

BLOCK 2

COMMUNICATION TECHNOLOGY

**THE PEOPLE'S
UNIVERSITY**

UNIT 4 FUNDAMENTALS OF COMMUNICATION TECHNOLOGY

Structure

- 4.0 Objectives
- 4.1 Introduction
- 4.2 Communication Systems
 - 4.2.1 Electrical Communication
 - 4.2.2 Optical Communication
 - 4.2.3 Radio Communication
 - 4.2.4 Satellite Communication
- 4.3 Signal, Noise and Attenuation
 - 4.3.1 Signal
 - 4.3.2 Pure Waves
 - 4.3.3 Complex Signals
 - 4.3.4 Noise
 - 4.3.5 Attenuation
- 4.4 Digital Communication
- 4.5 Communication Functions
- 4.6 Summary
- 4.7 Answers to Self Check Exercises
- 4.8 Keywords
- 4.9 References and Further Reading

4.0 OBJECTIVES

The developments in communication technology have ushered in a new era for accessibility of information. It is, thus necessary that you should be conversant with its basics. In this Unit we will be discussing the basic aspects of communication technology.

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

- information transfer process;
- different types of communication systems;
- signal and noise in communication;
- importance of signal to noise power ratio;
- what are pure waves and complex waves;
- importance of digital communication;
- different functions of a communication system;
- coding for different purposes;

- what is modulation and why it is required;
- different modulation schemes;
- what is multiplexing; and
- why synchronisation and its techniques.

4.1 INTRODUCTION

The primary purpose of communication systems is information transportation, i.e., to transfer information from one place to another over a distance. Information that is to be transferred may originate in different forms. Human speech, a still picture, an action scene, a discourse, a message and a document are examples of original forms of information. Information in its original form can rarely be transported over a communication system. It needs to be represented in the form of **signals** that can be carried over communication systems. In today's technology, communication systems are capable of carrying signals in three energy forms: **electrical, optical or electromagnetic**. Hence, if we wish to transfer across any piece of information over a communication system, then we need to represent it as signals in one of these three energy forms. Communication technology deals with representation of information in the form of signals, transportation of these signals *efficiently* and *reliably* to the specified destination and convert the signals back to the original form of information at the destination.

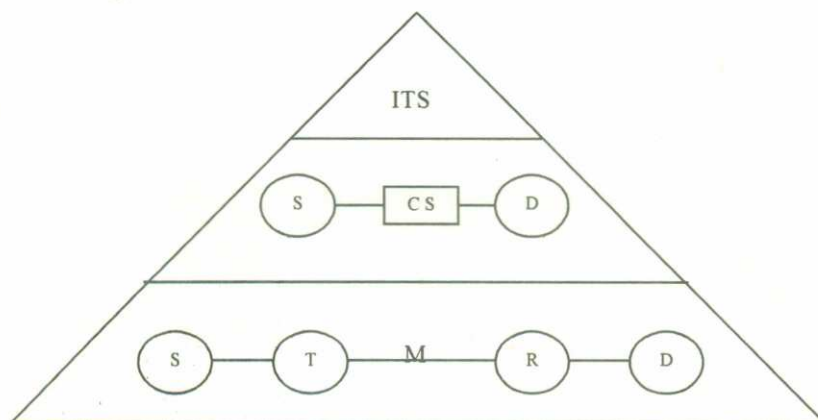
Efficient information transfer deals with four aspects: *low power consumption, minimal bandwidth utilisation, minimal transfer delay* and *low cost*. Reliable transfer means *error free* transmission of signals in the presence of **noise** and **attenuation**. By noise we mean unwanted disturbances that affect and distort the signals. By attenuation we mean the loss of signal strength caused by a communication medium. Reliable transfer also means *guaranteed delivery* to the specified destination.

Direct transmission of source signals, i.e., signals representing information, via a communication medium is usually not efficient and reliable. Hence, communication systems do not directly transmit the source signals. Instead, they operate upon the source signals in some ways using certain **mathematical functions** and superimpose them on another signal, called **carrier**, in order to achieve efficient and reliable information transfer. The source signals present themselves in two fundamental forms: **analog** and **digital**. Correspondingly, the communication systems operate upon the signals in the analog or digital domain. The present trend is to use digital signals and **digital communication systems**. Since source signals are operated upon using mathematical functions at the transmitting end, the equivalent inverse functions need to be carried out at the receiving end to recover the original signals. In communication system terms, these functions are known by special names such as **coding, multiplexing** and **modulation** at the transmitting end. The corresponding inverse functions at the receiving end are known as **decoding, demultiplexing** and **demodulation**.

A communication *link* provides point-to-point connection between two stations. In a communicating environment, there is a large number of stations that communicate amongst themselves from time to time. For this purpose, each station needs access to all the other stations. Establishing separate communication links connecting each station to every other station is very expensive and is not done. Instead, one link from each station is terminated on a centrally located **switching centre** which is capable of establishing connections between pairs of end stations as required from time to time. Many such switching centres are then interconnected to enable connectivity between stations across switching centres. Interconnected switching centres each having a cluster of end stations form a **telecommunication network**. In this Unit, we learn about the communication link related aspects and in Unit 5 about the network related aspects.

4.2 COMMUNICATION SYSTEMS

An information transportation system may be exploded to reveal its component systems as shown in Figure 4.1.



ITS = Information Transportation System

S = Source CS = Communication System

D = Destination T = Transmitter R = Receiver M = Medium

Fig. 4.1: Components of an information transportation system

At the broad level, information transportation involves an information source and an information destination connected via a communication system. The communication system has three major components: a transmitter, a receiver and a communication medium. A variety of functions are performed in some order by the transmitter and the inverse of these functions are performed by the receiver in the reverse order. At present, four media are in use in communication technology. Depending on the energy form and the medium used, four types of modern communication systems are prevalent. The type of communication system, form of energy and the corresponding communication media are shown in Table 4.1.

4.2.1 Electrical Communication

Electrical Communication, transfer of information using electrical energy, is the major form of communication in existence today. Traditionally, copper cables have been used for electrical communication. Basically, there are three types of copper cables:

- Unshielded twisted pair (UTP)
- Shielded twisted pair (STP)
- Co-axial cables.

Table 4.1: Types of Communication, Energy and Media

Communication Type	Energy	Media
Electrical	Electrical	Metallic conductor
Optical	Light	Optical fibre or free space
Radio	Electromagnetic	Free space
Satellite	Electromagnetic	Satellite and free space together

Twisting and shielding of pairs of wires serve to improve the reliability of information transfer. Twisting reduces electromagnetic interference at low frequencies. Shielding the

wire pair with metallic braid reduces the effect of noise and electromagnetic interference at high frequencies. Twisted pair cables are ideally suited for low capacity applications over short distances whereas co-axial cables are used for high capacity applications and long distance transmission. UTP cables are used in applications like data transfer whereas co-axial cables are used in applications like video transfer. For example, local area networks (LANs) use UTPs whereas cable TV operators use co-axial cables. Six categories (Cat) of UTP cables, identified as Cat 1 to Cat 6, have been developed and standardised. Cat 1 and 2 are no longer used. The speed capabilities of other categories are shown in Table 4.2.

Table 4.2: Speed Capabilities of UTP Cables

Category	Speed	Application
Cat 3	10 Mbps	Telephone, 10 Mbps LAN
Cat 4	16 Mbps	16 Mbps LAN
Cat 5	100 Mbps	100 Mbps LAN
<i>Enhanced Cat 5</i>	155 Mbps	ATM
Cat 6	1 Gbps	Gigabit LAN

LAN = Local area network

ATM = Asynchronous transfer mode

MBPS = Mega Bits per Second

GBPS = Giga Bits per Second

4.2.2 Optical Communication

Light traverses through optical fibres or in free space. Optical fibres are made of glass or plastic through which light can travel. The optical fibre acts like a tunnel. When light is launched at one end of the optical fibre it reaches the other end. Light energy has a much higher capacity than electrical energy to carry information. Accordingly, optical fibres are said to have a much higher information carrying capacity than copper cables. Optical fibres have several advantages over the electrical cables, but are not without problems. The merits of the optical fibres stem from the fact that the basic material used in their construction is non-metallic and electrically non-conductive. The problems are mostly due to their delicate structure.

The conductive metallic paths of electrical cables act as antenna and pick up electromagnetic radiation from radio frequency sources, power line currents and other industrial machinery. They also give out electromagnetic radiation when carrying high-speed signals, and thus produce cross talk in cables running adjacently. Electrostatic discharges like lightning find a path in the metallic conductor, thereby causing damages in many ways. In contrast, the non-metallic fibre optic cables are immune to radio frequency and other electromagnetic interference. In optical cables the information is transmitted by packets of photons that have no charge and hence do not lead to electrostatics discharges. There is no possibility of spark or short circuit when a fibre is cut. Signal transmission in fibres is contained totally within the fibre. Therefore, no electromagnetic radiation takes place from the optical fibre.

Fibre cables are the thickness of a human hair. Any dirt obstructing the optical port causes poor transmission. The thin dimension results in a low weight for a given length when compared to electrical cables. However, being thin and somewhat brittle in nature, fibres tend to break easily if bent beyond a certain limit. While the fibre cables are safe from electrical discharge and shock, a direct viewing into the optical point can be harmful to the eyes.

A simple fibre optic communication system consists of a light source as transmitter, a fibre as medium and a photodetector as receiver as shown in Figure 4.2.

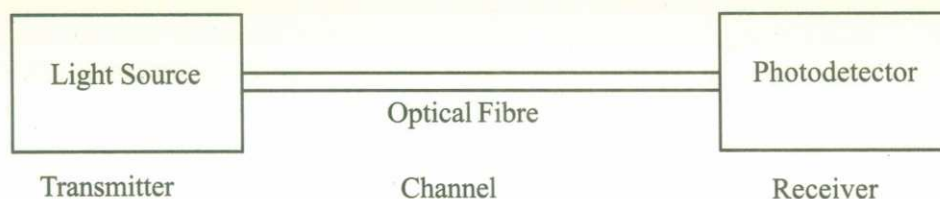


Fig. 4.2: Basic fibre optic communication system

Two principal light sources are used in fibre optic communication systems; **light emitting diodes (LEDs)** and **injection laser diodes (ILDs)** or simply **laser diodes**. LEDs are low-cost devices, but the light beam of LEDs is divergent and has many frequencies of emission. LASER stands for *light amplification by stimulated emission of radiation*. Light radiation is in the form of photons. By stimulating the radiation of photons, the light output can be amplified. This is the principle used in lasers. Light output from a laser is sharp and is concentrated around a single frequency. At the receiving end of a fibre optic communication system is an optical detector or a photodetector. It senses the luminescent power incident upon it and converts the same into a corresponding electric current. Although a variety of photodetectors are known only semiconductor photodiodes are used extensively as photodetectors in fibre optic systems. They have a small size compatible with optical fibres, high sensitivity and fast response time. There are two principal types of semiconductor photodiodes: p-i-n photodiode (PIN diode) and avalanche photodiode (APD). LEDs, ILDs and semiconductor photodiodes are suitable for fibre optic systems for the following reasons:

- 1) Physical dimensions of these devices suit that of the optical fibres.
- 2) Representation and retrieval of information is easily done with these devices.
- 3) They function efficiently in the low-loss regions of optical fibres.
- 4) They are capable of supporting a wide range of applications.
- 5) Their operating and fibre coupling efficiencies are high.

Physically, an optical fibre is composed of these layers: a core, a cladding and a protective coating as shown in Figure 4.3.

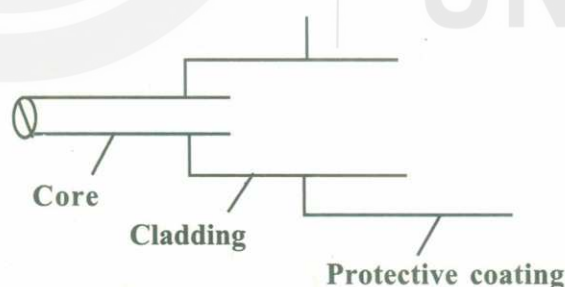


Fig. 4.3: Structure of an optical fibre

The core may be made of glass or plastic. Glass is used when fibre attenuation is to be minimised. When the core material is glass, the cladding is either glass or plastic. Fibres which use plastic for core have higher attenuation, but are cheap. Being economical, the fibre with plastic core and cladding are used in many applications. Cladding serves two purposes: 1) since cladding is manufactured in a controlled environment, the interface between the core and the cladding has a uniform refractive index throughout the cable; 2) cladding protects the core from absorbing surface contaminants. The protective coating gives mechanical strength to the fibre.

The refractive index of the cladding is less than that of the core. Depending on the material used, the refractive index of the core may be uniform or it may gradually decrease

from the centre towards the surface of the core. If the index is uniform in the core, there occurs an abrupt change in the refractive index value when we move from the core to cladding or vice versa, i.e., at the core-cladding interface. Signifying this abrupt change, such fibres are known as **step index** fibres. When the refractive index changes gradually, the fibres are known as **graded index** fibres. The cladding has generally a uniform refractive index.

The propagation of light along an optical fibre may be described in terms of a set of guided electromagnetic waves called the modes of the optical fibres. Only a certain discrete number of modes are capable of propagating along the fibre, depending upon the geometry of the fibre. The physical dimension and the configuration of the fibre may be chosen such that only one mode propagates through the fibre or many modes propagate. Accordingly, fibres are classified as **single mode** fibres or **multimode** fibres. Mode types and refractive index types together give rise to four different types of fibres, as shown in Table 4.3. Of these four, only three are practically realisable. Restricting the propagation to a single mode demands that the core dimension be relatively small. When the core dimension is small, it is difficult to introduce graded refractive index. As a result, the graded index single mode fibres are not practicable as technology stands today.

Table 4.3: Fibre Types

Fibre	Single Mode	Multimode
Step index	Yes	Yes
Graded index	No	Yes

Optical fibre transmission systems are currently being favoured for a variety of applications in telephone and data networks throughout the world. Major telecommunication administrations are opting for fibre optic links as an alternative to co-axial and high frequency cable pair systems. Some of the current applications are as follows:

- Undersea cables
- Intercity long-haul trunks
- Inter-office trunking in metropolitan areas
- Community antenna television loop
- Very high-speed data transmission
- Metropolitan area voice/data networks
- Local area networks

The other mode of optical communication is to use free space as medium. When light is beamed in free space like a pencil ray, there should be no obstruction in its path before reaching the receiving point. It is necessary that the communicating parties are in sight of each other with no obstruction in between them. Such a form of communication is known as *line-of-sight* (LOS) communication. The two communicating parties may be at very far off places and not being able to see each other with naked eye, but the condition is that there should be no obstruction between the sending and receiving points. We may call this LOS as *electromagnetic* LOS.

4.2.3 Radio Communication

Electromagnetic energy radiates through atmosphere in the form of waves. This energy is more commonly referred to as **radio waves** which are invisible to the naked eye and inaudible to human ears. Both radio and satellite communications make use of electromagnetic energy for information transfer. Radio broadcasting and mobile telephony are examples of radio communication. Radio frequencies range from 30 kHz to 300 GHz, i.e., 30×10^3 Hz to 300×10^9 Hz. Within this range of frequencies, there are bands of

frequencies in which the radio waves tend to exhibit a particular way of propagation from one place to another. Depending on the way in which radio waves propagate, they may be placed under four broad categories:

- Ground waves
- Troposcatter waves
- Sky waves
- Line of sight waves

At the lower end of the radio frequency (RF) range lie the ground waves. They tend to follow the terrain of the earth when propagating. The atmosphere that extends up to 500 km altitude is composed of three layers: troposphere at the bottom, stratosphere in the middle and ionosphere at the top. Certain ranges of RF frequencies are scattered by the troposphere and propagate as a result. They are called troposcatter waves. Another band of higher frequencies is reflected back to the earth by the ionosphere and these are called sky waves. Radio waves above the sky waves are unaffected by the atmosphere and pierce through it. They require LOS for communication. The frequencies in the range 3 GHz to 30 GHz are known as **microwave** frequencies. Radio communication systems that operate in this region are called **microwave radio**. Microwave radio needs LOS. Hence, two microwave towers on the ground can at best be separated by distances that are determined by earth's curvature.

Two parameters, *transmitter power* and *receiver sensitivity*, are important in radio communication systems. Larger the transmitter power, farther is the distance at which the signals can be detected. Receiver sensitivity is an indication of the smallest signal power that can be recognised by a receiver. Higher receiver sensitivity means that the receiver can detect even very low signals. For example, if a mobile station transmits larger power it can cover a longer range. Similarly, if a mobile handset has higher receiver sensitivity it can *catch* the signals better and at longer distances.

