons attempt to derive benefits by exploiting others' information resources. Attempts may also be made to destroy other's information resources to cause harm to the owner. Hence there is the need to secure information resources. In the context of information security, one may associate four properties to any piece of information. These properties are:

- Availability
- Confidentiality or Secrecy
- Integrity
- Authenticity.

If information is not secured, one or more of these properties may be compromised. For example, if a communication link is broken or a hard disk stolen, then information becomes **unavailable**. Information must be available to bona fide users during specified times. Genuine accesses should not be denied. An attempt to make information unavailable is formally referred to as **denial of access** attempt. If an unauthorised person gets access to some information, then the **confidentiality** associated with that resource gets lost. This amounts to illegal copying of information even though it is not physically erased. If an unauthorised person modifies the contents in an information resource, then the **integrity** of information is lost. In such a case, the internal consistency of information is lost. For example, the contents of a philosophical website dealing with the axiom 'Unity in Diversity' is modified to include biodiversity and eco systems. In such a case the consistency of information is lost. Finally, if an unauthorised person is able to insert some counterfeit information objects into the system, then the **authenticity** of information is lost. This is equivalent to committing fraud. For example, an impostor may post a circular in the name of the Government of a country. Thus, information security is essential to preserve all the four properties.

People who tend to violate information security provisions are generally known as **intruders**. They are usually put under two categories: **attackers** and **hackers** of information resources. Attackers are more harmful than hackers. Attackers are further classified as **passive** or **active** attackers. Similarly, hackers are also classified as **white hat** or **black hat** hackers.

The main aim of hackers is to gain unauthorised access to information resources. White hat hackers attempt to expose security loopholes in a system and usually do not cause serious harm. They tend to annoy the owners of the system. Black hat hackers exploit security loopholes for personal gains or to harm someone. A typical example is an employee who tampers pay records to get monetary benefits for self or to cause monetary loss to a colleague. Another example is to change the contents of a website to make fun of the site. A reported case includes that of a hacker who modified the contents of a medical website dealing with abortions to preach religious teachings against abortions.

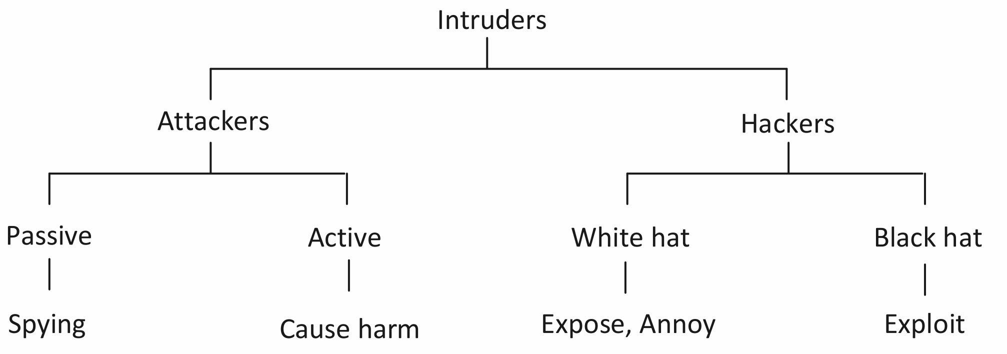


Fig. 12.1: Intruders of Information Systems

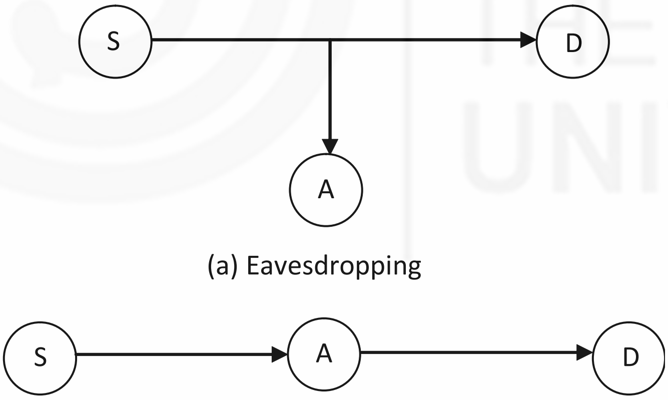
Attackers tend to break the security system and cause serious harm to the owners of the system. Their scale of operation is usually large and often borders espionage. Attacks on military sites and sensitive technology sites fall in this category. We will learn more about the type of attacks in the next section. Figure 12.1 summarises the intruder classification. The dividing line between black hat hackers and attackers is somewhat blurred. It is usually the scale of operation that distinguishes between the two. While a black hat hacker causes limited damage, an attacker could cause massive damage to the owner of the information resource.

12.3 TYPES OF ATTACKS

Passive attacks involve eavesdropping, listening to the traffic, monitoring the transmission parameters such as time and frequency of transmission, message length, and destination address. The contents of the message are not accessed. In the case of active attacks, the message and the transmission are tampered with in a variety of ways. Usually passive attacks precede active attacks. In the context of both passive and active attacks, the potential security threats may be listed as follows:

- Unauthorised access
- Eavesdropping
- Masquerade
- Modify
- Delete
- Delay
- Replay
- Reroute
- Misroute
- Flood network

Before we elaborate on the above security threats, let us see how attackers organise their attacks. Attackers use basically two techniques to organise their attacks and break the security system. These techniques are illustrated in Fig. 12.2.



(a) Eavesdropping

(b) Interspersed attacker

Fig. 12.2: Security attack modes

The normal flow of information is between a source (S) and destination (D). A passive attacker (A) just listens to the channel as shown in Fig. 12.2(a) and does not disturb the normal flow of information. An active attacker positions self in between the source and destination as shown in Fig. 12.2(b) and causes a variety of harmful effects. An incoming message may be blocked or deleted causing an interruption in communication between the source and destination. The incoming message may be modified and transmitted to the destination. The attacker may copy the credentials of the source and masquerade (pose oneself) as the sender and originate new messages on behalf of the sender. The message may also be delayed if time is critical in delivery. Replaying the message may cause the same action to be taken more than once, e.g. double payment. Rerouting or misrouting can be done when the attacker has access to alternative communication channels. Rerouting is usually done to send the message via the attacker's message analysis centre. The message may be copied there, analysed and even modified. In networks, routes are often chosen dynamically depending upon the traffic conditions. Hence, the destination may not be able to detect the deliberate rerouting at all. Misrouting is done to misguide the recipient and possibly cause a wrong action to be taken. For example, an army command to move the troops intended for western sector may be misrouted to eastern sector causing troop movement in the wrong sector. Rerouting and misrouting are serious forms of attacks.

An unprotected information system that is exposed to threats faces many **risks**, some of which are listed in the following:

- Information theft
- Unauthorised use or misuse of system resources
- Theft of services, i.e., steal and offer the same service at a lower cost
- Denial of service, i.e., genuine users is denied of services. This may happen if the attacker floods and chokes the network with unwanted information flow.

In view of the above risks, an information system must be protected. But, protection does not come free. There is a cost associated with implementing security measures. Ideally, one would like to have the best security measures. But the cost of such a solution may be prohibitive. Hence, the cost of security measures must be evaluated against the probable loss that may occur in the event of an attack. If the loss due to an attack is not very high, simple security measures will do. But if the loss is perceived to be significant, then stringent security measures are required. For example, the loss due to copying of a public information resource is not as serious as stealing the same resource. In this case, security measure against theft is more important than one for preventing copying.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

1) Identify each of the following incidents as caused by hacking or by attack. Also identify if the incident is caused by a white-hat or black-hat hacker or by a passive or active attacker. In the case attacks, identify the type of attack. Also identify in each case, which one of the four information properties is compromised.

a) A confidential governmental report is published in the media.

b) An employee's pay record is tampered to give him more than his genuine salary.

c) A literary website page is replaced with pornographic material.

d) An instruction to a stockbroker for investment is delivered late to lose market advantage.

e) An instruction to include four new users to a sensitive computer system is delivered with two additional bogus users.

f) An online money transfer transaction gets executed twice.

g) A mutual fund receives instructions to change investment portfolio from a customer who actually never sent one.

2) Why do attackers reroute messages?

....

....

....

....

....

....

....

....

12.4 AAA SECURITY

AAA security is predominantly concerned with access to information resources residing on a server connected to a network or on a standalone computer. The access may come from a user or a computer program running on another computer or server. AAA stands for:

- Authentication
- Authorisation
- Auditing

Authentication is the process of ensuring that users who access the system are genuine. While genuine users are given access to the system, bogus users denied permission. Authorisation is the process by which the access to the information resources by a genuine user is restricted as per the access policy of the organisation. For example, a user may be permitted only to read the contents of a file but may not be allowed to modify the contents. Auditing is the process by which the use of the system is monitored. Audit raises alerts when a user makes an attempt to access resources in a way that s/he is not permitted to. Audit detects any unusual activity on the system and raises an alarm. Thus audit plays an important role in detecting attacks in the early stages.

AAA security may be implemented at many levels. The levels are:

- User
- Message
- Client machine
- Server
- Process
- Session

The most widely implemented AAA security is at the user level. Implementation at other levels is undertaken for specialised applications. For example, client machine authentication is implemented in stock market trading application. Server level authentication is necessary while accessing web sites that deal with financial transactions. In some applications, the authentication process needs to authenticate both the communicating entities, e.g. client machine and the server. In critical applications, the entire communication session is authenticated, i.e. every transaction or message exchange is authenticated. This is to ensure that no intruder comes on to the network after the initial authentication. In this unit, we confine our discussions only to the user level AAA security.

The most widely implemented user level AAA security is based on passwords. Password authentication involves three aspects:

- Password design
- Password policies
- Authentication process

Password must be designed to be complex but at the same time easy to remember for the individual user. Password should always be designed by the user and memorised by him/her. The designed password must be such that an intruder would not be able to guess or break the password. A complex password is designed by following five guidelines as given below:

1) Mix and use both upper and lower case letters
2) Use numerals as part of the password
3) Use special characters
4) Passwords must be at least eight characters long. The longer the password, it is more difficult to break it.
5) Password should not contain a string from the user name (ID).

The complexity of a password is usually rated as *weak, medium* and *strong*. For example, *boyandgirl* is a weak password. The password *boy12girl* is rated as medium and *Boy&12Girl* is a strong one. It may be noted that *Boy* and *Girl* are not usually parts of the user name. We have chosen strings that are not part of user name, mixed both upper and lower case letters, and included numerals and a special character in designing this password. It is also easy to remember this password. But it is difficult for a hacker or attacker to guess and break the password.

After having discussed the password design guidelines, we now look at the password policies. A typical set of policies is given below:

The password must have a minimum of 8 characters and a maximum of 32 characters.

1) The password must be changed every 30 days.
2) The system shall alert the user of the need to change the password four days prior to the expiry date. The alert shall appear every time a user logs on.
3) If the user fails to change the password by the due date, the system shall force the user to change the password the next time s/he attempts to log on. Unless the password is changed, the user shall not be allowed to log on.
4) If a user fails to enter his/her correct password within three attempts, his/her account shall be locked out. Only contacting the system administrator can then open the account.

The above policies constitute a typical set. The actual set may vary from organisation to organisation. For example, one organisation may enforce password change every 15 days whereas another may specify 45 days.

We now look at password authentication process. The main concentration of password authentication process is to hide the password from intruders. The passwords are stored in the system in an encrypted form. The encryption is often made irreversible so that the original password cannot be deciphered from the encrypted one. When the password is entered for authentication, it is encrypted and compared with the stored one. If a match occurs, the authentication becomes valid. When the password is transmitted on the network, it is encrypted, but this time using a reversible encryption process. The receiver system decrypts the received string to obtain the password. In this case, if the intruder catches the encrypted string, s/he would be able to crack the password or use the string to pose as the genuine user. To avoid this, sophisticated authentication algorithms are used and they produce a different encrypted string every time. Even if an intruder catches an encrypted string, it would be useless for the next log on.

We now consider the authorisation process. Once a user is authenticated, his/her access to the system is controlled by means of an **access control list (ACL).** ACL is the mechanism for limiting access to certain items of information based on user's identity. The ACL specifies the resources that a user can access and the permitted access for each resource. The resources include servers, directories, specific objects, folders and files. Access permissions range from full access at one extreme to total denial at the other extreme. A typical set of access permissions is given in the following:

- Full access
- Total denial
- Read only
- Read and Print
- Read and Copy
- Read & Write (Modify)
- Append
- Create
- Delete
- Change ownership
- Deny any

While read & print allows a user to get a printed copy of the resource, read & copy allows the user to save the resource as a file in the local user system. Append allows a user to add own material to the resource but prohibits him/her from modifying the existing contents. Change ownership permission is only given to system administrators and not to users. As opposed to total denial, deny any permits the system administrator to prohibit a particular access. For example, a user may be denied delete operation explicitly.

Auditing process tracks the activities of users and records specific events that appear to pose security threats. The events to be recorded and reported to the system administrator are specified in audit policies.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

3) What is the importance of auditing in AAA security?

4) How passwords are rated? Give one example of for each category of rating. Explain the reasons for the rating you have given for each password.

5) Distinguish between ‘total denial’ and ‘deny any’ access permissions.

6) Which access permission gives the user the power to modify the contents of an information resource?

....

....

....

....

....

....

....

....

12.5 FIREWALLS AND PROXY SERVERS

The literal meaning of firewall is a wall or a partition that is designed to inhibit or prevent the spread of fire. In the context of computer networks, firewall is a device that is designed to protect an organisation’s local network from external attacks while permitting the local network users to safely access the external resources. It is a wall that prevents *electronic fires* (external security attacks) from spreading to the local network. It acts as a gatekeeper for the local network. No one can enter, (i.e., no access can be made to) the local network without the permission of the gatekeeper.

Firewalls are designed to operate at packet level, i.e., at the network layer level or transport layer level in the OSI reference model. On the other hand, **proxy servers** are computer systems that perform firewall functions at the application level i.e., at the sessions layer or applications layer level. Hence, proxy servers are also called as **application level firewall**.

Both firewalls and a proxy servers act as single entry/exit point where security checks and audit inspections are imposed. This is illustrated in Fig. 12.3. They are part of the internal local network but can connect to the external resources safely. They are designed to be immune to attacks or any penetration from outside or even inside. They use highly secure operating systems. In this sense they are called **trusted systems**.