Signals within a transmitter are almost always in the form of electrical energy. For radio transmission, electrical energy signals need to be converted to electromagnetic energy signals. This is done by a properly designed antenna or aerial. When high frequency electrical signal is fed to a transmitting antenna, the antenna radiates into space electromagnetic signals that are exact replica of the electrical signals. Similarly, when electromagnetic waves cut a receiving antenna, it produces exact replica electrical signals that flow to the receiver set via the cable attached to it. Antennas are of different types. Examples include *dish*, *dipole*, *rhombic* and *omni directional* antennas.

4.2.4 Satellite Communication

Satellite communication also uses microwaves as carrier for source signals. Signals are beamed to the satellite from a ground station using a microwave carrier, called **uplink** carrier. Satellite receives the signals and retransmits them to the ground over a different carrier, called **downlink** carrier. Satellite communication is used for broadcast applications like TV or radio and also for two-way communication like telephone conversation. A TV studio uplinks information to the satellite and the downlink transmission is picked up by cable operators and distributed to TV sets. A satellite acts like a big microwave repeater in the sky. It receives the uplink signal, filters out the noise from the signal, amplifies the clean signal and retransmits the same to the ground at downlink frequency. The functions of noise filtering, signal amplification and frequency translation are performed by a device called **transponder** inside the satellite. Usually there are many transponders in a satellite each having a certain capacity to handle signals. For example, **INSAT-2** (Indian National Satellite System 2) series of satellites have 18 transponders each and a transponder is capable of putting through about 10,600 telephone calls at a time. At present, three frequency bands in the microwave region are in use in satellite communication as shown in Table 4.4. Within the band, higher frequencies are assigned to uplink and lower to downlink.

Table 4.4: Satellite Frequency Bands and Range

Nomenclature	Uplink Range	Downlink Range
C-Band	5.9 – 6.4 GHz	3.7 – 4.2 GHz
Ku-Band	14.0 – 14.5 GHz	10.9 – 11.7 GHz
Ka-Band	27.5 – 31.0 GHz	17.7 – 21.2 GHz

All the receiving stations within the line-of-sight (LOS) of the satellite can receive the downlink signals. This is how satellite naturally acts as a broadcast communication system. To get LOS between ground stations and the satellite, it is necessary that the satellite is visible electromagnetically to the ground segment for all the 24 hours in a day. This is achieved by placing the satellite in **geostationary orbit** that lies at an altitude of about 36,000 km from the earth's surface. In this orbit, the relative motion between the satellite and the earth is zero and the satellite appears stationary when seen from the earth.

An interesting aspect of satellite communication is that the cost of communication is independent of the distance between the source and the destination. It is immaterial whether the destination is 10 km or 5000 km away from the source. In either case, what is involved is one uplink and one down link transmission as far as the satellite is concerned. This is not the case with other forms of communication like optical fibre or microwave radio, where the cost increases proportionately with distance. As a result, satellite communication is economical for long distance communication. However, satellite communication involves much higher delay because the signals will have to travel a distance equal to twice the altitude at which the satellite is placed, i.e., 72,000 km approximately, even if the source and destination are separated only by a few kilometers.

A satellite in orbit experiences small disturbances due to some weak forces such as attraction from moon, sun and other planetary bodies, pressure of solar radiation and variations in the earth's gravitational force. These forces cause the satellite to drift from its orbit, retard or accelerate its motion and affect its orientation (attitude) towards the earth. To overcome the effect of these forces, every satellite is provided with an attitude and orbit control system (AOCS) on board. The rate of fuel consumption in carrying out the AOCS functions and the quantum of fuel available on board are two factors that determine the lifetime of a satellite. There are also other factors such as battery life and solar cell degradation. It is expensive in terms of fuel consumption to maintain the exact altitude and orbit positions of a satellite at every instant of time. By allowing a satellite to have an orbital plane that is inclined at a small angle with respect to the earth's equatorial plane and to drift marginally around its mean position, fuel can be conserved, thereby extending the useful life of the satellite. In this case, the satellite is said to be geosynchronous, i.e., having the same mean orbital period as the earth, but possibly with a small non-zero relative speed at any given instant, i.e., it is not truly stationary. A geostationary satellite is always geosynchronous, but the reverse is not necessarily true. Indian satellite, INSAT-1C, is operated in geosynchronous mode to save fuel. In the case of geosynchronous satellites, ground station antennas must have capability for minor orientation maneuver to keep them pointed towards the satellite.

Self Check Exercise

- 1) In one telecommunication system, a satellite link costs Rs. 2.5 per minute for certain application. For the same application, the cost in an optical fibre system is Re. 0.005 per km per minute. Determine the break-even distance between the source and the destination, beyond which the satellite communication is cheaper.

- Note : i) Write your answer in the space given below.
ii) Check your answer with the answers given at the end of this Unit.

.....

.....

.....

.....

4.3 SIGNAL, NOISE AND ATTENUATION

Signal, noise and attenuation are three fundamental conceptual quantities in communication systems theory. They govern the performance of communication systems. We discuss these quantities in this section to obtain a greater insight into the role they play in communication systems.

4.3.1 Signal

Signal means an electrical, optical or electromagnetic quantity that **varies with time**. When plotted as a graph with X -axis representing time, the signals appear like waves as shown in Figure 4.2. A signal may be periodic or non-periodic. The waveform of a periodic signal repeats itself after a certain *time period* T as shown in Figure 4.2. Communication systems predominantly deal with periodic waveforms. The inverse of the time period ($1/T$) gives the number of times the waveform repeats itself in a *second*, when T is specified in seconds. This quantity ($1/T$) is the *fundamental frequency* or simply the *frequency* f of the waveform. The frequency, i.e., the number of cycles per second, is denoted by unit **Hz (hertz)** in honour of the German communication scientist, Rudolph Hertz.

Signals take a definite amount of time to traverse through communication media. The time taken, called *propagation delay*, is usually so small that the transfer appears instantaneous to human beings for all practical purposes. However, the design and analysis of communication systems need to take into account the propagation delay more often than not. Maxwell, a British physicist, first showed that signals traverse at the speed of light in free space or vacuum. Maxwell's result leads us to a spatial property of the signals. To understand this, let us assume that the signal shown in Figure 4.2 is generated from a source and is propagated through free space.

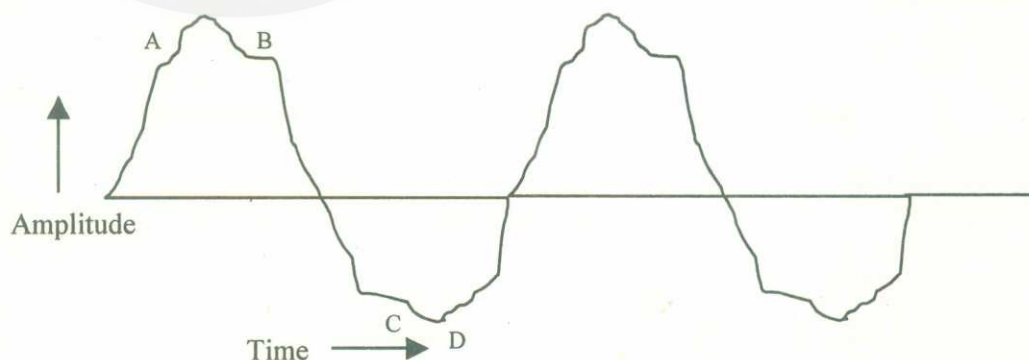


Fig. 4.2: A periodic signal

Consider the signal points A, B, C, D . Point A occurs first in time followed by B, C and D in that order. Now, consider a distant point P in space in the path of the signal. When signal point A arrives at P , signal points B, C and D would have travelled some progressively smaller distances and lie somewhere in space between the source and the spatial point P . If we consider all the points of the waveform, we can then visualise the entire wave pattern appearing in space and occupying a certain length in space. We call this as the *wavelength* λ of the signal. The wavelength and the frequency are related to the speed of propagation as shown in Equation 4.1.

$$f\lambda = c$$

(4.1)

where c is the velocity of light and has a constant value of 3×10^8 meters/second. It may be noted that frequency and wavelength are inversely related as per Equation (4.1). Hence, if the frequency is high, the wavelength is small and vice versa.

4.3.2 Pure Waves

There are certain waveforms whose parameter variations can be characterised by simple trigonometric functions like *sine* and *cosine* functions. Such waveforms are called **pure waves**. They are well understood mathematically. Almost all communication systems make use of pure waves at one stage or another in the process of information transfer. For example, wherever a carrier signal is used, it is invariably a pure wave. Whenever a signal is to be radiated in free space it is in the form of pure waves. Electronic circuits that deal with pure waves have been in vogue for a long time now. Oscillators, amplifiers and filters are all designed to operate upon pure waves. Light energy at a given frequency is also a pure wave. Thus, pure waves play an important role in communication systems.

As mentioned above, pure waves are *sine* and *cosine* waveforms. They are defined mathematically as in Equations 4.2 and 4.3. They are shown graphically in Figure 4.3 (a) and 4.3 (b).

$$v = V_m \sin (2\pi ft) \quad (4.2)$$

$$v = V_m \cos (2\pi ft) \quad (4.3)$$

The sine and cosine functions have degrees or radians as arguments. The quantity $2\pi ft$ is in radians. When t is equal to the period of the waveform T , the quantity $2\pi ft$ assumes a value 2π as $f = 1/T$. Thus, the waveform repeats itself after every 2π radians. Certain equivalence values of time, radians and degrees and the corresponding sine and cosine values are given in Table 4.5. The values of the sine and cosine functions lie in the range -1 to +1. Hence, the maximum value of the pure waves in Eqs. 4.2 and 4.3 is $\pm V_m$.

It may be noted from Table 4.5 and Figure 4.3, that the sine function is positive in the range $0 - \pi$ ($0 - 180^\circ$) and the cosine function in the ranges $0 - \pi/2$ and $3\pi/2 - 2\pi$. If we shift the sine function curve in Figure 4.3a to the left by $T/4$, i.e., $\pi/2$ or 90° , and extrapolate at the end, the resulting curve is identical to the cosine curve. The same effect may be obtained by shifting the cosine curve to the right by a time factor of $T/4$ and extrapolating at the beginning. We call this time relationship between sine and cosine curve as **phase difference**. We say that the cosine curve **leads** the sine curve or the sine curve **lags** behind the cosine curve by $\pi/2$ or 90° . Figure 4.3c shows an inverted sine wave. The phase difference

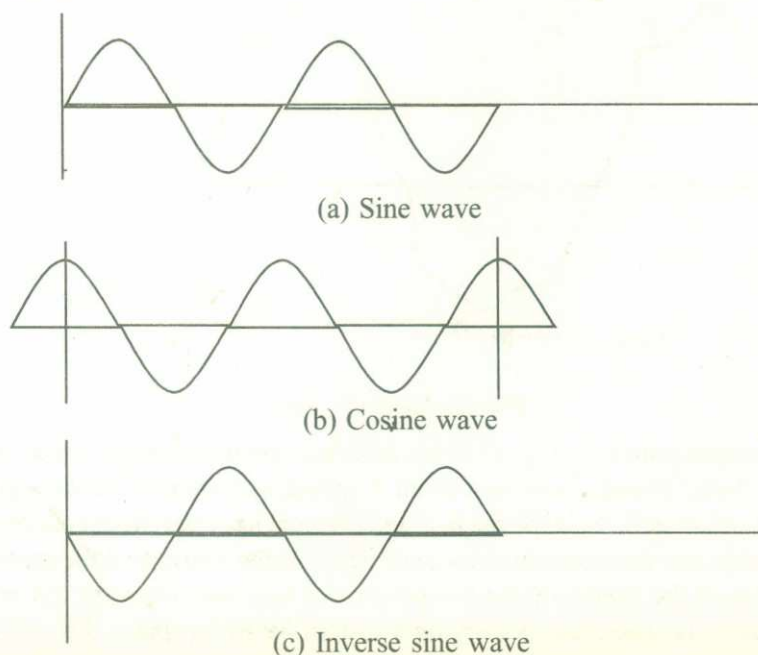


Fig. 4.3: Sine, Cosine and Inverse Sine wave patterns

between the sine and inverted sine waves is π or 180° . It is possible to obtain a family of pure waves by shifting the sine waves by different phase values. We may now say that a pure wave is characterised by three parameters: maximum amplitude, frequency and phase.

Table 4.5: Some Sine and Cosine Function Values

Time	Degrees	Radians, x	$\sin x$	$\cos x$
0	0	0	0	1
$T/12$	30	$\pi/6$	0.5	0.866
$T/8$	45	$\pi/4$	0.707	0.707
$T/6$	60	$\pi/3$	0.866	0.5
$T/4$	90	$\pi/2$	1	0
$T/3$	120	$2\pi/3$	0.866	-0.5
$3T/8$	135	$3\pi/4$	0.707	-0.707
$5T/12$	150	$5\pi/6$	0.5	-0.866
$T/2$	180	π	0	-1
$3T/4$	270	$3\pi/2$	-1	0
T	360	2π	0	1

4.3.3 Complex Signals

Complex signals are those which are not definable by simple functions as in the case of pure waves. We now introduce a mathematical concept concerning analysis of complex periodic signals. This concept is of fundamental importance in communication systems. The mathematics is due to a French mathematician, Fourier, and is known as *Fourier analysis* after his name. Fourier showed that any reasonably behaved periodic waveform $g(t)$ can be represented by an infinite series comprising a constant term A and a sum of *sine* and *cosine* terms as shown in Eq. 4.4.

$$g(t) = A + \sum a_n \cos(2\pi nft) + \sum b_n \sin(2\pi nft) \quad (4.4)$$

A , a_n and b_n are called *Fourier coefficients*. f is the fundamental frequency of the waveform and nf is the frequency of the n -th harmonics of the waveform. The harmonic frequency is an integer multiple of the fundamental frequency. Fourier results tell us that any complex signal is a combination of two or more pure waves. The power of Fourier analysis lies in the fact that any complex waveform can be studied easily by analysing its component pure waves. Most of the practical waveforms encountered in communication systems are complex and it would have been impossible to comprehend and handle them without the help of Fourier analysis.

4.3.4 Noise

Every communication process is prone to disturbance from both external and internal sources. For example, when two persons are conversing, if a plane passes by, the conversation is affected. This is external disturbance. If a person coughs while he is talking, the speech is interrupted. This is internal disturbance. Any disturbance to a signal, be it external or internal, is called **noise**. At the atomic level, the universe is in a constant state of agitation. This agitation is the source of a very great deal of noise particularly in electrical and electronic communication devices. Hence, it may be said that, in general, when a signal traverses through a communication system, it is affected by noise generated from within and outside the system. Noise is present in nature in different energy forms.

The most common form of noise that we encounter is in sound energy form. In order to affect the signal, the noise must be in the same energy form as the signal. For example, electrical signal can be affected by only electrical noise and not by optical noise. Similar is the case with optical and electromagnetic signals.

Noise affects the reliable transmission of signals. If the signal power is more dominant than the noise power, then the noise would have no appreciable adverse effect on the signal. This is the case in good communication systems. The ratio of signal power P_s to noise power P_N , therefore, acts as a quality index of communication systems. Since the value of signal power is orders of magnitude larger than the noise power, this ratio, denoted as SNR, is measured and expressed using a logarithmic scale and is defined as under:

$$\text{SNR} = 10 \log (P_s / P_N) \text{ decibels (dB)} \quad (4.5)$$

SNR, being a ratio, is just a number and has no dimensions like length or time. However, the measure is given a unit called *Bel* in honour of Alexander Graham Bell, who first demonstrated *telephony*. For the kind of numbers encountered in communication systems for signal and noise powers, *Bel* as a unit originally defined was found unsuitable and a smaller unit, *decibel*, ($1 \text{ Bel} = 10 \text{ decibels}$), was adopted as the standard unit for measuring SNR.

4.3.5 Attenuation

When signals traverse through a medium, the signal energy gets spent and the strength (power) of the signal gets reduced. For example, one's speech can be heard only up to a distance as the speech energy dissipates as it travels through air. The reduction in the signal strength on account of traversal through a medium or a device is called **attenuation**. Almost every communication medium has the property of attenuating the signal it carries. In other words, the media are lossy in nature. Lossless media are unknown at present. As a result, there is a limitation of the distance to which a signal of given strength can traverse in a medium. Beyond that distance, the signal power becomes so low that it is not possible to detect the presence of the signal. In case the signal is to be carried beyond this distance, it becomes necessary, at that point, to boost or **amplify** the signal. For example, speech can be conveyed over a distance by forming a human chain with individuals appropriately spaced. What is spoken at one end is relayed by the intermediaries until it reaches the destination. In this case, the intermediaries are acting as amplifiers of the small but detectable signal that reaches them. By the time the signal reaches the intermediaries, it is not only attenuated but also affected by noise. The intermediaries also weed out the noise associated with the incoming signal and send out a clear and loud signal to the next intermediary. In electronic communication, such intermediaries are called **repeaters**. Filtering of noise and amplification of signal are the two functions of the repeaters.