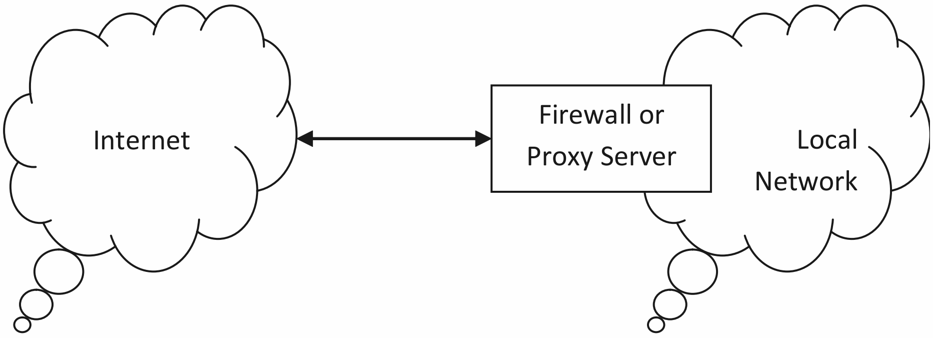


Fig. 12.3: Firewall or Proxy Server in the network

As mentioned earlier, firewalls are designed to operate at packet level. They use what is known as *packet filtering technology*. There are two versions of this technology:

- Static Packet Filtering technology
- Dynamic Packet Filtering technology

Static packet filtering (SPF) firewalls check the header information of incoming and outgoing service request packets and determine whether a service is to be established or not or whether a packet is to be forwarded or discarded. Forwarding or discarding decision is taken based on a set of security rules defined by the organisation. In general, SPF firewalls examine the following information for regulating the traffic:

- Destination address
- Source address
- Destination service port number
- Source service port umber
- Flag settings, where relevant

On the Internet, services are associated with certain port numbers. For example, web access takes place on port 80 and email receipt on port number 110. Over one thousand port numbers are reserved for specific services. Hence, by examining port numbers, a firewall can identify the service that is being requested. From the port number values and the corresponding security rules, the firewall can decide to block or allow a particular service.

Flag is a set of bits with each bit or a group of bits conveying a specific meaning. For example, one bit in the flag represents an acknowledgement for a request. If this bit is set, it means that the packet is a response to a request that has already originated under an established service. In this case, the firewall would pass the packet without any further checking. Once a bona fide service is established, there is a direct communication between the external and internal systems. In general, after the establishment of a service, SPF firewalls do not examine the subsequent packets of that service. This is to reduce the processing load and time overhead of examining the packets. Please note that every packet that is examined by the firewall contributes to delay in delivery of the packet.

Dynamic packet filtering (DPF) firewalls are more advanced than the SPF firewalls. In addition to establishing initial bona fide service connections, DPF firewalls continuously monitor the state of the service to ensure that no security lapses occur during the course of service communication. Obviously, the DPF firewalls add to delivery delay. To reduce this delay, they are designed to be more powerful than SPF firewalls. Consequently, they are more expensive.

Proxy servers are application level firewalls. They are also sometimes referred to as application level gateways. Unlike packet level firewalls, proxy servers do not allow any direct communication between internal and external systems. Instead they act like language interpreters between two speakers who speak different languages. The speakers carry on the conversation but never speak to each other directly. They only speak to the interpreter who translates what is communicated in one language into the other. Proxy servers function exactly this way. An interpreter tries to do a precise translation of what is being said unmindful of harmful or desirable effects the conversation may have. But, the proxy servers also examine the incoming and outgoing information to ensure that the security rules of the organisation are adhered to. In that sense, the proxy servers perform certain additional function when compared to an interpreter. A typical proxy server configuration is shown in Fig. 12.4.

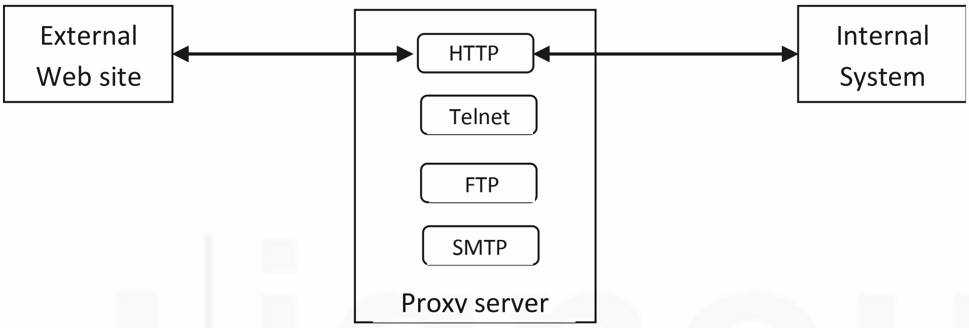


Fig. 12.4: Typical Proxy Server configuration

The proxy server shown in Fig. 12.4 supports four applications, viz. web access (HTTP), remote log in (Telnet), file transfer (FTP) and mail transfer (SMTP). A web access service connection is illustrated between an internal and an external system using hypertext transfer protocol (HTTP). Note that there is no direct connection between the internal system and the web site. The connection is via HTTP application in the proxy server. When the internal system wishes to access the web server, it sends a request to the proxy server. The proxy server, after ensuring compliance to the security rules formulates its own request to the web server. The web server replies to the proxy server and it does not even know the existence of an internal system. The proxy server examines the reply to ensure conformity to the security rules and then passes the reply on to the internal system. Thus all internal systems are protected from external systems.

Apart from the packet filtering firewalls and proxy servers, there is another type of firewall known as circuit type firewall. This is not used widely. A popular brand of DPF firewall uses the terminology S*tateful Inspection* to denote dynamic packet filtering. Finally, some packet level firewalls have other security features like anti-virus and anti-spam built in.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

7) What is the main function of a firewall?

8) How do firewalls and proxy servers perform their respective functions?

9) Distinguish between static and dynamic packet filtering.

10) What are the fundamental differences between packet level firewalls and proxy servers?

....

....

....

....

....

....

....

....

12.6 WEB SECURITY

These days, a large number of business houses, government agencies and many individuals have their own web sites. The number of web sites is growing day by day. Many business houses are currently upgrading their web sites to facilitate electronic commerce. At the same time, attacks on Internet and web sites are becoming serious. Almost all the four security properties of associated with information resources can be compromised by security attacks on the web sites. If an attacker gains read/write access to a web site, the integrity of information is lost. If an attacker is able to steal client data or business data from the web site, confidentiality is lost. Client data can be used to impersonate a client leading to the failure of authentication process. Hence, it has become imperative to design web sites that are secure and are immune to security threats.

Web security may be implemented at three levels:

- Network layer level
- Transport layer level, i.e. transmission control (TCP)
- Application layer level

At the network level, the emphasis is to make Internet Protocol (IP) secure. The secure Internet protocol is known as **IPSec**. When implemented at the transmission control level, the solution places emphasis on secure transport functions. Two most widely known solutions at this level are **Secure Socket Layer** (SSL) and **Transport Layer Security** (TLS). Application level security solutions are specific to each individual application. The solutions are implemented as part of the applications. Two well-known application level solutions are **Secure Electronic Transaction** (SET) and **e-mail security**. The SET is used widely in electronic commerce and financial applications like on-line banking, electronic stock trading and purchasing merchandise using credit cards. In this section, we will study about SSL and SET. E-mail security is dealt with later in another section. Equivalent of IPSEC in wireless networks is **Wi-Fi Protected Access** (WPA). Both IPSec and WPA are complex and their discussion is beyond the scope at BLIS level.

A company called Netscape Communications Corporation originated Secure Socket Layer (SSL) concept. The SSL solution has evolved over the years and is widely used presently. SSL is being adopted as Internet standard under the nomenclature Transport Layer Security (TLS). The discussions about SSL equally apply to TLS.

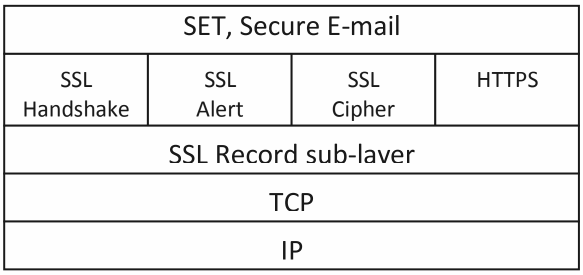


Fig. 12.5: SSL Protocol Stack

Socket is a connection end-point. SSL creates dynamically one connection end-point for each of the communicating entities. These sockets are then used for secure communication. SSL is designed to function as two sub-layers above the TCP layer in TCP/IP protocol stack. This is shown in Fig 12.5.

Immediately above the TCP layer is the SSL Record Protocol layer that provides basic security services to other higher level SSL and HTTPS functions. As you know, HTTP is the protocol used for web interaction. When HTTP is secured using SSL, the protocol is known as HTTPS, i.e. Secure HTTP. SSL handshake protocol is used for authenticating both the server and the client. SSL alert protocol generates SSL related security alerts. SSL cipher protocol deals with message encryption. We study encryption in a later section of this unit.

Secure Electronic Transaction (SET) is an application level service designed to protect credit card transactions on the Internet. SET specifies encryption and security protocols and standards to enable the users to utilise the existing credit card payment infrastructure on an open network like the Internet. The initiative for SET came from the major credit cad provider institutions like MasterCard and Visa. SET specifications support three functionalities:

- Provide a secure communication channel among the parties involved in a transaction, like the user, the merchant and the card issuer.
- Offer a trusted environment by issuing digital certificates.
- Ensure privacy among the transacting entities.

Digital certificates are explained in a later section of this unit.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

11) Why do we need web security?

12) What are the different levels at which web security is implemented?

13) What is a socket? How is it used by SSL?

14) How credit card transactions are secured on the Internet?

....

....

....

....

12.7 MALICIOUS SOFTWARE

A malicious computer program is one that harms the normal functioning of a computer. Malicious computer programs enter the computer systems via Internet, and removable mass storage devices such as floppies, pen drives and CDs. The rate of growth of malicious programs has been on the rise in the last ten years. It has also been observed that the cost of recovering from a malicious program attack is rising rather significantly. Downtime time for critical systems increase and services become unavailable for prolonged periods. In view of this, considerable attention is being given at present to prevent malicious program attacks. As you know, *prevention is better than cure*. Among the prevention measures are the use of malicious software removal tools and anti software like anti-virus.

A variety of malicious computer programs exist, virus being one of them. Malicious programs tend to exploit the vulnerabilities in system software, in utility programs such as editors and compilers and in application software. They fall under two broad categories:

- Programs that can exist on their own and run on their own
- Pieces of code that need a host program to run

Under the first category are programs called **bacteria** and **worms.** Under the second category are programs called as **trapdoors, logic bombs, Trojan horses** and **viruses.** Of these, bacteria, worms and viruses are capable of replicating themselves. Of the six different malicious programs mentioned above, we briefly discuss five of them in this section. Viruses are discussed in greater detail in the next section.

Bacteria are programs that generally cause **denial of service** attack. They do not affect the existing files or programs. Since they are independent stand alone programs, they replicate themselves and start growing exponentially, viz. one copy to two, two to four, four to eight and so on. Eventually, these replicated programs (say, 100s in numbers) take up all the resources of a computer system, viz. processor capacity, memory and disk space. The result is that no resources are available for normal programs and the regular services come to standstill.

Worms are malicious programs that spread from system to system via network connections. To travel from one system to another, they use one of the following methods:

- A copy of the worm program is sent as email from one system to another. It then executes itself on the new system. Then it further propagates to other systems in a similar fashion.
- A worm may execute itself on another computer via the remote execution feature of the operating system. For this purpose, the worm looks up tables to see which other computers are accessible and which of those allow remote execution. Then it selects a suitable system and moves to that system.
- A worm may log on to a remote system as a user and then use commands to copy itself on to the remote system. Then it travels to another system in a similar manner. This uses **Telnet** feature of the network operating system.
- A worm may also use file transfer facility (**FTP**) to copy its source code from one system to another.

Once active within a system, a worm can perform disruptive or destructive action within the system like bacteria or virus.

Often, programmers build in special entry points in programs at the time of development. Such entry points help them debug and diagnose programs. These entry points usually skip many standard procedures to enter the program. Some of the steps skipped may be security steps. Although many special entry points are closed before releasing the program in the field, some points are left open for field level diagnosis and debugging. In the context of information security, such open entry points are called **trap doors.** A person who is knowledgeable about the existence of such entry points can exploit the feature to enter the program bypassing security provisions and wreck the operation of the system. The existence of trap doors is verified by performing what is known as **penetration testing.**

Logic bombs are codes that get activated when certain conditions are met. Such conditions include particular date and time, the presence or absence of certain files, running of a particular application or the appearance or disappearance of particular user ID. A well-known example of logic bomb is the one that explodes on Friday, the thirteenth. If the date 13 happens to be a Friday, this bomb explodes. Another interesting example is that of a contractor who built in a logic bomb to explode if his payment was not released by a particular date.

A **Trojan horse** is a seemingly useful program that contains hidden code, which when invoked performs unwanted harmful functions. Attackers who are otherwise unable to get access to systems that they want to attack use Trojan horses. They would supply seemingly beneficial program (often free of cost) to a bona fide user. When the user executes the program, the malicious code becomes active. To avoid suspicion, the code actually causes harm after sometime (say, a day or two) so that the user does not suspect the supplier of the program.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

15) What are the commonly known malicious programs? Which of these need support to exist and execute?

16) How do bacteria cause denial of service attack?

17) What are the different network utilities through which worms spread?

18) Discuss the useful and harmful effects of trap doors.

....

....

....

....

....

....

....

....

....

12. 8 VIRUSES

A computer virus is a malicious program code designed to cause harm to the normal functioning of computer systems. A virus is not an independent program. It needs a vehicle to ride and cause damage. The vehicle could be an executable program file or a data file like word document or a picture file. Much like the biological virus, a computer virus also spreads from program to program, data file to data file and from one system to another.

How does a virus get into a computer in the first place? People obtain programs, utilities and data from a variety of sources. These are imported into the system in a variety of ways. They are brought in through floppy disks, pen drives, CDs, downloaded from Internet or copied from another system using some network. It is this process of bringing in external program and data that launches a virus into a system. Obviously, day-to-day functioning calls for external interaction that is unavoidable. The only thing that one can do is to guard one's system from being infected by a virus. One needs to ensure that only virus-free programs and data are brought into the system. This is done by using what are known as **anti-virus** programs that keep a continuous watch on what enters the system. Anti-virus programs scan all incoming code and data to ensure virus-free import. They also scan every removable storage device whenever they are mounted on the system. Viruses, if found are neutralised or healed. If this is not possible, the import process is rejected. The user is given an appropriate message when scanning is done. The results of scanning are also displayed to the user. Despite all this, virus attacks take place. This is because the writers of virus code continuously innovate new techniques to befool the existing anti-virus programs. They produce new type of virus that is not previously known to the anti-virus programs. Such viruses enter the system. Then the anti-virus program is updated to deal with the new virus. This is a continuous process and that is the reason why it is essential to keep anti-virus programs updated on a daily basis and scan the system periodically. However, if one is using the system for surfing only trusted sites, downloading data or programs from certified sources only and using a mail system that has virus scanning built in, then the system can function safely without anti-virus programs. But the usage has to be restrictive and care should be exercised in choosing sites. Usually, the operating systems like Windows provide their own software firewalls that offer a good amount of protection.