In Section 4.3.4, we learnt that decibel (dB) represents 10 times the logarithm of the ratio of signal power to noise power. Mathematically speaking, dB represents the ratio of any two power quantities, not necessarily that of signal and noise only. This fact is used to measure and express attenuation also in terms of dB. Attenuation A in dB is defined as

$$A = 10 \log (P_R / P_T) \text{ dB} \quad (4.6)$$

where P_R = signal power at the receiver and P_T = signal power at the transmitter. P_R being smaller than P_T on account of attenuation, the ratio P_R / P_T is less than unity and its logarithm has a negative value. In a fashion similar to that of the attenuation of a communication medium, the gain of an amplifier may also be expressed in dB. Thus dB as a measure of relative power plays a very important role in communication systems.

Self Check Exercises

- 2) Given that the fundamental frequency of a signal is 200 kHz, calculate its period, wavelength and the frequency of its 5-th harmonic.
- 3) Should SNR have a large or small value for a good communication system? Give reasons for your answers.
- 4) In a communication system, the noise power is 0.1 mW and signal power is 1 W. Determine the SNR in dB.

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

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

.....

.....

.....

.....

.....

.....

.....

4.4 DIGITAL COMMUNICATION

Communication systems may be analog or digital. Analog systems accept analog signals as input and digital systems accept digital signals. Analog communication has been known for a long time (100 years or more) whereas digital technology is relatively new (40 – 50 years). Early telephone networks used analog signals and even today a major part of the telephone networks around the world is analog. Both analog and digital systems coexist in different parts of the world and are designed to interwork with each other. Digital communication is preferred to analog communication for reasons of efficiency and reliability. Almost all communication systems world over are shifting to digital technology. At present, TV and radio broadcasting remains predominantly analog. Here also, the trend is towards digital broadcasting.

Analog and digital representations of information are equivalent of one another. Analog signals represent information as a continuous time varying waveform. For example, pure waves discussed in Section 4.3.2 are analog signals. On the other hand, digital signal represents information as discrete quantities. Discrete quantities are numbers like 4.7 V (volts), 5.2 A (amperes) and 34.5 °C. When a point value is read from a curve, it is a number and is discrete. When a number of discrete values are taken and interpolated between the successive points, a curve is obtained. Thus we can convert analog representation into digital and vice versa. We often need to do this conversion. These conversions are done by devices known as **analog to digital converter (ADC)** and **digital to analog converter (DAC)**.

At the core of digital representation of information are *sampling theorem* and *coding theory*. A discussion of these aspects appears in Unit 8, Block 2 of MLI-101 on Digital Information. Briefly, a continuous analog waveform can be sampled at close enough intervals such that the original waveform can be reconstructed by interpolating the sampled values. This, then, allows us to deal with sampled values only for communication. The sampled values which are *discrete quantities* are coded using the binary number system

having only two digits 0 and 1. Such coded values are called *digital quantities*. The sampled values appear as pulses. Since the pulses are coded in binary system, the entire process of digitisation is called *pulse code modulation* (PCM). When the above system of sampling and coding is adopted, any signal, video, audio, text or picture, can be represented by a string of *ones* and *zeros*. We then need only to design systems that are capable of communicating strings of *ones* and *zeros*. Digital communication systems do exactly that. They are very rugged and reliable as only two signal values are handled.

Within the communication systems the signals are in the electrical domain. Hence digital quantities are represented by two discrete voltage values, say 0 V and 5 V, inside the communication system. One of the voltage levels represents a binary '1' and the other a binary '0'. If the digital voltage levels are of the same polarity, i.e., both positive or both negative, then the representation is called **unipolar**. It is desirable to choose levels from both polarities, say -2.5 V and +2.5 V, so that the average voltage level is zero. Such a representation is called **bipolar** and more widely used.

Digital signals are less affected by noise and hence lead to more reliable communication. This is the reason why digital communication is preferred over analog communication. In general, it may be said that the noise is analog in nature and hence adds/subtracts with analog signals easily. The value of the analog signal is invariably changed by noise and hence the signal at the receiving end is erroneous. The magnitude of error is directly proportional to the noise level in comparison to the signal level. Smaller signal values are affected more severely. On the other hand, digital signals have only two levels that are well separated. An error occurs in digital signals only when the noise level is strong enough to change one of the voltage levels to the other. This requires a very high noise level which generally does not occur. Thus, digital signals are more rugged and less prone to noise.

Self Check Exercise

- 5) In a digital communication system, + 2.5 V and -2.5 V are used to represent binary '1' and '0' respectively. What noise amplitude levels can cause errors in this system?

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

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

.....

.....

.....

.....

.....

.....

.....

.....

4.5 COMMUNICATION FUNCTIONS

As mentioned earlier, communication systems perform a variety of functions to improve the efficiency and reliability of information transfer. The basic functions performed by a communication system at the transmitting end and the corresponding inverse functions performed at the receiving end are listed in Table 4.6. The functions performed at the transmitting end are collectively known as **signal design** and those at the receiving end as **signal detection**.

Transmitting end functions	Receiving end functions
Signal generation	Signal recovery
Encoding	Decoding
Modulation	Demodulation
Multiplexing	Demultiplexing
Time reference indication	Time reference locking
Transmission	Reception

Signal generation, i.e., representation of information to be transported in the form of a signal, is the first step in the communication process. Digital communication being superior to analog communication in terms of efficiency and reliability, digital representation of information is almost becoming universal. Readers are advised to consult Unit 8, Block 2 of MLII-101 on Digital Information for further details on digital representation of information.

Encoding or simply **coding** is the process of modifying the signal representation to achieve efficiency and reliability. Coding is prevalent throughout the communication process and is performed at four different stages:

- 1) Source coding
- 2) Media coding
- 3) Channel coding
- 4) Line coding

Source coding takes into account the characteristics of the source to improve efficiency. For example, human speech has about 40% pauses and a telephone conversation is always half-duplex, i.e. only one person speaks at a time. Such facts can be used to achieve better utilisation of speech channels, particularly on long-distance routes. Media coding deals with standards for recording and exchanging audio, video, text and picture information in digital form. Channel coding is concerned with error detection and error correction to ensure error-free transfer of information. Finally, line coding is used to make sure that the transmitter and the receiver are in synchronism. With the advances in integrated circuit (IC) technology, chips have been developed to implement certain coding and decoding techniques and such chips are called **codec** chips. Codec stands for coder and decoder.

Modulation is the process of selecting an appropriate carrier and varying one or more of its characteristics according to the source signal. *Demodulation* is the process of extracting the source signal from the varying characteristics of the carrier signal. In the context of modulation, the source signal is often called the *baseband* signal. Two common forms of modulation used in broadcasting, *amplitude modulation* and *frequency modulation*, are familiar to us by their respective abbreviations **AM** and **FM**. The frequency of the carrier signal is usually much higher than that of the baseband signal. Hence, modulation is also described as a process of *frequency translation*.

The need for modulation stems from nature itself. The spectrum of frequencies present in nature is very large ranging from 1 Hz to 10^{22} Hz or more. The human ear is sensitive to frequencies in the range 20 Hz to 20 kHz. This range of frequencies is known as the **audio spectrum**. Any frequency less than 20 Hz or greater than 20 kHz is not heard by the human ear. Radio broadcasting is concerned with transmission of audio frequencies over long distances. The baseband signals in this case, i.e., the audio signals, cannot travel more than a few hundred meters. As we know, sound dies down after a short distance. On the other hand, the frequencies used in AM broadcasting, 550 kHz to 1600

kHz have the property of traversing along the surface of the earth for distances of about 50 Km without significant attenuation. Hence, baseband audio signals are used to modulate the amplitude of a carrier signal in the AM broadcast range and the modulated carrier is launched into space. The radio receivers are tuned to the carrier frequency and the receiver electronics extracts the baseband audio signal from the carrier and plays the same via a loudspeaker. To quote an example, Delhi B radio station uses a carrier frequency of 666 kHz.

In digital communication, we need modulation to transmit 1s and 0s. We have seen that 1s and 0s are represented by two different voltage levels in digital systems. Transmitting these voltage levels directly on the cable or free space is not efficient. On the other hand, the technology for transmitting pure waves is well developed, reliable and efficient. Hence, digital modulation systems represent 1s and 0s by two different pure waves that differ in some way. As brought out in Section 4.3.2, there are three characteristic parameters of a pure wave. By choosing different values for any one or more of these parameters we can obtain different pure waves. Accordingly, there are three basic digital modulation techniques:

- 1) Amplitude Shift Keying (ASK)
- 2) Frequency Shift Keying (FSK)
- 3) Phase Shift Keying (PSK)

In ASK, two pure waves of the same frequency and phase but with differing maximum amplitudes are used. One amplitude denotes a binary '1' and the other binary '0'. In FSK, pure waves of two different frequencies with the same maximum amplitude and phase are used. In PSK, two pure waves having the same maximum amplitude and frequency but with differing phases are used. IC chips and circuits that implement modulation and demodulation functions have been developed. Such circuits are called **modem** circuits or simply modems.

Effective and full utilisation of communication resources is a way of lowering costs. The information carrying capacity of communication media like copper or optical cables is much larger than the capacity demanded by a single information transfer application. Use of a communication medium on a dedicated basis for a single application implies only partial utilisation of the available capacity. If more than one information transfer can be carried simultaneously on a single medium, then the medium capacity will be better utilised. *Multiplexing* is the technique of organising multiple information transfers simultaneously over a single resource. There are two fundamental ways in which multiplexing is done, one in the frequency domain and the other in the time domain: *frequency division multiplexing* (FDM) and *time division multiplexing* (TDM) respectively. In the context of optical signals, a technique similar to FDM known as *wave division multiplexing* (WDM) is used. An example of multiplexing that is right in our homes is in the co-axial cable of cable TV. A single cable carries the video signals of several channels in multiplexed mode.

In any communication process, the transmitter and the receiver have to work in time synchronism. For example, in order to receive a programme on your radio or TV, you have to tune in your radio or TV at the same time when the programme is being broadcast. In other words, the transmitter and receiver have to be active at the same time in order for information transfer to take place. For this purpose, the transmitter establishes a time reference before information signal transfer begins and the receiver locks on to this time reference to receive the signals. There are two ways in which time reference is established between the transmitter and the receiver of a communication system: **asynchronous** and **synchronous**.

The final act in a communication process is the actual *transmission* of information signals. Transmission via an electrical or optical cable is relatively straightforward. The flow of signals needs to be enabled. It is almost as simple as turning on a light switch. In radio and satellite communication, the process involves converting electrical energy signals to electromagnetic energy signals, which is done via antennas.

Self Check Exercise

- 6) A co-axial cable has a bandwidth of 200 MHz. If a video channel requires 5 MHz band, how many such video channels can be multiplexed on this cable?

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

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

.....

.....

.....

.....

4.6 SUMMARY

In this Unit, the fundamental aspects of communication technology have been discussed. First, a perspective of a modern communication system is presented. The main objective of modern communication system, i.e., to transport information from one place to one or more places, has been explained. Four types of modern communication systems viz. electrical, optical, radio and satellite communication systems are then discussed. The fundamental aspects of signal, noise and attenuation are then brought out. Signals, as pure waves and complex waves, are explained. Fourier analysis of periodic complex waveforms is presented. The effect of noise on signals and the importance of signal to noise ratio are brought out. The importance of digital communication is then highlighted. Finally, the communication functions performed by a transmitter and a receiver are explained. These functions include coding, modulation and multiplexing at the transmitter end and the corresponding inverse functions at the receiver end.

4.7 ANSWERS TO SELF CHECK EXERCISES

- 1) The cost of satellite communication is independent of the distance between source and destination whereas the cost of fibre optic communication is directly proportional to the distance.

Let the break-even distance be X km. Then, we have $2.5 = 0.005X$.

$$X = 2.5/0.005 = 500 \text{ km}$$

Therefore, satellite communication is cheaper for distances above 500 km.

- 2) Fundamental frequency $f = 200 \text{ kHz}$

$$\begin{aligned} \text{Period } T &= 1/f \\ &= 1/200 \text{ kHz} \\ &= 0.5 \times 10^{-5} = 5 \times 10^{-6} = 5 \mu\text{s} \end{aligned}$$

Wavelength is given by Eq. (4.1).

$$\begin{aligned} \lambda &= c/f = 3 \times 10^8 / 200 \times 10^3 \\ &= 1.5 \times 10^3 \text{ m} = 1.5 \text{ km} \end{aligned}$$

- 3) SNR should have a large value. SNR is the ratio of signal power to noise power. For a good communication system, the signal power must be larger and the noise power lower. Hence SNR must be high.

- 4) SNR in dB = $10 \log (P_s/P_n)$
- $$\begin{aligned} &= 10 \log (1000/0.1) \\ &= 10 \log (10000) \\ &= 10 \times 4 = 40 \text{ dB} \end{aligned}$$

- 5) Binary '1' = +2.5 V Binary '0' = -2.5 V Mid-point = 0 V

Any voltage level above zero will be seen by the receiver as a binary '1' and any voltage level below zero as binary '0'. At exact zero level, there is an uncertainty. It could be interpreted either way. An error means '1' is received as '0' and vice versa. This can happen only if +2.5 V is affected by a noise level of more than -2.5 V so that the resultant voltage is below 0 and hence interpreted as '0'. Similarly, a noise level of more than +2.5 V can change a '0' as '1' (+ve value). Hence the minimum noise level required to cause an error is +2.5 V for '0' and -2.5 V for '1'.

- 6) Video bandwidth = 5 MHz

Total bandwidth available = 200 MHz

Therefore, the number of channels that can be accommodated = $200/5 = 40$ channels.

In order to avoid interference between two adjacent channels, a frequency gap called *guard band* is provided between them. If we assume a guard band of 1 MHz, then the number of channels that can be multiplexed is $200/6$, i.e., 33 channels.

4.8 KEYWORDS

Attenuation	: Loss of signal power caused by the communication media.
Bipolar	: Representation of digital signals using voltages of different polarities.
Carrier	: A high frequency signal that carries information signals.
Coaxial cable	: Pair of conductors placed in the form of concentric cylinders along the same axis.
Communication media	: Media like copper and optical cables used in telecommunication.
Communication system	: Any telecommunication system like electrical or satellite communication system.
Decoding	: Performing mathematical operations on coded signals to detect and possibly correct errors if any.
Demodulation	: Extraction of base signals from the high frequency carrier.
Demultiplexing	: Extraction of different information signals from a multiplexed stream of signals.
Digital communication	: A form of communication in which information is represented and transmitted by one's and zero's.
Encoding (coding)	: Performing some defined mathematical operations on digital signals to improve reliability and efficiency.
Energy form	: Electrical, optical or electromagnetic energy.
Fourier analysis	: Analysis of complex waveform in terms of simple pure waves.
Frequency spectrum	: Range of frequencies in telecommunication systems.
Geostationary satellite	: A satellite whose instantaneous angular velocity and mean orbital period are the same as that of earth.
Geosynchronous satellite	: A satellite whose mean orbital period is the same as the period of revolution of earth, but its instantaneous angular velocity may differ slightly from that of earth's rotational speed.

Injection laser diode (ILD)	: An optical source that produces sharp light beam when current is applied to it.
Light emitting diode (LED)	: A low-cost optical source that produces light when electric current is applied to it.
Modulation	: Embedding base signals in high frequency carriers.
Multiplexing	: Sending more than one information signal on the same communication medium.
Noise	: Disturbance in a communication system.
Photodetector	: A device that accepts light as input and produces electric current as output.
Repeater	: A device used in microwave radio communication which filters noise from the incoming signal, amplifies and retransmits the signal to the next station at the same frequency.
Shielded Twisted Pair (STP)	: Shielded pair of wires that are twisted.
Signal	: A time varying electrical, optical or electromagnetic quantity representing information, speech, video, etc.
Transponder	: A device inside the satellite that receives uplink signals, filters out noise, amplifies and retransmits at the downlink frequency.
Unipolar	: Representation of digital signals using voltages of the same polarity.
Unshielded Twisted Pair (UTP)	: Unshielded pair of wires that are twisted.
Wavelength	: The physical distance occupied by one cycle of a signal wave while propagating in free space.

4.9 REFERENCES AND FURTHER READING

Brewster, R.L. (1986). *Telecommunications technology*. New Delhi: Affiliated East-West Press.

Carne, Bryan. E. (1987). *Modern telecommunication*. New York: Plenum Press.

Cherin, A.H. (1983). *An introduction to optical fibers*. Tokyo: McGraw Hill.

Couch, Leon W. (1998). *Modern communication systems: principles and applications*. New Delhi: Prentice Hall of India.

Viswanathan, Thiagarajan. (2004). *Telecommunications switching systems and networks*. New Delhi: Prentice Hall of India.