There are two broad categories of viruses:

- Those attach themselves to executable program files
- Those attach themselves to data files.

The latter category of viruses is also known as **macro viruses**. Macro is a piece of code that is placed in data files. Macros execute themselves when the data files are open. If a virus is placed as a macro in a data file, it gets executed as soon as the data file is open. Virus then copies itself to other data files and thus spreads. There are many viruses that belong to the first category. The important ones are as follows:

- Parasitic virus
- Memory resident virus
- Boot sector virus
- Stealth virus
- Polymorphic virus

Parasitic viruses are the most common. They attach themselves to executable files. When an infected program file is executed, the virus becomes active and replicates itself by infecting other executable files. Memory resident viruses lodge themselves in main memory and infect every program that executes. Boot sector virus find place in boot record of the system and spreads when the system is booted. A stealth virus is explicitly designed to hide itself from anti-virus programs. Polymorphic viruses mutate (change) with every infection making it difficult to locate replicated copies even if the original has been detected.

During its lifetime a typical virus goes through the following four stages:

- Dormant phase
- Propagation phase
- Triggering phase
- Execution phase

When a virus enters a computer system, it usually remains idle (dormant). It then replicates itself (propagates) whenever the infected file is executed or opened. A virus gets activated (triggered) to perform its intended function based on some event, much like the logic bomb. Finally, the virus performs its destructive function. It is not essential that every virus goes through all the above four stages. For example, a virus may execute itself as soon as it enters the system.

Antivirus programs usually go through three phases of operation:

- Detection
- Identification
- Removal

When an infection takes place, the anti-virus program scans the system files to detect the presence of the virus. Once detected, the program identifies the specific type of virus and determines what is called the **signature** of the virus. The system is then scanned for the occurrence of the signature and the virus is removed wherever it occurs. Since polymorphic viruses change the signature at every replication, they are difficult to remove at one go. Multiple scanning may be required.

Antivirus technology has advanced significantly over the years. Starting with simple scanners, they now incorporate advanced features like heuristic scanners etc. The antivirus programs are considered to have evolved over four generations. The current most sophisticated programs are called fourth generation anti virus programs.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

19) How does a virus enter a computer system?

20) What are the functions of anti-virus programs?

21) What is a macro? What are macro viruses? Why are they most common?

22) What are the stages through which a virus moves during its lifetime?

....

....

....

....

12.9 SPYWARE, SPAM, PHISHING AND COOKIES

There is yet another class of malicious programs that do not cause harm directly to the functioning of the computer system but are of nuisance value or exploit personal data in other ways. Spyware is a malicious program that launches itself in the start up menu of a computer and gets executed every time the system is booted. This program tracks the activity of the users of the system and passes the collected information over the Internet to its parent site for analysis. In this sense, the program behaves like a spy. The rouge parent site may misuse the information collected to cause harm to the user. Some spyware programs can actually interfere with the functioning of the computer. They may change the Internet browser settings resulting in slow connection speeds or redirect access to non-genuine websites.

Spam is unsolicited advertising via e-mail. A spammer sends and advertisement to tens of thousands of e-mail addresses and mailing lists. A recent study shows that 95% of Internet traffic comprises spam e-mails. It is not completely possible to stop spam. Spam guards can be used identify spam mails and ensure that they do not carry virus or other malicious programs. The source address of the spammer can be input to the spam guard, which then will block all the mails from that source. However spammer will generate new source addresses and keep sending spam. As a result spam adds to the nuisance value.

Phishing is a type of deception to steal valuable personal data, such as credit card number, bank account number, e-mail ID, passwords etc. Attackers pose like genuine sites and seek secret information. For example, an attacker site may pose as your bank and ask for your bank account details citing some convincing reason. If provided, the attackers would then use the information to the customer's disadvantage. Phishing filters are used to guard against stealing of valuable information. Phishing filters monitor the web sites accessed to ensure that sites are genuine.

Cookies are packets of text sent by a web server to a web client. The client returns them unchanged during subsequent access to the web server. The web server uses cookies to keep track of the earlier accesses. Cookies are not programs and are usually harmless. They are neither viruses nor spyware. However, it is better to delete the cookies once a web access is closed. This can be done by appropriate tool available as part of the Internet browser.

There are two types of cookies:

- Temporary
- Persistent

Temporary cookies are deleted by the web site when the site access is closed. Persistent cookies stay in the system for a long duration. Some of them may have a specific lifetime (say, 15 days) after which they are automatically deleted. Others may stay permanently unless specifically deleted by the user. The persistent cookies are used to arrange the content of the accessed web site according to the user preference. Once the user indicates his/her preference, the required information is stored in the persistent cookie and is used by the web site to arrange the contents when it is opened next time. Cookies are sometimes called as web cookies, tracking cookies or HTTP cookies.

12.10 ENCRYPTION

Encryption is the art of hiding the meaning contained in a message. Decryption is the reverse process of encryption to extract the hidden meaning. The general field of study of encryption and decryption and related aspects is known as **cryptography**. Cryptography is an age-old technique for communicating secretly between two parties. Military personnel and persons in love have been using this technique for ages. There are two fundamental ways of encryption:

- **Substitution** where each character or a group of characters in a message is substituted with another.
- **Transposition** where the characters are jumbled but not changed.

When each character is substituted by another, it is called **monoalphabetic** substitution. When a group of characters is substituted by another group, it is called **polyalphabetic** substitution. We illustrate the substitution and the transposition techniques by taking a simple example. Consider the word 'COMPUTER'. A substitution encryption for this word is FRPSXWHU. Here, each letter has been substituted by the letter that occurs two letters after the current one in English alphabet, i.e. the third letter. The encrypted text, i.e. FRPSXWHU in this case, is often referred to as **cipher text**. Julius Cesar, the Roman emperor, first used monoalphabetic substitution encryption. Cesar's cipher used a shift of three letters, i.e. A is substituted by D, B by E, C by F and so on. A transposed encryption for the word COMPUTER is 'UTERCOMP', where the characters are shifted in a round robin fashion by four positions. The characters are the same as in the original word, but their positions have been changed, i.e. they are transposed.

Obviously, the simple substitution and transposition ciphers as illustrated above are easy to be broken by an attacker. By simple trial one can easily determine the shift value in these cases. The substitution and transposition can be made complex by using keys. For example, let us use the numeric key 12340123 to transpose the word COMPUTER. The key is as long as the text to be transposed. Each digit in the numeric key specifies the number of positions by which the corresponding character must be shifted in a round robin fashion. The transposition operation leads to the cipher ECROUMTP. The letters are jumbled and their order of occurrence is changed. But the letters themselves have not changed. Now it is more difficult to break the cipher. If only the sender and the receiver know the numeric key, then secret information can be exchanged between the two. The attacker will not be able to break the coded message unless he/she can figure out the secret key used for encryption and decryption.

In this example, the key length is the same as the text length, i.e. 8 characters. In practice, one cannot have very long keys. Hence, the message is broken into small blocks and the blocks are then encrypted by using the same key. Sophisticated encryption algorithms have been designed by using these two fundamental techniques and keys. Among them are Data Encryption Standard (DES), Triple DES (3DES) and International Data Encryption Algorithm (IDEA). A general cryptography scheme is shown in Fig. 12.6. User message is encrypted using the encryption key and the encryption algorithm to produce cipher text. The cipher text is decrypted using the decryption key and decryption algorithm to obtain the original user text.

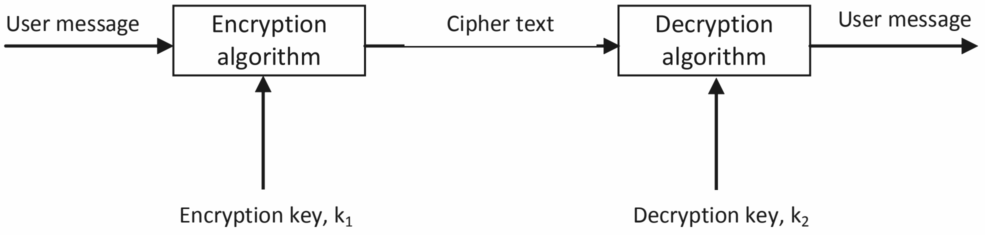


Fig. 12.6: Cryptography Scheme

We can think of encryption as the process of putting the user message in safe box and locking it with the key. The safe box is then transported to the receiver. Because the box is locked (cipher text), it is safe during the passage. Decryption is the process of opening the safe box with the key and retrieving the message.

Depending on the properties of keys used, two cryptographic schemes are in use:

- Symmetric cryptography where key k_1 = key k_2
- Asymmetric cryptography where key $k_1 \neq$ key k_2 but the two form a special pair.

Symmetric cryptography is like the one where a safe box has two identical keys. The same key can be used for both encryption and decryption. One key each is available with both the communicating parties. Both of them keep the key secret. Whenever one wants to send a secret message, he/she encrypts the message using the key. The receiver decrypts the message with the same copy of the key. The main problem with symmetric cryptography lies in the exchange of key between the communicating parties. Consider the case where the two parties are in different countries. The key should not be exchanged over the network because an attacker may copy the key. In high-risk situation, voice exchange over telephone line may not also be safe. The telephone line may be tapped. Postal communication may have the same constraint. Such serious security concerns apply only to military and other sensitive applications. Exchange of keys over telephone would work for most situations. However, many key exchange algorithms have been designed to ensure safe key exchange over the network itself. Because the keys are private to the two communicating parties, this scheme is also called **private key cryptography**.

Asymmetric cryptography can be likened to a safe box with two special keys that form a pair. The keys are such that if the safe is locked with one key it can be opened only with the other key. Both keys can be used to lock and open the safe except that the locking and the opening keys have to be different but from the same pair. Consider a key pair (k_1, k_2). If the safe is locked with k_1, it can only be opened with k_2 and no other key, not even k_1. Similarly, if the safe is locked with k_2, it can only be opened with k_1 and no other key, not even k_2. That is the special property of the keys. Every user on the network has this key pair generated. The user keeps key k_1 secret and makes k_2 public. If anyone wants to send a secret message, then the message can be encrypted using k_2 and sent to the user. The user decrypts the message using the key k_1. Since no key other than k_1 can decrypt the message, only the recipient with key k_1 can decrypt the message, which means the message has been passed on to the receiver secretly. Since one key is made public, this asymmetric scheme is also called **public key cryptography**.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

23) The string given below is encrypted using mono alphabetic substitution. Spaces between words have been deleted and the capitalisation has been hidden to make deciphering difficult. The encrypted string is:

ymnxnxgqnxhtzwxjtknlstz

Determine the original unencrypted message including spaces and capital letters.

24) Why is the symmetric cryptography called private key cryptography?

25) What is public in asymmetric cryptography?

26) In asymmetric cryptography, a message is encrypted by key k_2. How can this be decrypted?

....

....

....

....

12.11 DIGITAL SIGNATURE

Why do we use signatures? A signature authenticates the contents of a document. A signed document is an authentic source of information. Digital signatures are a means of authenticating electronic documents. We saw in the previous section as to how the asymmetric (public key) cryptography can be used to convey secret information. In this section, we will show that the scheme has inherent capability to authenticate documents as well. We will then describe how digital signatures are formed.

Consider an asymmetric cryptography key pair of some user *A*, (Ak_1, Ak_2). The key Ak_1 is kept secret and the key Ak_2 is made public. Anyone who wants to convey secret information to user *A* encrypts the message with Ak_2 and transmits the same to user *A*. This message can be decrypted only by using key Ak_1 that is kept secret by the user *A*. (Recall the key properties of asymmetric scheme). Even if an intruder intercepts the message while it is being communicated, he/she cannot read the contents, as he/she does not know the secret key Ak_1. Thus, information is conveyed safely and secretly to user A.

Now, consider the case when user A wants to communicate an authenticated document to someone on the network. He/she then encrypts the document using key Ak_1 and transmits the same on the network. Now, by key pair properties, this document can be decrypted only by key Ak_2. The key Ak_2 is publicly known to everybody and hence the recipient can decrypt the message. There are two things that you must note now. One, this message is not secret and an intruder who intercepts the message can read the contents using key Ak_2 that is public. Two, the fact that this message can be decrypted only by Ak_2 implies that it has been encrypted by Ak_1 that is known only to user *A*. Hence the message has been originated by user *A*. In other words, this is an authentic message from user *A*.

In the examples considered above, either a message secret or authenticated but not both. Often we need to send an authenticated message secretly. How is this possible? In fact, the asymmetric scheme is capable of supporting both secrecy and authentication simultaneously. This is achieved by carrying out two levels of encryption and decryption. We shall now see how this happens? In addition to user *A*, let us consider user *B* with asymmetric key pair (Bk_1, Bk_2). Let us suppose that user *A* wants to send an authenticated message secretly to user *B*. The two-level encryption and decryption process is depicted in Fig. 12.7. The user A performs two levels of encryption in the following order:

- First, encrypt the original message *M* with Ak_1. Let the encrypted message be *X*.
- Second, encrypt *X* with Bk_2 to get *Y* and transmit the same to user *B*.

The first encryption ensures authentication and the second secrecy. The original message goes through the following transformation: *M* ’! *X* ’! *Y*. At the receiving end user *B* performs two levels of decryption in the following order:

- First, decrypt *Y* with Bk_1 to get *X*.
- Second, decrypt *X* with Ak_2 to get the original message *M*.

The received message *Y* goes through the following transformation: *Y* ’! *X* ’! *M*.

Because the transmitted message *Y* is encrypted with the public key of user *B*, only user *B* can open it with his secret key. Hence secrecy is achieved. After first decryption, *B* gets message *X*. Since this is encrypted with the secret key of *A* this can be opened only using the public key of *A*. The fact that the message is decrypted with AK_2 implies that the sender is *A*, i.e. the message is authentic from user *A*.

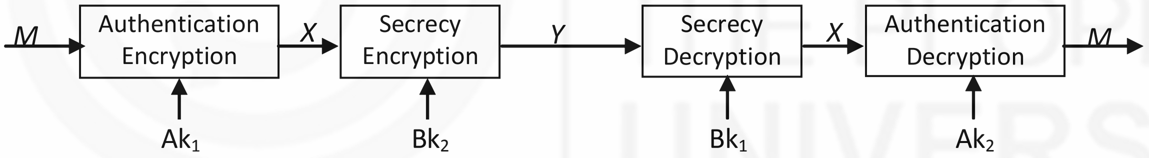


Fig. 12.7: Two-level asymmetric cryptography

Thus, the public key cryptography scheme is capable of transporting authenticated documents secretly. In other words, both authentication and secrecy are achievable in asymmetric cryptography. The problem of key exchange also does not exist. There is no secret key that needs to be exchanged as in the case of symmetric scheme.