UNIT 5 NETWORKS AND NETWORKING

Structure

- 5.0 Objectives
- 5.1 Introduction
- 5.2 Networking Basics
 - 5.2.1 Fully Connected Networks
 - 5.2.2 Switched Networks
 - 5.2.3 Network Subsystems
- 5.3 Switching Techniques
 - 5.3.1 Circuit Switching
 - 5.3.2 Packet Switching
 - 5.3.3 Cell Switching
- 5.4 Major Telecommunication Networks
- 5.5 Telephone Networks
 - 5.5.1 Landline Networks
 - 5.5.2 Mobile Networks
 - 5.5.3 Data Transmission in Telephone Networks
- 5.6 Integrated Services Digital Network
 - 5.6.1 Motivation for ISDN
 - 5.6.2 New Services
 - 5.6.3 ISDN Channels
 - 5.6.4 User Interfaces
- 5.7 Broadband ISDN
- 5.8 Summary
- 5.9 Answers to Self Check Exercises
- 5.10 Keywords
- 5.11 References and Further Reading

5.0 OBJECTIVES

The developments in networks and networking have increased the accessibility of information. It is, thus, necessary that you should be conversant with their basics. In this Unit we will be discussing the basics aspect of Networks and Networking.

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

- what a telecommunication network is;
- what a switched network is;
- different subsystems that make up a network;
- different forms of signalling;
- different switching techniques;

- differences between circuit and packet switching;
- cell switching as a new technique;
- major telecommunication network types;
- landline and mobile networks;
- how data communication is done on landline and mobile networks;
- what is ISDN and the new services that it offers;
- motivating factors for ISDN;
- ISDN channels and user interfaces; and
- broadband ISDN.

5.1 INTRODUCTION

In Unit 4, we learnt about the modern communication systems that are used extensively in forming telecommunication networks. We studied about wire (electrical) communication, optical fibre communication, wireless (radio) communication and satellite communication. Radio communication may be further classified as **short-haul** and **long-haul** radio. Mobile communications and wireless linkage to a telephone exchange are examples of short-haul radio. High capacity microwave and short wave broadcasting fall in the category of long-haul radio. All these communication systems are used in networks. In this Unit, we learn about the basics of networks and networking techniques. We shall discuss aspects like what constitutes a network, how networks are formed, how networks function, etc. We would also see how some of the communication functions like multiplexing and coding are used in networks. We will learn about the major telecommunication networks.

5.2 NETWORKING BASICS

Networks are not new to us. We are familiar with postal networks, railway networks, airline networks, etc. In a postal network, a post office collects letters from a number of post boxes in the locality, sort them based on addresses and despatch the same to other post offices for distribution in their locality or onward routing towards other destinations. In a railway network, junctions switch tracks for trains to go to different destinations. In this section, we discuss the basics of telecommunication networks which also have features similar to that of postal or railway networks. Information collection and distribution, address processing and path switching for information flow form part of telecommunication networks. A fundamental difference between telecommunication and other networks is that telecommunication networks operate in the electrical, optical or electromagnetic domains.

5.2.1 Fully Connected Networks

Telecommunication involves persons in one part of the world being able to contact and communicate with persons in other parts of the world. For example, a telephone subscriber speaks to different persons at different times. It is not just fixed one-to-one connection. It is multi-point communication connecting different destinations at different times. For this purpose, a subscriber needs access to all the other subscribers that he wishes to contact. The different types of communication *links* that we discussed in Unit 4 provide only point-to-point or direct connection between two stations, i.e., between a transmitter and a receiver. If a subscriber were to be connected to other subscribers using only such communication links, we need as many links as there are other subscribers. If every subscriber is to be connected to every other subscriber, then we need a very large number of links. A network formed this way for five subscribers is illustrated in Figure 5.1. Here, every subscriber is directly connected to every other subscriber. Networks with this kind of connectivity are called **fully connected networks**. There are 10 links in Figure 5.1. The links are assumed to be duplex in the sense that they are capable of transporting

information both ways. If the links were unidirectional, we would need twice the number of links. The number of links required in a fully connected network becomes very large even with moderate number of subscribers. For example, we require 1225 links for fully interconnecting 50 subscribers.

In a general case with N subscribers, $N(N-1)/2$ links are required as reasoned in the following. Let us consider the N subscribers in some order. In order to connect the first subscriber to all other subscribers, we require $(N-1)$ links. With this, the second subscriber is already connected to the first. We now need $(N-2)$ links to connect the second subscriber to the others. For the third subscriber, we need $(N-3)$ links, for the fourth $(N-4)$ links, and so on. The total number of links N works out as follows:

$$L = (N-1) + (N-2) + \dots + 1 + 0 = N(N-1)/2 \quad (5.1)$$

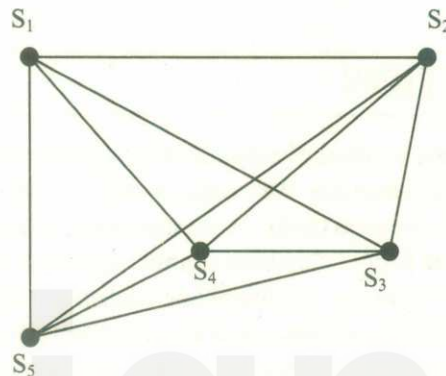


Fig. 5.1: Fully connected network with five subscribers

Establishing separate and direct communication links connecting each subscriber to every other subscriber as shown in Figure 5.1 is very expensive and is impracticable. Hence, this is usually not done.

5.2.2 Switched Networks

The number of communication links can be reduced significantly by using a **switch** along with communication links for establishing connection between subscribers. Switch, as you know, is a device that is used to make and break electrical connections. Networks formed with one or more switches and communication links are known as **switched networks**. A switch is variously known as a **node**, a **switching node**, a **switching system**, a **telephone exchange**, a **switching office** or a **switching centre**. A switched network with one switch is illustrated in Figure 5.2. Here, one communication link from each subscriber is terminated on a centrally located switch. Hence, the total number of communication links is equal to the number of subscribers in the system. This represents a significant reduction in the number of links when compared to a fully connected network. For example, for a 50-subscriber system, we need only 50 links in the switched network as against 1225 links in the fully connected network. The link that connects a subscriber to the switch is called **access link** or **subscriber link** or **local loop**. Subscriber link may be wireless or a landline. In mobile communication, we use wireless access links.

The switch in a switched network is capable of establishing connections between pairs of subscribers as required from time to time. A subscriber who initiates a connection is known as the *calling subscriber* and the destination subscriber is known as the *called subscriber*. Since networks are used not only for human communication but also for machine communication, the terms *source entity* and *destination entity* are used in place of calling subscriber and called subscriber respectively. We would use the term *subscriber* or *entity* whichever is appropriate in a given context. Whenever a subscriber initiates a connection, he needs to convey the identity of the called subscriber to the switch. The identification is a telephone number in the case of a telephone network and an IP address in the case of Internet. You are familiar with telephone numbers and IP addresses are discussed in Unit 6. The switch analyses the address and establishes a connection between the calling and called subscribers. If the called subscriber is local,

the connection is established within the switch. If the called subscriber is non-local, i.e., connected to a different node, the switch forwards the address to the appropriate neighbouring node for further processing. This function is identical to sorting, delivering and forwarding letters in a post office. In the context of telephone networks, address processing is known as **digit processing**.

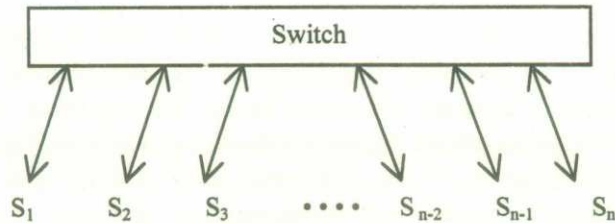


Fig. 5.2: A switched network with one switch

Forwarding or **routing** as it is more commonly called is an important function of the switch. Usually, there is more than one path available to reach a given destination from any node in the network. A node must then decide which one of the paths is to be followed. This routing decision is taken dynamically by the switch usually based on the traffic conditions at that instant. This is unlike the postal network where forwarding is done following a fixed path.

5.2.3 Network Subsystems

Telecommunication network comprises a number of switching centres that are interconnected and each switching centre has a cluster of entities connected to it via access links. Such a network enables connectivity between entities across switching centres. A typical telecommunication network is shown in Figure 5.3. The links that connect switching systems are known as **transmission links**. Transmission links are of very high capacity when compared to access links. Information flow from/to a number access links is multiplexed on one transmission link. Typically, 32 or multiples of 32 access links are multiplexed on one transmission link. The basic speed of an access link is usually 64 or 128 kbps whereas that of a transmission link is 2048 kbps and above. Transmission links are called *trunks* in the context of telephone networks.

You are familiar with dial tone, ringing tone, engaged tone, etc. in a telephone network. These tones are sent out by the switching system to the calling or called subscriber as is appropriate. This function of the switch is known as **signalling**. Signalling is an integral part of any telecommunication network. There are three main forms of signalling:

- 1) Subscriber-Network signalling
- 2) Network-Network signalling
- 3) End-to-end signalling

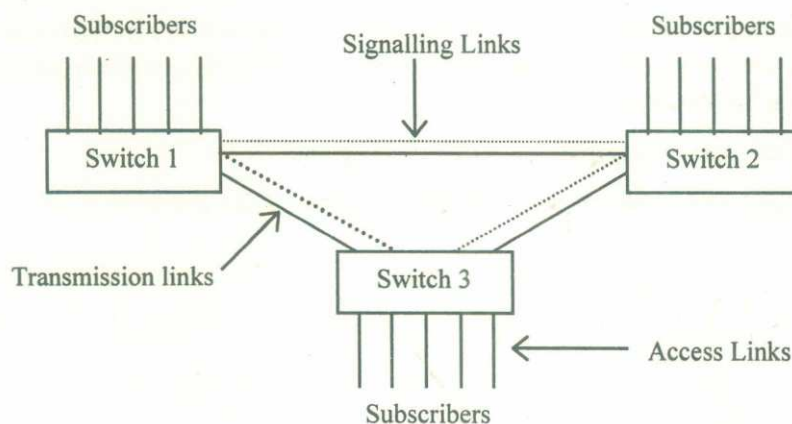


Fig. 5.3: Telecommunication Network

Network-Subscriber signalling, also known as subscriber loop signalling, refers to the exchange of signalling information between the subscriber and the network. Network-

Network signalling refers to the exchange of signalling information between two switching nodes or two different networks. End-to-End signalling is essentially a facility provided by the signalling subsystem of the network to allow calling and called subscribers to exchange signalling information. In fact, end-to-end signalling is the basis for computerised voice answering system that we commonly come across in railway, airline or telephone enquiry services these days. We shall discuss this aspect more in Section 5.5.1.

Signalling techniques fall under two major classes: **inchannel signalling** and **common channel signalling (CCS)**. Inchannel signalling uses the same channel that carries user information. CCS uses a separate common channel for signalling. CCS is a recent development in telecommunications. Figure 5.3 shows separate signalling links depicting the use of CCS. The concept of CCS is a powerful one and is the basis for many value-added services provided by modern telecommunication networks. Most of the present day telephone networks are equipped with inchannel signalling systems. There is, however, little doubt that all networks will gradually evolve to CCS.

From the foregoing discussions, we may observe that a telecommunication network has the following five major subsystems:

- 1) Subscriber end equipment
- 2) Access links
- 3) Switching nodes
- 4) Transmission links
- 5) Signalling links.

Subscriber end equipment may be a telephone (landline or mobile), a computer, a fax machine, etc. It must be remembered that address processing which is an important function in telecommunication network is done by the switching systems. Most of the modern switches are electronic in nature and are built around a computer. In such cases, **switching software** forms a major subsystem of the telecommunication network.

It is not essential that all nodes in the network have subscribers connected to them. Some of the nodes may act as intermediate nodes performing the function of routing only. They are called *transit nodes*. They are known as *transit exchanges* in telephone networks. Transit nodes have only transmission links and no access links terminated on them. The nodes that support subscribers are sometimes called *end offices* or *subscriber exchanges*. End offices have both access links and transmission links terminated on them. When a subscriber exchange routes an incoming trunk channel to an outgoing trunk channel, it functions as an intermediate node for that particular connection.

Self Check Exercises

- 1) Calculate the number of unidirectional communication links required in a fully connected network with 100 subscribers, and in a switched network with one switch.
- 2) One hundred and twenty eight PCM digital voice channels are multiplexed on a trunk line. Determine the data rate on the trunk line. What form of multiplexing do you suggest for this trunk?

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

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

.....

.....

.....

.....

.....

.....

5.3 SWITCHING TECHNIQUES

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

- 1) Circuit Switching
- 2) Packet Switching
- 3) Cell Switching

Circuit switching is the oldest technique and has been in existence for over 120 years. Packet switching is about 50 years old. Cell switching is the most recent one evolved during mid 1990s.

Before we proceed to discuss these techniques definitions of two terms are in order: *channel* and *circuit*. We have been using the term *channel* so far informally. We now define the same. 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 information channels in a multiplexed mode.

5.3.1 Circuit Switching

In circuit switching, a circuit, i.e., channels both ways, is established between the source and destination entities before information transfer begins. In other words an end-to-end connection is established first. The circuit may be established using copper wires, optical fibres, radio links or satellite links or a combination of these media. These media may be used in multiplexed or non-multiplexed modes. Multiplexing may be based on FDM or TDM, or WDM in the case of optical links (See Unit 4). The information pipe is established via a sequence of transmission links between nodes and the access links at the two ends. If four nodes are involved in establishing a connection, then there are three transmission links and two subscriber links. The circuit thus established remains dedicated to the communicating pair until released on request from any one of the communicating entities. Thus, information transfer using circuit switching involves three distinct phases:

- Circuit establishment
- Information transfer
- Circuit release

Information transfer phase begins once the circuit is established and lasts until the release request is received. This implies that once established, the circuit remains dedicated whether there is any actual information transfer or not. The beginning and end of information transfer phase is in the hands of the subscriber. Hence, the charging is based on the duration for which the circuit is held rather than the volume of information transferred. Circuit switching is most commonly used in telephone networks.

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 control associated with information transfer is entirely with the end entities. The network is also freed of any responsibility in information transfer. There is no delay caused by the network in information transfer and the information flows smoothly over the circuit from one end to another. Only propagation delay is experienced.

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 underutilised. Consider a telephonic conversation using circuit switching. First, telephone

conversation is half-duplex. If we assume that each subscriber speaks for half the duration of the call, then the utilisation of the channels and therefore the circuit is only 50%. Second, human speech has about 40% pauses. If we take this into account, the effective utilisation of the circuit works out to be 30%.

Now consider a user accessing a server on the Internet from a client using circuit switching. When a user sits at a terminal and works with a computer interactively, he/she spends time thinking, keying in the query command to the computer, and waiting for a response from the computer before proceeding further. In essence, the user is in one of three states: *thinking*, *keying-in* or *waiting*. During thinking and waiting states no information transfer takes place. During the waiting state the computer is busy processing the user command. Typically, the circuit is busy only for 5–15% of the session time. Hence, circuit switching performs worse for computer-based traffic that is basically bursty in nature. There are some other problems associated with circuit switching. All the links that make up the circuit operate at the same speed. The maximum speed at which the circuit can operate is determined by slowest link even if segments of the circuit are capable of operating at higher speeds. The source and destination systems must run at the same speed determined by the slower of the two. This is not an ideal situation for computer-to-computer traffic where one system may be a PC and the other a super computer. Such considerations led to packet switching as an alternative switching technique particularly for data traffic.

5.3.2 Packet Switching

Packet switching is a special form of a broader class of switching known as *Store & Forward* (S&F) switching. S&F switching is precisely what is practised in the postal system. A post office collects letters from post boxes, *stores* them in the post office for sorting and *forwards* the sorted bags to other appropriate post offices. A similar function is performed by nodes in S&F switched networks. The switching nodes store user messages, analyse destination addresses (sorting) and forward the sorted messages on different output links towards the intended destination. For this purpose, each node is equipped with a processor and some buffer storage. No end-to-end link is set up prior to information transfer. The user deposits his/her message to the switching node via access links. Then on, the network takes the responsibility of delivering the message to the destination node via intermediate nodes as required. The message travels from node to node in the general direction of the destination node. The message is stored in each node on the way for address processing and then forwarded on the appropriate output link. The last node, i.e., the destination node actually delivers the message to the user. This process is identical to the way post office handles letters.

An S&F node processor performs the following functions:

- Receive the full user message and store the same.
- Check the message for data transmission errors and perform error recovery if required.
- Determine the destination address from the user message.
- Choose an appropriate link towards destination, based on certain routing criterion.
- Forward the message to the next node on the chosen link.