Encrypting a full document is expensive in terms of processing time and communication overheads. In practice, there are numerous instances where the contents of a document are not secret but they need to be authentic. Examples include government circulars, court orders and department notices. All of them are public documents. However, their content must be authentic. This is where digital signature comes handy. Use of digital signatures reduces the encryption overheads. Here, the main contents of the document are in plain text and the signature is encrypted using asymmetric scheme.

Consider the example of a government circular to be issued on the network. The official who issues the circular appends his/her name encrypted using his/her own secret key k_1. Any recipient can verify the signature by decrypting it with the public key k_2 of the official concerned. This procedure does not ensure that the contents are safe. An intruder may intercept the circular, change the plain text contents but retain the encrypted signature. To safeguard from such an eventuality, the official concerned generates what is called a

message digest and encrypts the same along with his/her name. Message digest is short string of characters or numerals that unique to the message. Even if one letter is changed in the message, the digest value will change. Now recipient can generate the same message digest and match it with what is sent by the official in encrypted form. If a match occurs, the message is safe. Otherwise, the message has been tampered. The procedure for generating the message digest is public. Anybody can generate the digest and verify with what is sent.

Digital signatures can be used in private communication too. In other words, every user of the network may have one's own digital signature and send authenticated documents across. There is one problem here. What happens if user *X* poses as user *Y*? To safeguard from this problem, **digital certificates** are used. These certificates are issued by well known certified authorities (CA). These certificates are digitally signed by the CA and confirm that the sender is the genuine person and not a fake. Digital certificates are required only in sensitive applications where security threats are high.

Self-Check Exercise

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

27) User *A* encrypts a message with key Bk_2 and sends the same on the net. Who can read this message and how? What does this operation ensure?

28) User A encrypts a message with key Ak1 and sends the same on the net. Who can read this message and how? What does this operation ensure?

....

....

....

....

12.12 E-MAIL SECURITY

Electronic mail is one of the most widely used applications on the network. The power and speed of electronic mail has prompted users and businesses to seek ways of sending confidential information over the network. Thus was born the need for secure electronic mail. There are two most popular schemes used for sending secure e-mail:

- Pretty Good Privacy (PGP)
- Secure Multi-purpose Internet Mail Extension (S/MIME)

PGP was largely developed by a single person, Phil Zimmermann in 1991, who made it available freely on the Internet. It is open source software that can be downloaded from the net. PGP provides the following features:

- Digital signature
- Message encryption for secrecy
- Compression
- Compatibility to e-mail systems
- Segmentation.

We have already studied digital signature and encryption. PGP offers facility to compress (zip) messages to economise on storage and transmission. Most e-mail systems put restriction on the number and size of the files that can accompany e-mail. Segmentation and reassembly at the receiving end are used to handle large messages that exceed the size permitted by e-mail systems.

Most of the electronic mail systems use Simple Mail Transfer protocol (SMTP) for transporting mail from one system to another. SMTP is an Internet approved protocol. There are some limitations with SMTP. To overcome these limitations, Multipurpose Internet Mail Extension (MIME) was developed. Security features have been added to MIME and S/MIME has been developed. S/MIME also has all the features of PGP and is more sophisticated than PGP in some ways. Prior to S/MIME, Internet task force developed e-mail security software known as **Privacy Enhanced Mail** (PEM). PGP and PEM are very similar. But, PEM never became popular. S/MIME was developed as a sophisticated replacement for PEM.

12.13 SUMMARY

This unit has dealt with information security issues in a networked environment. Information, being a valuable commodity, needs to be secured from hackers and attackers. If the information security is violated, the availability, confidentiality, integrity and authenticity of information may be compromised.

Access to information resources is secured by authentication, authorisation and auditing (AAA security) at different levels. User authentication is via passwords. They need to be designed to be complex but easy to remember. The complexity must be strong enough so that attackers cannot break the passwords easily. An organisation must enumerate password policies to ensure better security. The networked environment comprises LAN and Internet. Firewalls and proxy servers protect the organisation's LAN from attacks originating from the Internet.

World wide web (WWW) is used extensively in a networked environment. A number of applications involving financial transactions are being supported on the web increasingly. Such applications require security. Secure socket layer (SSL) and transport layer security (TLS) are two standards that support web security. Secure electronic transaction (SET) is an application-level standard that ensures secure credit card transactions on open networks like Internet. Wireless networks are becoming popular these days. Security in wireless networks is an important issue. Unsecured wireless networks are vulnerable for misuse. There are reported cases of unsecured wireless networks being hacked and used for terror activities. It is essential to install Wi-Fi Protected Access (WPA) feature with wireless networks.

In a networked environment, malicious software like virus, bacteria and worms enter user systems and cause damage to information resources. Some of the malicious programs exist on their own and others ride on executable user files or data files. The ones that ride on data files use macros and are called macro viruses. Macro viruses are more common than viruses that attach themselves to executable files. There is another class of malicious programs that do not directly damage information resources but are of nuisance value or attack indirectly by collective sensitive and confidential information. They include spyware, spam mails, Phishing attacks and cookies.

Encryption is the most important security tool used extensively in all security applications. Substitution and transposition are two fundamental encryption techniques. They are made very powerful by using keys. Encryption can be implemented in symmetric or asymmetric form. Symmetric cryptography is also known as private key cryptography and the asymmetric scheme as public key cryptography. While symmetric scheme ensures secrecy, asymmetric scheme can ensure both secrecy and authenticity. Hence, it is used more extensively on the Internet. Digital signatures affixed using asymmetric cryptography principle allow public documents to be sent in plain text while ensuring integrity of information that is authenticated. To facilitate secure digital signature use in private domain, trusted certificate authorities (CA) issue digital certificates.

Electronic mail is the most extensively used service on the network. In sensitive applications, information needs to be exchanged securely via e-mail. For this purpose, pretty good privacy (PGP) and secure multipurpose Internet mail extension (S/MIME) protocols have been developed. Both of them support encryption, digital signature and compression. They are compatible with most mail systems.

12.14 ANSWERS TO SELF-CHECK EXERCISES

1) The answers to different parts of this question are:

   a) White hat hacker. Confidentiality.

   b) Black hat hacker. Integrity.

   c) White hat hacker. Integrity and authenticity.

   d) Attacker. Delayed delivery. Availability.

   e) Attacker. Modify. Integrity and authenticity.

   f) Attacker. Replay. Authenticity.

   g) Attacker. Masquerade. Authenticity.

2) Attackers reroute the messages to analyse and cause an active attack. Usually, rerouting is done via the attacker's own server, where the message is analysed in detail. An attack strategy is then worked out. The attack is then carried out as per the strategy. The strategy may be to delay, modify or substitute the message. The onward transmission of the message may be blocked resulting in apparent breakdown of communication between source and destination.

3) Auditing brings out potential attacks before they happen. For example, attacker trying to crack a user's password may be caught if he does not succeed in, say, 3 attempts. Audit generates alerts about all unusual activities. These alerts can help prevent security attacks. Audit also alerts when an employee attempts to carry out activities that he/she is not expected to perform. These are raised as security alerts.

4) Passwords are rated as: weak, medium and strong. Examples: IGNOUBLIS is a weak password. This is weak because it uses two well known strings without numerals or special characters. I1g2n3o4U5 is a medium password. This is medium because it uses both letters and numerals. I$g5N6o&U* is a strong password. This is strong because it uses both upper case and lower case letters, numerals and special characters. It is easy to remember as the upper and lower cases alternate and special characters and numerals occur in a sequence on the keyboard.

5) 'Total denial' to an information resource means that the concerned user cannot access the resource at all. 'Deny any' means a particular access. For example, 'deny copy' prohibits a user from copying the contents of a resource but may allow him to read the information.

6) ‘Write” permission would allow the user to modify the contents.

7) The main function of a firewall is to protect the internal systems of an organisation from external security attacks.

8) Firewalls and proxy servers act as a single point of entry to and exit from the organisation for information flow. They are part of the internal network and provide safe access to external resources. Refer to Fig. 12.3.

9) Static filtering examines the packets whenever a new connection is requested. In general, after the establishment of a service, SPF firewalls do not examine the subsequent packets of that service. Dynamic packet filtering (DPF) firewalls are more advanced than the SPF firewalls. In addition to establishing initial bona fide service connections, DPF firewalls continuously monitor the state of the service to ensure that no security lapses occur during the course of service communication.

10) Packet level firewalls permit direct connection between the internal user and the external resources. Proxy servers do not allow internal users to interact with external resources. Firewalls operate at packet level whereas the proxy servers at the application level.

11) These days, a large number of business houses, government agencies and many individuals have their own web sites. The number of web sites is growing day by day. Many business houses are currently upgrading their web sites to facilitate electronic commerce. At the same time, attacks on Internet and web sites are becoming serious. Almost all the four security properties of associated with information resources can be compromised by security attacks on the web sites. If an attacker gains read/write access to a web site, the integrity of information is lost. If an attacker is able to steal client data or business data from the web site, confidentiality is lost. Client data can be used to impersonate a client leading to the failure of authentication process. Hence, it has become imperative to design web sites that are secure and are immune to security threats.

12) Web security may be implemented at three levels:

- Network layer level
- Transport layer level
- Application layer level

13) Socket is a connection end-point. SSL creates dynamically one connection end-point for each of the communicating entities. These sockets are then used for secure communication.

14) Credit card transactions on the Internet are secured by the use of an application level service called Secure Electronic Transaction (SET) service. SET is designed to protect credit card transactions on the Internet. SET specifies encryption and security protocols and standards to enable the users to utilise the existing credit card payment infrastructure on an open network like the Internet. SET specifications support three functionalities:

- Provide a secure communication channel among the parties involved in a transaction, like the user, the merchant and the card issuer.
- Offer a trusted environment by issuing digital certificates.
- Ensure privacy among the transacting entities.

15) There are six commonly known malicious programs: bacteria, worms, trapdoors, logic bombs, Trojan horses and viruses. Bacteria and worms are self-supporting. Trapdoors, logic bombs, Trojan horses and viruses need some executable or data file to attach themselves.

16) Since bacteria are independent stand alone programs, they replicate themselves and start growing exponentially, viz. one copy to two, two to four, four to eight and so on. Eventually, these replicated programs (say, 100s in numbers) take up all the resources of a computer system, viz. processor capacity, memory and disk space. The result is that no resources are available for normal programs and the regular services come to standstill. Hence genuine users are denied of service, which is a denial of service attack.

17) There are four network utilities through which worms spread to other systems: e-mail, remote login (Telnet), remote execution and file transfer (FTP) facility.

18) Trap doors are built-in special entry points in programs. Such entry points help them debug and diagnose programs. This is useful. A person who is knowledgeable about the existence of such entry points can exploit the feature to enter the program bypassing security provisions and wreck the operation of the system. This is harmful.

19) People obtain programs, utilities and data from a variety of sources. These are imported into the system in a variety of ways. They are brought in through floppy disks, pen drives, CDs, downloaded from Internet or copied from another system using some network. It is this process of bringing in external program and data that launches a virus into a system.

20) Anti-virus programs keep a continuous watch on what enters the system. They scan all incoming code and data to ensure virus-free import. They also scan every removable storage device whenever they are mounted on the system. Viruses, if found are neutralised or healed. If this is not possible, the import process is rejected. The user is given an appropriate message when scanning is done. The results of scanning are also displayed to the user.

21) Macro is a piece of code that is placed in data files. Macros execute themselves when the data files are open. Viruses that attach themselves to a data file as a macro are called macro viruses. Macro viruses are more common because people import a lot more data files into the system than program files. Macro viruses come in as part of data files.

22) During its lifetime a typical virus goes through the following four stages:

- Dormant phase
- Propagation phase
- Triggering phase
- Execution phase

23) The decrypted string is: **This is BLIS course of IGNOU**.

24) In symmetric cryptography, the encryption and decryption keys are identical and are private and secret to thee sender and receiver. Hence, the nomenclature.

25) In asymmetric cryptography, the key k_2 of the key pair (k_1, k_2) is made public.

26) A message encrypted by key k_2 can be decrypted by using key k_1.

27) Since the message is encrypted using Bk$_2$, i.e. the public key of user *B* it can only be decrypted using the secret key of user *B*, i.e. Bk$_1$. Only user *B* can read this message. Hence this process ensures secret message being passed on to user *B*.

28) Since the message is encrypted using Ak$_1$, i.e. the secret key of user *A* it can only be decrypted using the public key of user *A*, i.e. Ak$_2$. The key Ak$_2$ is public and is known to everyone on the network. Hence anybody can read this message. This process ensures that the message is an authentic one from user *A*.

12.15 KEYWORDS

Access control list : A list used to control the access to information resources such as files by any user

Attackers : Those who cause harm to information resources like modify contents, deny access, steal data, replace the resource etc.

Authenticity : The guarantee that information is actually from the source it is claimed to be.

Availability : Availability of information resource to genuine users.

Cipher Text : The text obtained after encryption.

Confidentiality : Privacy of information.

Cookie : A piece of text message sent by a web site to user machine to keep track of browsing history by the user.

Cryptography : The branch of study covering encryption, decryption and other related aspects.

Decryption : The process of recovering hidden information from an encrypted text. It is the reverse process of encryption.

Digital certificate : A certificate issued by Certifying Authority assuring that the user of a public key is genuine.

Digital signature : An electronic signature to a document that ensures integrity of the contents and affirms that the source is authentic.

Encryption : The process hiding useful information from those oother than the intended recipient.

Firewall : A hardware device or software on a computer, which protects the internal network of an organisation from external attackers.

Hackers : Persons who violate security provisions to expose loopholes or for personal gains.

Integrity : A property of information resource that assures the contents are not corrupted.

Intruders : Those who enter the network with mala fide intentions.

Macro : A piece executable code attached to a data document like word document.

Macro virus : A virus that attaches itself to a data resource.

Malicious programs : Programs that enter user computer systems and cause harm to information resources or spy or cause nuisance.

Masquerade : Pose as a genuine user by deceitful means.

Pretty Good Privacy : PGP. It is a security service for e-mail.

Phishing : The process of collecting sensitive and confidential user information by deceitful means.

Proxy servers : Systems that hide internal users from external network.

S/MIME : Secure Multipurpose Internet mail Extension used for sending secure e-mail.

SET : Secure Electronic Transaction. Used for securing credit card transactions on open network like Internet.

Spam : An unsolicited e-mail often containing advertisements.

Spyware : A malicious program that spies on user activity.

12.16 REFERENCES AND FURTHER READING

SSL : Secure Socket Layer. SSL is a suite of services, standards and protocols for securing access to web sites.

Brenton, Chris. *Mastering Network Security*. New Delhi: BPB Publications, 1999. Print

Stallings, William. *Network Security Essentials*. Delhi: Pearson Education, 2003. Print

Tanenbaum, A. S. *Computer Networks*. 4th Ed. New Delhi: Prentice Hall of India, 2002. Print

Viswanathan, Thiagarajan. *Telecommunications Switching Systems and Networks*. New Delhi: Prentice Hall of India, 2008. Print