Message switching has certain drawbacks. For long messages, it becomes important to ensure that there is adequate storage space on the receiving node before the transmission is initiated. Otherwise, the buffer storage may become full, and part of the message may not be stored, thereby requiring retransmission of the entire message. Often, the message length is not known to the receiver until the message is actually received at least partially. Similarly, if an error occurs during transmission, the entire message may have to be retransmitted. Retransmission of long messages result in large communication overheads in the network. If a high priority short message arrives while a long message is in

transmission, it will have to wait until the transmission of the long message ends. Such drawbacks are overcome in packet switching.

In packet switching, messages are split into a number of packets of some chosen size. For example, a message of length 10000 bytes may be split into 10 packets of 1000 bytes each. The packets are then transmitted in S&F fashion. This is as if the post office forwards its mail in a number of smaller bags instead of one big bag. This, as we know, is not the case. We may therefore say that while message switching is identical to operations in a postal network the packet switching is not. However, the basic mode of operation of storing and forwarding is common to all the three. Consider two end systems communicating over a packet switched network. Messages are split as packets at the source and delivered to the network in order. The packets are numbered in sequence as 1 2 3, etc. to indicate the order. The network transports the packets to the destination. The packets are reassembled based on sequence numbers as messages at the destination. Since multiple messages may be transferred simultaneously, many packets with the same sequence number but belonging to different messages may be in transit at the same time. Therefore, a message number is associated with every packet to indicate the message to which the packet belongs.

Packet switched networks adopt two different approaches for transporting packets from the source to the destination:

- Datagram approach
- Virtual circuit approach

In datagram approach, each packet transmission is treated independently. Every packet carries the complete identification information, viz. destination address, source address, message number, and the packet number, in addition to the actual user data. When a packet arrives at a node, its destination address is processed, one among the many available alternative routes chosen and the packet forwarded on the chosen route. It is possible that there are many packets at a node that need to be sent on the same route. In this case, a queue is formed and the packets are despatched according to a queue discipline.

Since the packets are treated independently in datagram approach, the packets belonging to one message may travel via different routes and arrive at the destination with different delays. This may lead to the situation where the packets of the same message arrive out of sequence at the destination node. The source host delivers the packets of a message in sequence to the network and it is natural to expect that the network deliver the packets to the destination host in proper sequence. Therefore, if the packets arrive out of sequence at the destination node, it becomes the responsibility of the network to **resequence** the packets in proper order before delivery to the destination host. This calls for considerable overhead in terms of buffer storage and processing power at the network nodes. There is yet another overhead in datagram approach. The need to carry elaborate identification information with each packet contributes to data transmission overhead. As a result, it may turn out to be somewhat expensive if a user attempts to transfer large messages comprising many packets using datagram service. Datagram service is ideally suited for transmitting short messages of one or two packet lengths. The main advantage of datagram service is that there is no overhead of circuit establishment and release.

Virtual circuit approach overcomes the overhead deficiencies of datagram service. No resequencing is required and the addressing overhead per packet is significantly reduced. The virtual circuit approach draws upon the idea of circuit establishment as in circuit switching. Instead of establishing a dedicated circuit, it preselects a route from the source to destination, which remains fixed for all packets of a message. There is no separate path selection for each packet and all packets of a message follow the same preselected route. Since the packets are delivered to the network in sequence and they follow the same route on a first-come-first served basis (FCFS), they arrive in sequence at the destination. Hence, there is no need for resequencing.

While the route remains fixed for the duration of information transfer, no channels along the route are dedicated for any particular connection. A channel may be used for transmitting packets of different messages and packets belonging to a particular message may be transmitted via different channels depending on the availability. Since the route is established and released, the term **switched virtual circuit (SVC)** is used to denote such connections. It is important to make a distinction between packet reassembling and resequencing. Packet reassembling is always required to reconstruct the original message but resequencing is required only if the transportation scheme causes out of sequence reception at the destination.

Now let us see how the addressing overhead is reduced. As soon as a virtual circuit 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. Hence, the transmission overhead is reduced significantly.

It must be obvious that virtual circuit connection has the overhead of preselecting the route. Packet transmission does not start until a route is chosen and finalised between the source and the destination. Some users may want to avoid this overhead and use the same virtual circuit for transmitting a number of messages over a period of time. Such a connection is called **permanent virtual circuit**.

A major disadvantage of packet switching is its inability to support real time information transfers. Since the packets are moved in an S&F fashion, they experience considerable delay on account of address processing and possible queuing in each node of travel. The queue is a dynamic phenomenon and the queuing delay may vary from packet to packet. This means that the packets do not reach the destination at regular intervals. The time duration between packets is not fixed, it varies. Delay and delay variations are unacceptable in real time services like audio transmission or video distribution. Packet switching is ideally suited for data services.

5.3.3 Cell Switching

Cell switching is the most recent switching technique evolved during 1990s. The main objective of cell switching has been to arrive at a technique that combines the desirable aspects of both circuit and packet switching and caters to all types of services taking into account the advances in networking technology. As discussed earlier, in circuit switching, the network causes no delay and only propagation delay is experienced in information transfer. Hence, circuit switching is ideally suited for real time services. Packet switching is more flexible and allows more efficient use of network resources. Cell switching is designed to cause minimal network delay ensuring at the same time efficient utilisation of network resources.

Packets are of different sizes in different networks. Therefore, internetworking calls for packet matching operations. If a large packet arrives at a network that supports a smaller packet size, then the large packet needs to be segmented at the entry and reassembled at the exit. Similarly, if a small packet arrives at a network that supports a larger packet size, then the large packet is padded with dummy data. In cell switching, this problem is completely avoided. All networks use a standard cell size of 53 bytes.

Sizing packets has been a major issue in packet switched networks. Large packets mean fewer overheads but they tend to exhibit problems similar to those experienced in message switching. They also perform poorly in the context of real time services. Smaller packets call for too many S&F operations and cause address processing overheads. Some networks, for example Ethernet use variable size packets subject to a maximum size. In such cases, actual size of data needs to be counted in advance and communicated as part of control information in the header part of the packet. Cell switching overcomes these problems by using advances in switching and transmission technology. Cell switching is

extremely fast and works on a very reliable network infrastructure. Reliable infrastructure reduces the S&F overheads and increases the speed of switching. For this reason, cell switching is sometimes called **fast packet switching**. By using a small size of 53 bytes for cells, cell switching ensures support for real time services.

Self Check Exercises

- 3) A circuit switched connection involves five switching nodes. Each node takes 1.0 and 0.1 second for connection set up and release respectively. Once established, the path operates at a speed of 100 kbps. Determine the time required for one information transfer session involving 10 MB of data.
- 4) A datagram packet traverses through five switching nodes before it is delivered to the destination computer. The datagram size 2.5 kB. The average S&F delay in each node is 0.25 sec. The data speed of access links is 100 kbps and that of transmission links is 1 Mbps. Determine the time it takes to transport the datagram from source to destination.

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

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

.....

.....

.....

.....

.....

.....

5.4 MAJOR TELECOMMUNICATION NETWORKS

Modern telecommunication started with telegraph network followed by the telephone network. Since the connections are established by switching in a telephone network and the network is open to public, it is called **public switched telephone network (PSTN)**. When digital computers came about, a new class of networks, known as **data networks** were developed to support computer interconnections. The data network has today evolved as **Internet**. With the maturing of digital technology, a further advanced telecommunication network called **Integrated Services Digital Network (ISDN)** was evolved. ISDN is capable of supporting all types of services like telephony, facsimile, data transfer and video transmission. ISDN has been implemented in some developed nations. It is also being used in India. Recently, wireless has been introduced in telephone access links giving birth to **mobile communication networks**. Mobile networks are also public switched and designed for voice communication. Hence, they may be considered as part of PSTN. Thus, at present, we have three major classes of telecommunication networks in operation:

- 1) Telephone Network (Landline and Mobile)
- 2) Data Network (Internet)
- 3) Integrated Services Digital Network (ISDN).

All the networks are meant for public access. Interestingly, the three major telecommunication networks are built around the three switching techniques that we discussed in Section 5.3. Table 5.1 shows the networks with the corresponding switching technique and the type of switching node that is used in these networks.

Today's telecommunication network in the world is a massive one. The massive nature of the network is obvious when one realises that today's network has over one billion telephone connections, 800 million Internet connections, 900 million mobile connections and about 100,000 exchanges. Data networks are the subject matter of Unit 6. Hence, we discuss PSTN and ISDN in this Unit.

Table 5.1: Networks, Switching Techniques and Switches

Network Type	Switching Technique	Switch Type
Telephone	Circuit Switching	Telephone Exchange
Data	Packet Switching	Internet Node
ISDN	Cell Switching*	ISDN Exchange

* refers to Broadband ISDN as discussed later

5.5 TELEPHONE NETWORKS

Worldwide telephone network is the largest telecommunication network in existence today. A marvellous feature of the telephone network is that every piece of equipment, technique or procedure that has evolved in the last 120 years from a number of giant corporations is capable of working with each other. Compare this with the fact that it is almost impossible to interface the first IBM computer with its own latest system. The enormous complexity of the telephone network is managed by using a hierarchical structure, worldwide standardisation, and decentralisation of operations, administration and maintenance. With the evolution of mobile communications predominantly for voice communication, the telephone network may be viewed in two different classes: landline networks and mobile networks.

5.5.1 Landline Networks

Landline telephony was born in March 1876 when Alexander Graham Bell demonstrated long distance voice communication over a pair of wires. In 1877, operator manned switching systems were set up with the subscribers connected to the operator switchboard giving birth to the concept of telephone networks. Limitations of manual system were quickly recognised and automatic switching systems were invented and put into operation by early 1890s. The first automatic switching system was an electromechanical one and is known as **Strowger System** after its inventor Alan B. Strowger. These systems were designed to establish connection in a step-by-step manner. They were in use for about 40 years. In the late 1930s, another electromechanical switch known as **crossbar** exchange came into operation. With the advent of digital computers, electronic switches came into being. The control in electronic exchanges was called **stored program control (SPC)** signifying the use of computer for control purposes. The switching scheme in electronic exchanges may be **space division switching** or **time division switching**. In space division switching, a dedicated physical path is established inside the switch between the inlet and the outlet for the entire duration of information transfer. In time division switching, sampled values of speech signal are transferred at fixed intervals via shared common resource within the switch. If the speech samples are discrete values, then the switching is called *analog*. There is further classification in digital time division switching. Figure 5.4 summarises the classification of automatic switching systems. The subscriber end equipment in the case of a landline network is the traditional telephone. As mentioned earlier, addressing in telephone network is the phone number of the called subscriber. The number that is dialled by the subscriber is conveyed to the exchange using one of the two different techniques:

- Pulse dialling
- Multifrequency dialling.

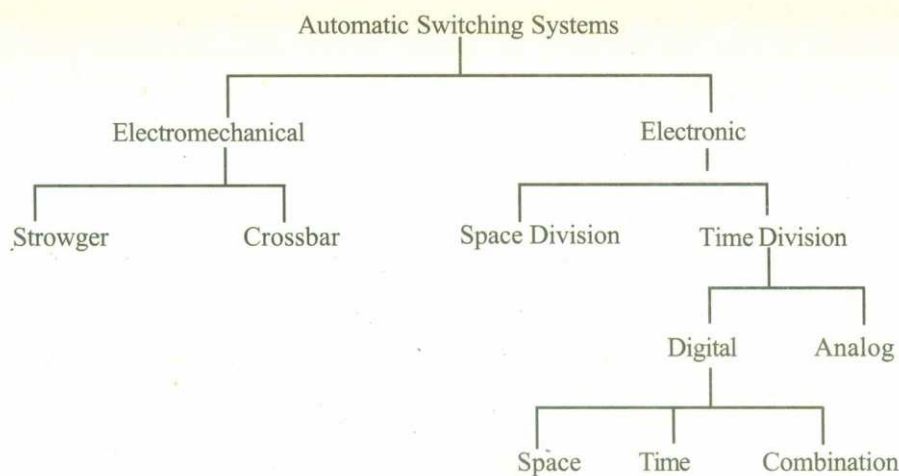


Fig. 5.4: Classification of automatic switches

Pulse dialling is the older technique that originated with Strowger switching. After having been in use for about 100 years, it has been replaced with multifrequency dialling almost universally. In multifrequency dialling, pressing a number button on the phone generates a signal that is a combination of two frequencies or tones. Since 'touching' a button generates a tone, this system of dialling is known as **touch tone dialling**. As two different frequencies are combined to produce the signal, the technical term is *dual tone multifrequency* (DTMF) signalling. The two frequencies are in the same audio range as the speech. This means that once the call is established, pressing a button would send the signal to the other end. If the receiving end has a suitable instrument, the received signal can be decoded to determine the number pressed. A powerful application of touch-tone dialling is the **data in voice answer (DIVA)** system. A customer calling an airline may receive voice announcements like "dial 1 for reservation" and "dial 2 for flight information", etc. Based on the voice announcements, the customer dials further digits which may result in further instruction for additional dialling. Thus, digit dialling and voice response can be interspersed to any level. DIVA system is a typical example of the use of touch-tone dialling that enables interaction between a telephone user and a service provider. This system of digit dialling and voice response is also popularly known as Interactive Voice Response System (IVRS).

5.5.2 Mobile Networks

Mobile communication service is a recent phenomenon. As a major commercial venture, it started in the early 1990s. Rapid expansion has taken place since then. There are as many as one billion mobile subscribers in the world as of year 2004. From the network point of view, the most significant aspect of mobile communication technology is the ease with which subscriber connectivity can be established. From the subscriber's point of view, the personal mobility that enables a subscriber to be in touch with anyone at any time from anywhere is the most impressive aspect of mobile technology. Mobile communication *per se* was first introduced in 1946 in USA to support communication amongst car-borne telephone sets. Early systems used a central radio station with a large antenna mounted at a high point, like a hill, which radiated high power to cover the entire target area within which mobility was supported. Today's personal mobile system (PMS) is based on *cellular communication* principle in which a distributed network of smaller radio stations, called *base stations* (BS), is used to cover the target area instead of one single large central station. The mobile phones need to communicate with only the nearby base station of the distributed network. As a result, they are small in size, lightweight and hand-held type.

Cellular communication has evolved over three generations of mobile systems. The first generation cellular system went into operation in 1983 in the United States and was known as *advanced mobile phone system* (AMPS). This was followed by an implementation of a similar system in UK in 1985. Both the systems used analog communication technology with frequency modulation (FM) for speech transmission.

European telecommunication standardisation forum adopted digital technology and ISDN principles in evolving standards for cellular mobile communications with a view to supporting multimedia services. European standards are known as Global System for Mobile (GSM) communications standards. India adopted GSM standards in 1992. Digital cellular systems like GSM are called *second-generation systems*. An integrated system of personal communication via satellites and mobile networks supporting subscriber access both via satellites and base stations would result in a truly universal communication facility. Such systems are categorised as *third generation mobile systems* and are currently being studied and standardised.

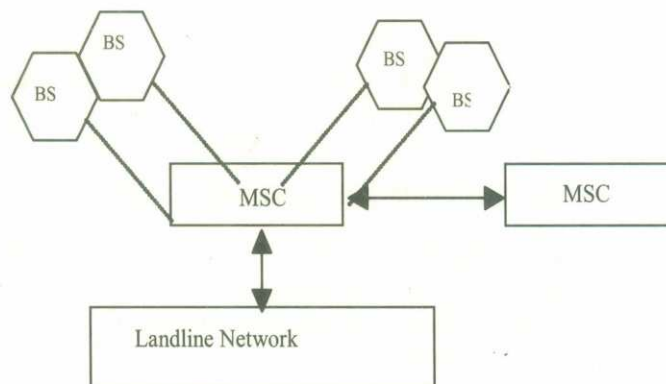


Fig. 5.5: Architecture of a mobile network

The architecture of a cellular mobile network is shown in Figure 5.5. In theory the cells are of the hexagonal shape, but in practice, they are less regular in shape. The cell sizes may not be uniform throughout the cellular region. For example, the cell sizes may be small in the crowded central part of a city and become larger and larger towards the outer areas. Each cell has a base station. Mobile phones in a particular cell communicate via the base station of that cell. The base station functions like a telephone exchange for the mobile phones. Each base station is equipped with a low power transmitter whose power output is such that satisfactory reception is achieved throughout the cell area but no interference takes place with other cells having the same frequency of operation. Obviously, neighbouring cells cannot operate at the same frequency. Cells with the same frequency of operation must be separated by a minimum distance in which the signals are attenuated to a significant level.

All base stations are connected to a mobile switching centre (MSC) via landlines. The MSC in turn is connected to the landline network as well as other MSCs. Mobile switching centre keeps track of the carrier frequency allocations by different cells and manages the overall functioning of the system. When a subscriber moves from one cell area to another, the carrier frequency changes from that of one cell to another. The process of changing over the carrier frequency is known as *hand over* operation. When a subscriber initiates a call, the mobile instrument sends a message to its base station announcing the called telephone number. The base station then allocates an available frequency to the telephone for communication. When the call is finished the frequency is deallocated and made available to another caller in the same cell.

5.5.3 Data Transmission in Telephone Networks

Interest in digital data transmission came about in late 1950s to allow remote job entry terminals to work with digital computers. The telephone network (PSTN) with its wide coverage was the obvious choice as the telecommunication infrastructure for this purpose. At that time, the telephone network worldwide was based on analog technology. Digital technology was yet to appear on the scene. Even today, landline telephone network is predominantly analog in good part of the world. Transmission of digital data signals over PSTN networks demands that the digital signals be converted to analog form at transmitting end and vice versa at the receiving end. A modem performs this function (Refer Unit 4 Section 4.5). Data communication using modems in telephone network is shown in Figure 5.6. The digital interface of a modem is connected to the computer and the analog interface to the telephone network. In addition to data exchange, the digital interface permits

transmission control signals to be exchanged between the modem and the computer. The transmission may be half-duplex or full duplex. Half-duplex transmission implies transmission in only one direction at a time. Full duplex transmission means both way transmissions simultaneously. In Figure 5.6, both terminal-to-computer and computer-to-computer connections are shown.

A telephone channel is basically designed to carry analog voice signals with a nominal bandwidth of 3.1 kHz. When the channel is to be used for data transmission, a natural question that arises is *what is the maximum data rate that a voice channel can support?* There are two results in communication theory, which help us answer this question. The first one is Nyquist theorem that states:

In a noiseless channel, the maximum data rate can be increased indefinitely by introducing an arbitrarily large number of levels in the signal.

Claude Shannon extended Nyquist's work to the case of noisy channels and showed that there is a maximum limit for the data rate that can be achieved in a noisy channel. Telephone channels are noisy and Shannon's limit works out to be about 30,000 bits per second (bps) for telephone channels. Modems are designed taking these two results into account for data transmission via telephone lines.

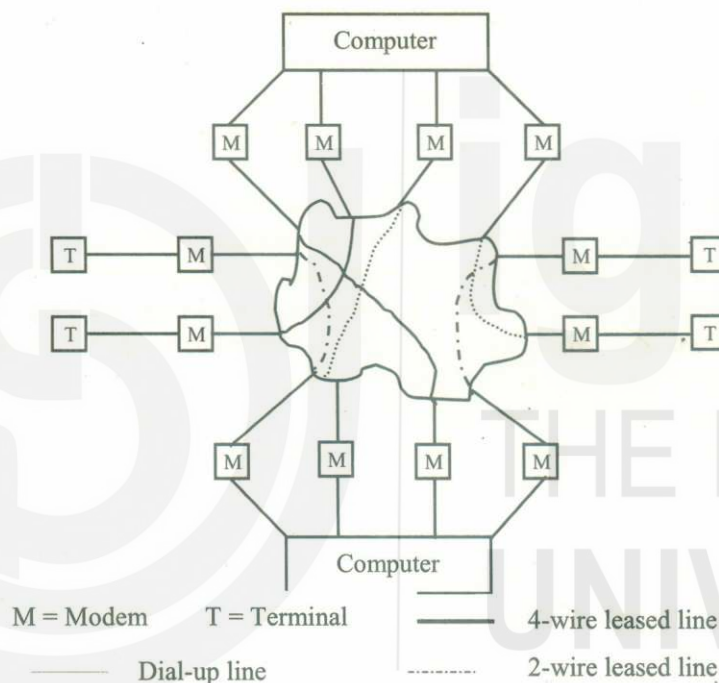


Fig. 5.6: Architecture of a mobile network

Different levels in a signal could be introduced by considering the three fundamental properties of any signal, viz. amplitude, frequency and phase. Multiple amplitudes or frequencies or phase values can be used to represent different signal levels. As discussed in Section 4.5, the corresponding techniques are called amplitude shift keying (ASK), frequency shift keying (FSK) and phase shift keying (PSK) depending on which one of the characteristics has multiple values. Too many levels in one property may make the signal prone to noise. Hence, when large number of signal levels like 16 or 128 are required to achieve higher speeds, a combination of these techniques is used. The most commonly used combination is the multiple phase and amplitude values. This combination is called **quadrature amplitude modulation (QAM)**. Modems are standardised under V-Series by International Telecommunications Union (ITU-T). The features of some of the standard modems are shown in Table 5.2.

Data communication in analog mobile networks takes place by a process similar to that of landline networks as described above, i.e., by using modems. In GSM networks, it takes place via the signalling channel. GSM mobile networks use common channel signalling (CCS), which means that there is a separate channel for signalling purposes. Information transfer channel and the signalling channel are separate and different. Information channel carries voice signals. The signalling channel is digital and uses packet switching technology.

Hence, small messaging system (SMS) messages are transmitted in datagram mode via the signalling channel. This concept, originally evolved in ISDN, is adopted in GSM networks.

Table 5.2: Some V-series Modem Recommendations

V-series	Speed (bps)	Modulation	Application
V.22	1200	4-ary DPSK	FD, D-up, 2-W
V.22 bis	2400/1200	6-ary QAM	FD, D-up, 2-W
V.23	600/1200	FSK	HD, D-up
V.29	9600	16-ary QAM	FD, 4-W
V.32	up to 9600	32-ary QAM	FD, D-up, 2-W
V.33	14,400	128-ary AM-PM	FD, 4-W

HD = half duplex FD = Full duplex D-up = dial-up connection

4-W = 4 wire leased circuit 2-W = 2 wire circuit

5.6 INTEGRATED SERVICES DIGITAL NETWORK

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. ITU-T the then CCITT has been pioneering and guiding the efforts towards the development of ISDN. The process of digitalisation of telecommunication networks started in early 1960s when PCM was introduced for digital transmission of voice signals. With the arrival of transistorised mini computers the possibility of digital switching opened up in late 1960s. ITU-T was quick to recognise the feasibility of digital telecommunication networks and set up a study group called **Study Group D** in 1968 to look at a variety of issues related to the use of digital technology in telephone networks. This study group is the forerunner of today's Study Group XVIII set up in 1976, and has the responsibility for all ISDN related activities within the ITU-T. The first formal definition of ISDN was given by the Study Group D in its recommendations G.702, issued and adopted by ITU-T in 1972:

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.

It is interesting to note that a study that started with the objective of exploring the use of digital technology quickly shifted its emphasis to an Integrated Digital Network (IDN) and within a period of four years even foresaw the emergence of ISDN. From then on, the progress towards ISDN has been steady and unrelenting. By mid 1970s, the concept of common channel signalling using digital technology emerged. Digital end equipments like digital telephone started appearing in the market. In 1984, a comprehensive and generic definition of ISDN was adopted as:

An ISDN is a network, in general evolving from telephony IDN, that 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.

The key points of the above definition have to be noted. First, ISDN is an infrastructure to support a wide variety of services and is not network designed for any specific service. Second, the end-to-end digital connectivity implies that the digitisation process begins right at the user premises. Third, it should be possible to support every conceivable service on ISDN, for any such service is either a voice or non-voice service. Finally, a small set of carefully chosen interfaces should enable the support of all possible services. The users of ISDN should not be burdened-with too many specialised interfaces, but at the same time, an expensive universal interface should be avoided.

5.6.1 Motivation for ISDN

Three factors are responsible for the developments towards ISDN:

- 1) Sociological or social needs
- 2) Economic necessities
- 3) Technological developments

Present day developments in various facets of society call for increasing and complex communication facilities. For example, a biotechnologist today would like to perform a variety of functions simultaneously via communication facilities: examine a blood sample remotely; compare the analytical results of other samples stored in a centralised database; consult his assistant who is presently in a different laboratory; and report the findings as the investigation progresses to his superior who is in another building. To meet such a demand, we need to electronically transmit the microscopic image of the blood sample and produce the same graphically on the computer screen of the biotechnologist, at a rate fast enough to faithfully reproduce the movements of living cells, etc. Simultaneously, the same communication medium is used to access a remotely located database and hold an audio conference with his assistance and superior.

As another example, a senior executive of a company, who often takes important decisions at home late in the evening or while on a holiday, would like give instant effect to his decisions. This may call for access to different computer systems connected in the form of a network, electronic banking facilities, facsimile transmission and desktop image processing facilities, all in one place where he is at present. In other words, automobiles, homes, boats, hotel rooms and railway compartments will need to house electronic workstations. Calling party identification and guaranteed transaction privacy and security are important signalling requirements of systems that can support such applications.

In effect, the society is demanding a telecommunication system that can support universal access to a host of services. In such a system, it should be possible for a user to attach to the network anywhere in the world the equipment of his choice to obtain a particular service. The user will be allotted a permanent identification number or code, like the income tax permanent account number or a social security number, which would be valid for his lifetime. No matter where a user lives, or how often he shifts residence, dialling the number assigned to the user would always ring his telephone, computer system or any other equipment. Similarly, irrespective of wherefrom a user obtains his services, the related charges are debited to his account. The society is looking forward to a telecommunication infrastructure that can carry voice, data, image, graphics, video, etc., simultaneously supported by sophisticated signalling systems.

Society may demand many things, but are these technologically feasible and if so, are these economically viable? We now discuss these points. Traditionally, network providers have put up separate and independent networks to support different services. Telex network, data network, telephone network and CATV networks are examples of such a development. Can the network providers continue the same path of evolution for the plethora of new services demanded by society? Independent networks call for separate administration, maintenance staff, and building for housing switching systems. The independent and duplicate infrastructural facilities lead to high capital cost, low maintenance efficiency and high management cost. In addition, network facilities are never fully utilised as the services are independently supported on different networks. The net result is that overheads turn out to be excessive, leading to economically unviable network services. Obviously, the independent network approach is not viable for future services. One should consider the possibility of sharing resources to keep the cost down. Thus, sheer economic necessity is forcing network providers to look for solutions where many different services may be integrated and supported on common network resources.

Searching for new solutions is of no avail unless technology developments make possible such solutions. In fact, it is the technology factor that brought about the independent network solutions earlier. The end equipments for different services were analog in nature and had different electrical, electronic, signal and communication characteristics. It was

necessary to design different networks to suit each of these devices. For example, the voltage, current and bandwidth requirements of a teleprinter, a computer and a telephone instrument are vastly different. Accordingly the switching systems, signalling systems and the transmission media have to be different. The desire of the network providers to use a common network infrastructure can fructify only if there are uniform signal and communication requirements for all types of services. Such a common substratum is provided by the digital technology. If the end equipment, switching, transmission and signalling systems can all function in the digital domain, the interface and communication requirements can be uniform for all the services. Today, the digital technology has matured to a level where all the above-mentioned functions of telecommunication networks can be realised in the digital domain.

It is interesting to trace the history of the process of digitalisation in the field of telecommunications leading to ISDN. Traditionally, digital technology always appeared first in the field of computers. In late 1930s and early 1940s, while the theory of digital communication (pulse code modulation, sampling theorem, etc.) was coming of age, digital computers were already being built. The first digitalisation step in telecommunication appeared when slow channel Pulse Code Modulation (PCM) technique was introduced for long distance transmission in the early 1960s. Transistorised or integrated circuit based mini computers opened up the possibility of digital switching in late 1960s. These computers were used in the exchanges, and switching was organised in space domain or time domain. Developments in the field of packet switching networks led to the possibility of digital signalling in mid 1970s. Since late 1970s, digital end equipments such as digital telephone and digital facsimile are being developed. ISDN became a reality in 1980s.

Each Plenary Assembly of ITU-T, which takes place once every four years, assigns specific topics related to ISDN to the concerned study groups for in-depth study during the ensuing four-year period. The result of such an organised study in the last 20 years is that today we have a clear picture of ISDN architecture and a plethora of services that can be supported in ISDN. However, the ISDN technology, design approaches and standards are still evolving and will remain the focus development in the coming decades.

Thus, we are now at a point of time when there exists a synergy of social demands, technological feasibility and economic viability with regard to ISDN. Consequently, an era of ISDN will be heralded all over the world during the next few decades or so. As must be obvious by now, ISDN is the apt nomenclature for the telecommunications network of the future in which all network functions will be handled in digital domain and a single network will support a set of integrated services.

5.6.2 New Services

ISDN will support a variety of services including the existing voice and data services and a host of new services. A short list of the important new services is:

- 1) Videotex
- 2) Electronic mail
- 3) Digital facsimile
- 4) Teletex
- 5) Database access
- 6) Image and graphics exchange
- 7) Document storage and transfer
- 8) Automatic alarm services, e.g., smoke, fire, security, police and medical
- 9) Audio and video conferencing

ISDN uses common channel signalling. Hence, there are separate information and signalling channels. The basic information channel is known as **B-Channel** and its speed of operation is 64 kbps corresponding to digital PCM speech signal rate. The signalling channel is known as **D-Channel** and its speed of operation is either 16 kbps or 64 kbps. Signalling network is packet switched and D-Channel carries signalling information in the form of **signalling packets**. A user is allowed to use the D-Channel for transmitting user packets subject to spare capacity being available. B-Channel can be operated in circuit switched or packet switched modes. Popular use of B-Channel is circuit switched. Every connection, either circuit switched or packet switched is established as virtual circuits. In addition to B and D channels, there are several high-speed channels known as **H** channels operating at different speeds.

5.6.4 User Interfaces

Comprehensive user-network interface definitions are key to ensuring worldwide ISDN compatibility. An excellent example of an interface standard that serves us so well and yet goes almost unnoticed is the electrical power user interface. We can purchase an electrical appliance almost anywhere in the world and plug it in our house socket and expect it to work. In fact, we are totally unmindful of the internal wiring, power distribution or generation scheme or the physical dimensions of the sockets, etc. as the final user interface conforms to a universal standard. By contrast, data communication suffers due to presence of a variety of interface connectors, pin assignments and software interfaces.

In ISDN, user-network interfaces have been given careful consideration to avoid potential inconsistencies that may arise. ISDN caters to a variety of services such as voice, data, telemetry and image. In a situation like this, one encounters conflicting requirements. On the one hand, a number of custom designed interfaces may ideally suit each service but would lead to proliferation of interfaces. On the other hand, one single multipurpose interface may turn out to be an 'overkill' for most of the services. One needs to adopt a via media approach.

Keeping such factors in mind, two information access interfaces have been standardised for ISDN:

- Basic rate access interface
- Primary rate access interface

In some sense, these rate access interfaces are analogous to 5A and 15A power sockets at home. The access interfaces cater to low and high speed data rates much as the same way the two power sockets cater to low-power and high-power electrical appliances. The basic rate access caters to low bit rate services. It is defined at a rate corresponding to $2B + D$, i.e., two **B** channels of 64 kbps each and one **D** channel of 16 kbps. This works out to a total speed of 144 kbps. The primary rate access interface caters to services that demand high bit rates and is defined as $nB + D + S$ where **n** may be any value in the range 3-30. In other words, the information channel capacity may be chosen by the user in the range of 192-1920 kbps. The speed of the **D** channel in this case is 64 kbps as signalling for up to 30 information channels has to be carried on the channel. In addition to the **B** and **D** channels, the primary rate interface has a synchronisation channel **S** that runs at 64 kbps. The total data rate of the primary access interface works out to be 2048 kbps for 30 information channels.

Self Check Exercises

- 5) What are the key features of ISDN?
- 6) An ISDN subscriber takes a primary rate access with 8 information channels. What is the data rate on his access link?

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

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

.....

.....

.....

.....

.....

.....

5.7 BROADBAND ISDN

Broadband ISDN (BISDN) is defined as a network that caters to speeds higher than 2048 kbps, i.e., 2 Mbps. Primary access interface runs at a maximum speed of 2 Mbps. Broadband ISDN supports data rates beyond what is supported by the primary interface. In the context of BISDN the original ISDN is referred to as *narrowband* ISDN (NISDN). 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 NISDN. This is because of the ambitious aim of BISDN 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 a 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 by BISDN. The services in BISDN are placed under two broad categories:

- Interactive services
- Distribution services

Interactive services support end-to-end information transfer on real time and bidirectional basis. The most commonly used service in this category is likely to be video telephony. In this service, the telephone instrument has the capability to image and transmit video signals. A dial-up connection brings about both audio and video transmissions both ways. An example of distribution service is electronic newspaper distribution.

The broad range of services envisaged to be supported in BISDN poses a challenge in designing a suitable architecture for the network. It is in this context that cell switching has been chosen as the switching technique for BISDN. The information transfer in BISDN uses **asynchronous transfer mode** (ATM). In ATM, each cell carries a virtual circuit identifier that enables the cells of a particular connection to be switched along the same route. Hence, there is no need for resequencing. Since cells are of small and fixed size, no information transfer can be fitted in one cell. Hence, ATM implements a function termed *segmentation and reassembly* (SAR). Information to be transmitted is segmented in the form of cells and reassembled at the destination. BISDN, ATM and cell switching are the technologies of the future telecommunications.

5.8 SUMMARY

In this Unit, we have covered the basics of switching, different switching techniques and the major telecommunication networks except data networks that are discussed in the next Unit. The Unit starts with a discussion on the basics of networking. The functioning of a switched network is explained. Functions of a switch are brought out. Different components of a network are enumerated. Three switching techniques are then discussed. As a prelude to this discussion, the terms *channel* and *circuit* are defined. Advantages

and disadvantages of each of the switching techniques are brought out. Major telecommunication networks are then placed in perspective. Different types of switches used in telephone switching are then briefly described. The architecture and the operation of mobile networks are then presented. The discussion moves on to ISDN thereafter. Motivation for ISDN, new services offered by ISDN, and ISDN channels and interfaces are discussed. Finally, the Unit discusses the upcoming broadband ISDN. The power of broadband ISDN and its design features are presented. In particular, cell switching and asynchronous transfer mode are emphasised.

5.9 ANSWERS TO SELF CHECK EXERCISES

- 1) The number of bi-directional links required in a fully connected network is given by Eq. 5.1. We need two unidirectional links as a substitute for one bi-directional link. Therefore, the number of unidirectional links = 9900.

The number of unidirectional links in a switched network = 200.

- 2) The data rate of a PCM voice channel is 64 kbps (See Unit 8, Block 2 MLI-101). Therefore, the data rate on the trunk is $128 \times 64 = 8192$ kbps or 8 Mbps. The natural choice for multiplexing is TDM as the information is digital.

- 3) Path set up and path release overhead = $5 \times 1.0 + 5 \times 0.1 = 5.5$ sec.

Data size = 1 MB (Mega Byte) $\times 8 = 8$ M bits (Mb)

Data transfer time = $8000 \div 100 = 80$ sec.

Total time = 85.5 sec.

- 4) Datagram size = $2.5 \text{ kB} \times 8 = 10$ kilo bits (kb)

Total S&F delay = $0.25 \times 4 = 10$ sec. There is no S&F delay at the last node as the datagram is delivered to the destination computer without any queuing.

Transfer time on the access links = $2 \times 10 \div 100 = 0.2$ s. There are two access links, one on either end of the connection.

Transfer time on transmission links = $4 \times 10 \div 1000 = 0.04$. There are four transmission links in between five nodes.

Therefore, the total transfer time = 10.24 s.

- 5) Digital infrastructure and hence universal;

End-to-end digital connectivity;

Support for a wide variety of services including voice and non-voice services and limited set of standard interfaces.

- 6) Primary rate access is $nB + D + S$. The value of n is given as 8.

Therefore, the data rate on the access link = $8 \times 64 + 64 + 64 = 640$ kbps.

5.10 KEYWORDS

Access links	: Communication links that connect the user to the nearest switch.
Asynchronous transfer mode	: Mode of information transfer in broadband ISDN.
Base station	: A wireless station covering a certain geographical area called <i>cell</i> .
Broadband ISDN	: An ISDN designed for very high-speed applications.
Cell	: In the context of mobile communication, a cell denotes a small geographical area. In the context of broadband ISDN, cell is a small digital packet that carries information.
Cell switching	: A switching technique in which information is cast in the form of small cells and transported over broadband ISDN networks.

Cellular communication	: A wireless communication technique in which a large geographical region is divided into a number of cells that are networked.
Channel	: An information pipe with some specified characteristics like bandwidth, capacity, level of attenuation and noise immunity.
Circuit	: A two-way link comprising two channels, one in each direction.
Circuit switching	: A switching technique in which a dedicated path is established between end points for the entire duration of information transfer.
Common Channel Signalling	: Signalling done using a separate channel that is common to many information channels.
Datagram	: An independent packet containing complete address information.
Digit processing	: Processing of a telephone number at the telephone exchange.
End systems	: Equipments at the subscriber premises that connect to a telecommunication network.
Fully connected network	: A network in which all communicating entities are directly connected with each other.
Packet switching	: A switching technique in which large size data is broken as packets and transferred across networks.
Permanent virtual circuit	: A virtual circuit assigned on a long-term basis.
Public switched network	: A switched network open to public.
Reassembling	: The process of assembling packets or cells to reconstruct the original message at the destination.
Resequencing	: The process of arranging the packets in proper order when they arrive out of sequence at the destination.
Signalling	: Indicative or control signals exchanged between a subscriber and a switching node or between two switching nodes.
Space division switching	: A technique in which a dedicated physical path is established inside a switch.
Subscriber	: A user connected to a telecommunication network.
Switch	: A device that establishes and releases connections between subscribers.
Switched network	: A network in which users are connected to a switching system.
Time division switching	: A technique in which a shared common resource is used for transferring information from an inlet to an outlet in a switch.
Transmission links	: Communication links that interconnect switching systems.
Virtual circuit	: A fixed route for information transfer.

5.11' REFERENCES AND FURTHER READING

Pandya, Raj. (2000). *Mobile communication services and systems*. New Delhi: Prentice Hall of India.

Sharda, Nalin K. (1999). *Multimedia information networking*. New Delhi: Prentice Hall of India.

Stallings, William. (2000). *Data and computer communications*. New Delhi: Prentice Hall of India.

Viswanathan, Thiagarajan. (2005). *Telecommunications switching systems and networks*. New Delhi: Prentice Hall of India.

UNIT 6 DATA NETWORKS

Structure

- 6.0 Objectives
- 6.1 Introduction
- 6.2 Data Units
- 6.3 Layered Communication Architecture
- 6.4 ISO-OSI Reference Architecture
- 6.5 Internet Architecture
- 6.6 Internet Protocols
 - 6.6.1 Internet Layer
 - 6.6.2 Transmission Control Layer
 - 6.6.3 Application Layer
- 6.7 Client-Server Architecture
- 6.8 Data Network Classification
- 6.9 Satellite Based Data Networks
- 6.10 Local Area Networks
- 6.11 Protocols for Libraries
- 6.12 Internetworking
- 6.13 Network Management
- 6.14 Summary
- 6.15 Answers to Self Check Exercises
- 6.16 Keywords
- 6.17 References and Further Reading

6.0 OBJECTIVES

In the previous Units you have learnt about basics of communication technology, networks and networking. Data networks have been using almost similar technologies, but have some distinct features. In this Unit we will be discussing the basic aspects of data networks.

After going through this Unit, you will be able to understand the concept of:

- data units and their structure;
- layered communication;
- open systems;
- open System Interconnection (OSI) Model;
- protocols;
- internet communication architecture;

- internet layer protocols;
- IP address classes and IP packet structure;
- client-Server model of computation;
- classification of data networks are classified;
- data communication via satellites;
- bus and ring local area networks;
- wireless LANs and bluetooth technology;
- protocol requirements for library applications;
- internetworking techniques and devices; and
- management aspects of networks.

6.1 INTRODUCTION

Data networks are otherwise known as **computer networks**. The term *data network* is more representative of the functionality of the network. The term *computer network* is more representative of the physical aspect of the network. The main purpose of the data networks is to transport **data** from one end entity to another and the term *data network* signifies this fact. The end entities are computers and data is transported from one computer system to another. Hence, data networks comprise a large number of computers that are interconnected and the term *computer network* signifies this fact. Starting from late 1960s, computer networks have evolved to a global network of computers, now popularly known as **Internet**. As you learnt in Section 5.5, data networks use packet switching as it is ideally suited for transfer of data that is bursty in nature. Irrespective of the switching technique used, data is always transmitted in the form of small units that we generically call as **data units**. We begin our discussions in this Unit with the structure of data units and go on to discuss other aspects of data networks.

6.2 DATA UNITS

Data networks are mainly concerned with transporting messages that are digital in nature. As we have seen in Section 5.3, large messages are never transported in one chunk for reasons of error handling and efficient utilisation of network resources. Large messages are spilt into smaller data units. In packet switched networks, data units are packets and in ATM networks they are cells. Even while using circuit switching for data transmission, large volume data is broken into smaller pieces called *frames* and transmitted frame after frame. Thus, *data unit* is a generic name for *packets*, *cells* and *frames*. A data unit (DU) has the general structure shown in Figure 6.1 with three parts in it: *Header*, *User Information* and *Trailer*. The header is like the envelope in a conventional letter, which carries the addresses. It consists of three fields: *destination address*, *source address*

Data Unit		
Header	User Data	Trailer

Fig. 6.1: Structure of data unit

and a *control field*. As you studied in Unit 5, in virtual circuit connections, the destination and source addresses are replaced by virtual circuit number. The control field has provision to specify *the type of data unit*, *sequence number*, *acknowledgement* and *flow control* of data units. The control field of the header is like the extra information such as *registered*,

insured, acknowledgement due, etc. that a user may inscribe on the envelope. Often the rate at which the destination can receive DUs and the rate at which source station may despatch DUs may not match. For example, if the destination system is slower than the source system, it is necessary to slow down the dispatch of DUs so that they are properly received at the destination. *Flow Control* provision in the control field is used for this purpose. The user data field contains the user information to be transported, i.e., the letter inside the envelope. It is usually much larger than the header and the trailer parts and is the actual user message to be communicated to the destination. The trailer has error check information that can be used for detecting and possibly correcting errors that may occur in the transfer of a data unit. The trailer may not always be present in a data unit.

6.3 LAYERED COMMUNICATION ARCHITECTURE

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 trunk dialling (STD) facility between the two cities. 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 and communicates the calling number, called number, nature of the call, the name of the particular person called, etc. Then, the local trunk operator calls up the other trunk operator in the other city and communicates the details. Remote trunk operator 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 Figure 6.2.

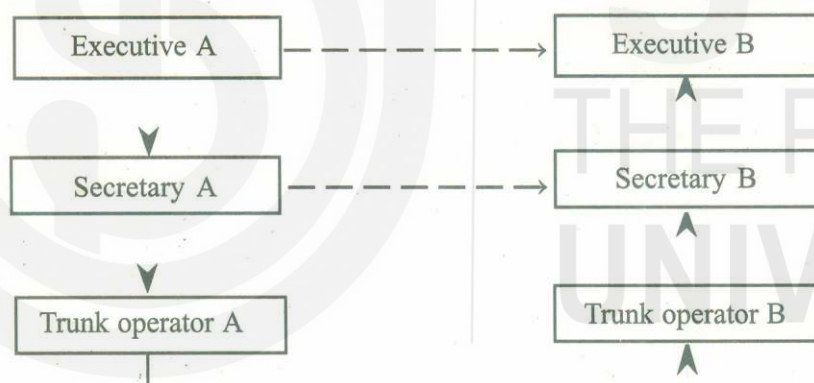


Fig. 6.2: A three-layer structure for a trunk call connection

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.

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 discuss the ISO-OSI architecture in the following section.

6.4 ISO-OSI REFERENCE ARCHITECTURE

Data communication among computers involves a number of functions such as physical transmission of bits, error control, 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 be interconnected as a network, an architecture which is used as standard by all the vendors is required. The heterogeneity covers the following aspects:

- 1) Systems of different vendors
- 2) Systems under different management
- 3) Systems of different complexities
- 4) 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 experience 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 Figure 6.3 taking

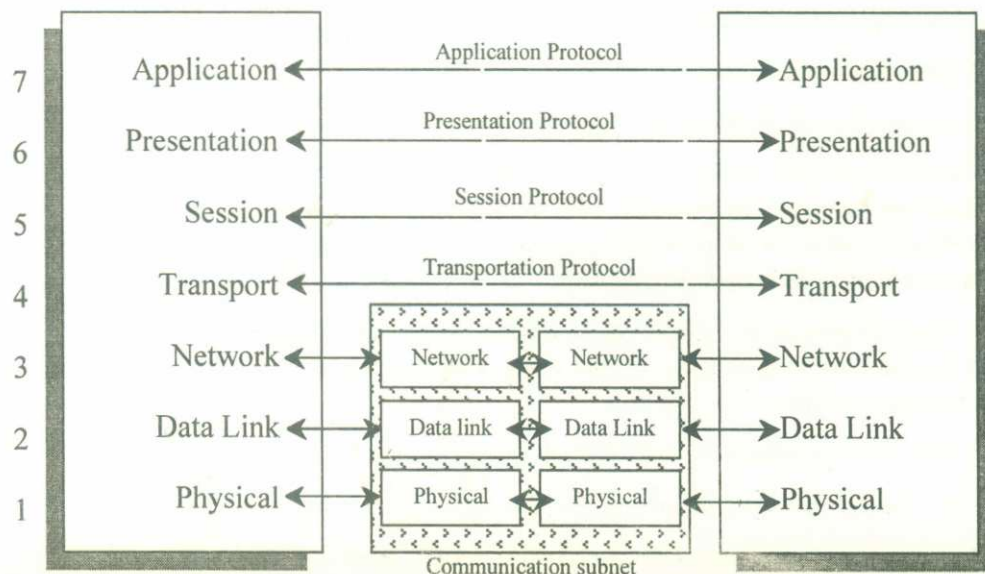


Fig. 6.3: ISO-OSI Reference Model

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**. The corresponding entities in the same layer but in different systems are called **peer entities**. The entities communicate by exchanging data units. The set of rules and standards used in exchanging data units is known as **protocol**. The set of rules for communication between entities in adjacent layers is known as **interface protocol** and the data unit exchanged between them as **interface data unit**. The set of rules for communication between peer entities is known as **peer protocol** and the data unit exchanged between them as **protocol data unit (PDU)**. The PDUs are generically identified as layer-1 PDU, layer-2 PDU and so on. Alternatively, the PDUs may also be identified by prefixing the name of the layer such as transport PDU, network PDU and so on. On different networks the PDUs are also known by some popular names. For example, layer-3 PDU is called a **packet** on Internet. Each layer has a set of defined services, interface protocols and peer protocols. These sets in all the layers are collectively known as **protocol suite**.

Figure 6.3 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 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 as discussed in Unit 4. The data link layer handles errors and organises reliable transport of layer-2 PDUs on a link-by-link basis. Layer-2 PDUs are popularly called as **frames**. The network layer is concerned with the processing of destination addresses, routing of data units and internetworking. 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 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 most layer is the application layer which enables a user to run different applications like electronic mail, accessing a web page, etc.

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.

6.5 INTERNET ARCHITECTURE

Internet architecture is a 5-layer architecture as shown in Figure 6.4. The bottom

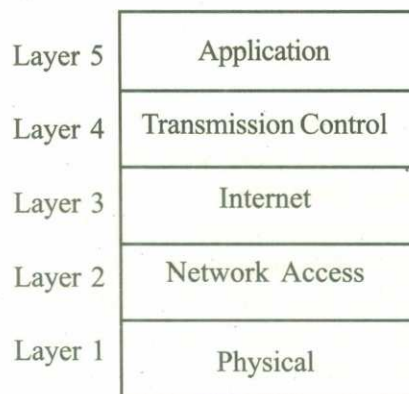


Fig. 6.4: Internet Architecture

three layers are largely as per ISO recommendations. Layer 2 is called as network access layer, layer 3 as Internet layer and layer 4 as transmission control layer. The layer 3 and layer 4 protocols are known as **Internet protocol (IP)** and **transmission control protocol (TCP)** respectively. Initially, ARPANET concentrated on IP and TCP only. As a result, its protocol suite is popularly known as **TCP/IP suite**. A fundamental difference between OSI and Internet architecture is that Internet assumes the underlying network to be a packet switched datagram network whereas OSI assumes a connection oriented virtual circuit network. The network access layer of Internet widened the scope of underlying networks to cover LANs and wireless networks. OSI largely concentrated on wide area networks built around the traditional telephone infrastructure. Over a period of time, a large number of application protocols were developed on Internet to support a wide variety of applications that would run over TCP/IP. We discuss some of the important protocols of Internet in the next section.

6.6 INTERNET PROTOCOLS

The main Internet protocols have been developed for the Internet layer, transmission control layer and application layer. We discuss the protocols in each of these layers starting with the Internet layer.

6.6.1 Internet Layer

The main purpose of the Internet layer is to deliver an IP datagram or a packet to the specified destination on *best-of-efforts* basis. The Internet layer does not guarantee the delivery of the datagram to the destination. It is the function of the transmission control layer to ensure guaranteed delivery of datagrams. The destination host may be on the same network as the source host or on an adjacent network or on a network that can only be reached by traversing through a number of other networks. Taking this fact into account the design of the Internet layer has placed a great emphasis on internetworking. The main services or functions of IP are address processing, routing datagram, congestion control and internetworking.

The first thing that is required to deliver an IP datagram to the destination is the unique and unambiguous destination address. Every computer connected to the Internet is called a **host** and is identified uniquely by an address known as **IP address**. 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 over 20 years and it uses a 32-bit address. With 32 bits we can uniquely address 2^{32} hosts, i.e., approximately 4 billion hosts. But the 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 20 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 256 billion 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 centuries. 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 limit our discussions to IPv4. Readers interested in IPv6 may refer to Further Reading [Stallings or Tanenbaum] 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 network addresses that contain many hosts. Each network is connected to the Internet via a **router** or a **layer-3 switch**. The router has an Internet network address and is aware of all the networks world over 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. Corresponding to three network categories, there are three address classes: **Class A**, **Class B** and **Class C** respectively. A large network identified by Class A address may have as many as 16 million hosts and IPv4 provides for 126 such networks world over. A medium sized network may have up to 64 K hosts and a small one up to 256 hosts. IPv4 provides for 16 K medium networks and two billion small ones. The general structure of IPv4 address has three fields as shown in Figure 6.5. A one-bit class field with a value '0' specifies a Class A address and a 2-bit field with value '10' and a 3-bit field with value '110' specify Class B and C

Class	Network	Host
-------	---------	------

Fig. 6.5: Structure of IPv4

addresses respectively. A 7-bit non-zero and non 7-ones pattern following the class bit specifies the address of a large network. That gives us 126 unique values. The network address fields for medium and small networks are 14 bits and 21 bits long respectively. With these lengths, we can uniquely specify 2^{14} (16 K) and 2^{21} (2 M) medium and small networks respectively. The remaining bits of the 32-bit address specify the host address in all cases. In addition to the above 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. Further, some special address patterns have been specified for addressing one's own computer, broadcast over local network, to get a loop back response from a remote host for testing the path etc.

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 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 183.241.065.179. With experience, it is felt that a hexadecimal representation would have served the purpose of clarity much better.

The large jump from a maximum of 256 hosts in small networks to 64 K hosts in medium networks has resulted in severe under utilisation of IP address space. Most of the organisations, like universities, research laboratories and large corporate houses felt that a provision for 256 hosts on their network is inadequate and, therefore, obtained Class B addresses. But the number 64 K turned out to be too large that these Class B networks remain heavily under utilised. Most of these organisations do not have more than a couple thousands of computers on their networks. The net result is the loss of address space. A

class definition with about 2-4 thousand 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 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. 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 of telecommunication 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.

Yet another development in the field of IP address is the concept of **subnet**. This concept came about when organisations registered as a network on the Internet found that their network is in itself a collection of small networks, typically many Local Area Networks (LANs). For example, a university may have a LAN in each department. Addressing in such an environment needed another level of structuring wherein each of these small networks is treated as a subnetwork to which the hosts are connected. Hence, the host address space is divided into two parts with higher order bits representing the subnet address and the lower order bits the hosts on the subnet. The subnet concept led to a hierarchy in routing. The entry point to each subnet is a router that connects to the Internet router on the one side and to the subnet hosts on the other side.

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

We now discuss the IPv4 datagram structure presented in Figure 6.6. IP datagrams are popularly known as **packets**. The IP datagram structure and the interpretation of different fields in it define the Internet Protocol. IP datagram is a data unit with only

Version	Header Length	Service Type	Datagram Length		
Datagram Identifier			D	M	Fragment Identifier
Time to Live	Upper Layer Protocol		Header Error Control		
Source Address					
Destination Address					
Options					

Fig. 6.6: Structure of IP datagram (functional)

header and data fields. The first thing to realise about IP datagrams is that both the header and the data fields are of variable length. This in fact introduces complications and consequent overheads in processing. This is one of the main reasons as to why Internet runs rather slow and throws up many interruptions. The minimum header length is 20 bytes and the maximum 60 bytes. The maximum datagram size including both the header and the data fields is 64 K. Different networks are allowed to set their own limit for the maximum size well below 64 K. This maximum size set by a network is called the

maximum transfer unit (MTU) of that network. This provision further complicates datagram-processing calling for fragmentation and reassembly of datagrams with different sizes. 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 the network and reassembled at the exit point. In such a case, we need a provision to identify the datagram and its fragments. There may be certain applications where fragmentation may not be acceptable.

As mentioned earlier, in layered architectures implementation details of layers are not standardised and are left to the designer to decide. This approach promotes innovation. In the case of IP layer, a large number of routing algorithms have been designed using a variety of criteria. Some of these algorithms, in the name of robustness, tend to perpetuate the presence of packets on the network even though not intentionally. To safeguard against the possibility of packets living endlessly on the network, the IP header carries a field that specifies the maximum time for which a packet may exist on the network.

From the foregoing discussions, the purpose of most of the fields in the IP datagram must be obvious. The field *service type*, though not well defined in IP is meant to indicate priorities for transfer of datagrams. The 'D' field, when set denies permission for fragmentation. The 'M' field distinguishes between intermediate fragments and the last fragment of the datagram. The field 'upper layer Protocol' specifies the protocol to which this datagram must be delivered. Error free header being important for proper transfer of the datagram, there is provision for checking the datagram header for errors.

In addition to IP, there are several protocols in the Internet layer for controlling and monitoring the functioning of Internet. They include Internet Control Message Protocol (ICMP), Dynamic Host Configuration Protocol (DHCP) and Bootstrap Protocol (BOOTP). There are two protocols in IP layer that deal with address translation between the Internet layer and the network access layer. We discuss these in the context of local area networks in Section 6.8 as they largely pertain to LANs. The protocols are Address resolution Protocol (ARP) and Reverse Address Resolution Protocol (RARP). With the advent of wireless communications, access to Internet via mobile phones and laptop computers calls for special provisions on the Internet to keep track of mobility. This is achieved by having an agent system on the home location of the mobile entity. The issues related to access to Internet from mobile phones are studied under the subject **Mobile IP**.

Self Check Exercise

- 1) Given an IP address as 192.26.16.120, determine the class of network and the network address. Assuming that hosts are placed in 8 subnets, determine the subnet address and the host address in the subnet.

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

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

.....

.....

.....

.....

.....

6.6.2 Transmission Control Layer

The main purpose of transmission control layer is to ensure reliable, efficient and cost-effective transfer of user messages from source host to destination host regardless of the underlying physical network or the chain of networks. The transmission control layer of Internet has two main protocols:

- User Datagram Protocol (UDP)
- Transmission Control Protocol (TCP)

In the Internet layer, we discussed IP datagrams. IP datagrams are network datagrams whereas user datagrams are for applications that send out one short message at a time. As one might expect, UDP is used to provide a connectionless service. The datagram is sent out on the network without establishing any prior connection with the destination host. On the other hand, TCP is a connection-oriented protocol with three phases of operation much as in circuit switching: connection establishment, data transfer and connection release phases. It is used with applications that transfer large volume of data such as file transfer application. There is a large number of applications that exchange short requests and short responses. Such applications use UDP for message transport. In this section, we discuss UDP at some length to bring out how applications run on networks. Readers interested in TCP details may refer to Further Reading [Stallings or Tanenbaum] listed at the end of this Unit.

The Transmission Control Layer (TCL) acts as an interface layer between the underlying physical network and the applications run by the user. Several applications are resident on a host and many of them may be running simultaneously. All the applications use the services of the TCL. When an IP datagram is delivered to the TCL by the IP layer at the destination host, the TCL must know as to which application the packet must be delivered to. The process of binding a datagram to an application is done via what are known as **ports**. Every active application is assigned a port number for reception and transmission of messages. The binding of an application with a port number may be permanent or assigned dynamically whenever an application is invoked. Port numbers from zero to 1023 are reserved for standard services. Some permanent port assignments for standard applications are presented in Table 6.1. The source port and the destination port addresses are stored in the TC header.

The structure of the protocol data unit (PDU) of UDP is shown in Figure 6.7. The PDU has an 8-byte header followed by data. The UDP PDU is handed over to IP layer for

Source Port	Destination Port
UDP Length	UDP Error Control
UDP Data	

Fig. 6.7: Structure of UDP PDU

transmission as an IP datagram. The entire UDP PDU becomes the data part of IP datagram. The destination port is used by the TCL to deliver the UDP data to the correct application. The application may use the source port value for sending a response to the source application. The value of the UDP length field specifies the total length of the PDU 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. The UDP does not perform error control at the datagram level. This must be taken care of at the application level. The UDP does not guarantee delivery. If the datagrams are lost, the applications must recover from the consequences. UDP is a simple protocol with very little overhead. This feature of UDP has made it very attractive for applications that do not require a very high level of reliability.

6.6.3 Application Layer

A large number of application protocols have been developed on the Internet. Some of the important ones are listed in Table 6.1 along with the application for which they are used, standard port number where assigned and the TCL protocol used for the application. The file transfer protocol (FTP) is used to send files from one host to another under user

command. The protocol supports transfer of both text and binary files that contain figures, etc. The protocol has provision that allows the user to transmit his ID and password to the remote host. Access permission provisions associated with files and the capability given to a particular user regulate the access to files. The user specifies the file and the file actions required.

TELNET supports remote log in facility. A user may log on to a remote computer via a simple dumb terminal or a personal computer using TELNET. Simple Mail Transfer Protocol (SMTP) provides basic electronic mail transfer facility. SMTP features include mail forwarding, acknowledgement and mailing lists. Multi-purpose Internet Mail Extension protocol (MIME) is an extension to SMTP for supporting multimedia information transfer. MIME operates via SMTP, which in turn uses TCP. HyperText Transfer Protocol (HTTP) is used for accessing World Wide Web (WWW) servers that store web documents. Web documents are prepared and stored using HyperText Mark up Language (HTML) which is a mark-up language for electronic documents.

Table 6.1: Important Application Layer Protocols

Protocol	Application	Port	TCL
FTP	File Transfer	21	TCP
TELNET	Remote Login	23	TCP
SMTP	Electronic Mail Application	25	TCP
HTTP	World Wide Web	80	TCP
SNMP	Network Management	-	UDP
RTP	Real Time Applications	-	UDP
MIME	Multimedia Mail Application	-	SMTP

Simple Network Management Protocol (SNMP) is used for Internet management. It uses simple commands to assess the status of host and monitor the performance of the network. Real time Transport Protocol (RTP) is used for transmitting audio and video in real time. It is used in multimedia applications like video conferencing, music on demand and Internet telephony.

6.7 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 model of interaction and computing underwent significant changes. The terminals were replaced with PCs that were capable of connecting themselves to data networks. The PCs could interact with different computers at different points of time. They could perform certain amount of computing at their end. This situation led to **peer-to-peer interaction model**. Distributed computing became the norm. Slowly, computers, called **servers** that provided different types of services started appearing on the networks. The services were accessible to other computers that were treated as **clients** of the service providing computers. Thus was born 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 6.8. There are two machines and a network in the configuration: a server machine, a client machine and a data network. The server and the client interact via the data network. The server provides a set of information and computational services, which are used by remote clients. In other configurations,

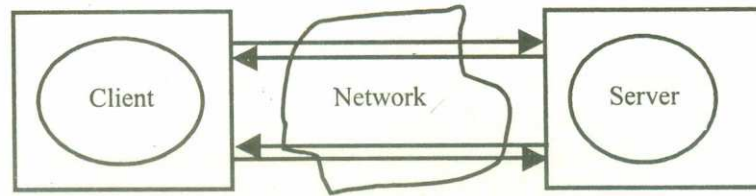


Fig. 6.8: Client-Server configuration

many clients may access one server simultaneously. Client-server interaction may take place at one of the following three levels:

- i) Human – Machine
- ii) Human – Human
- iii) Machine – Machine

The first case is the most popular one with human client and a machine server. A typical example of the first case is that of a user accessing information from a server. An example of the second case is a chat session between two users. Timed periodic file transfer or email transfer between two or more machines are examples of the third case. In all client-server interactions, it is the client that always initiates a session. The server is ready and waiting without doing anything. This is like a shopkeeper ready to sell with his shop open but the actual transaction takes place only when a customer arrives.

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 Intel 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 2000, Novell Netware and Linux. Most of 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 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. Some predict that in future, both PCs and workstations may run Linux instead of Windows or Unix. 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 as simple as a *cellular telephone*, *notebook computers*, *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 a mainframe as the number users rises. Interestingly, the upgradation 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 rolling 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. Two of the well-known browsers are Internet Explorer from Microsoft and Netscape Navigator from Netscape Communications.

Network applications are designed in one of two modes:

- File access mode
- Message passing mode

In the first case, the server only stores program and data files. For processing, program and data (if any) are copied across the network to the client machine and the program execution takes place there. A good example of this mode is the e-mail server. Here, the server only maintains the mailbox of the user. The user downloads the mail and processes the same on his/her machine. In the second mode, the processing load is shared between the server and the client by suitable exchange of messages. This mode is also called *request-response* mode. An example of this design is information retrieval server. In this case, the retrieval software runs on the server and further processing is done on the client.

6.8 DATA NETWORK CLASSIFICATION

Data networks are classified according to their geographical coverage:

- Wide Area Networks (WANs)
- Metropolitan Area Networks (MANs)
- Local Area Networks (LANs)
- Home Networks

Intercity, intercountry and intercontinental networks are known as WANs. Based on the communication infrastructure used, they may be classified as terrestrial data networks (TDNs) or satellite based data networks (SBDNs). TDNs may further be further subdivided as wired networks and wireless networks. In wired TDNs, data communication is organised using cables or fibre optic links, and in wireless TDNs using short-range

radio links. Synchronous optical networks (SONET) designed to operate at high speeds are often used as the communication infrastructure in WANs. SONET is fast replacing coaxial cables that were used as the main high-speed communication infrastructure hitherto. A geostationary or geosynchronous satellite is used for communication in SBDNs.

A metropolitan area network (MAN) interconnects computers within a metropolitan city. The concept of MAN as a data network has evolved from community antenna television (CATV) systems that were originally planned for areas where television signals were weak. A large antenna for the entire community is installed on a high rise building, a tower or a hill, if there is one nearby, for receiving television signals. The signal then is distributed via cables to homes. This, with the advent of satellite broadcasting evolved to cable television networks that exist in most major cities of the world today. When Internet attracted a large audience, cable operators recognised the business potential of offering two-way Internet services to homes via their cable infrastructure. With the addition of what is known as *set-top-box* (STB) to a TV set, the same may be used as a home computer to receive video on demand, access information from Internet and execute small programs on remote servers. The idea is to convert the TV set into a low-end computer. Yet another approach is to use what are called as **cable modems** that connect a computer to TV cables. The broadband capability of TV cables permits carrying voice, data and video simultaneously. Thus, MANs are usually multimedia networks. There have also been attempts to install high-speed optical networks in metropolitan areas, which then could be used to carry multimedia services.

Local area networks are confined to a single building or a group of buildings generally belonging to the same organisation. Examples include interconnected computers in a corporate office or in a university campus. Optical fibres, twisted pair and co-axial cables are used as the communication media for LANs. Fibre optic networks (FONs) are suitable for both LANs and MANs. Wireless as a communication medium in LANs has emerged recently.

LANs, MANs and WANs are generally interconnected in a hierarchical manner to form a global network as shown in Figure 6.9. LANs are often connected directly to WANs,

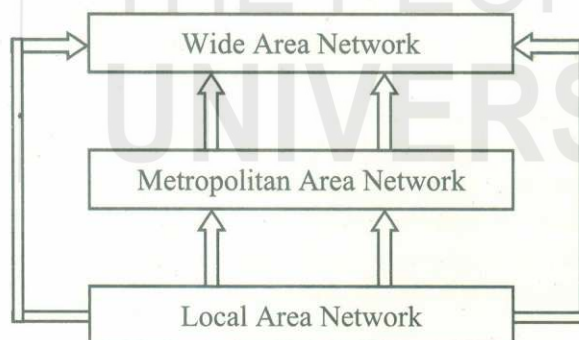


Fig. 6.9: Data network hierarchy

particularly in places where MANs are not installed or have not developed well. At present, most of the LANs directly connect to WANs. Apart from the different geographical coverages, the range of data rates supported on these networks also differs widely.

Home networks are in the offing. The idea behind home networks is very interesting. All electrically operated devices at home can be networked and made accessible on the data network infrastructure. They can then be controlled remotely. For example, a person on way to home can turn on the air conditioner from his car, or a baby left alone at home may be monitored via cameras by parents in an adult party, etc. For this purpose, it is predicted that most homes in future will be built with networking infrastructure. Home networks have begun in a small way already. Most of the home appliances are being designed with networking interface these days. Many homes have burglar alarms installed that alert a nearby police station in case any part of the house is break opened.

- 2) Data networks may be classified according to the communication medium used for networking. Enumerate such a classification.

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

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

.....

.....

.....

.....

.....

6.9 SATELLITE BASED DATA NETWORKS

As pointed out in Unit 4, satellite is inherently a broadcast medium. The geographical region reachable by signals from a satellite is called the *footprint* of the satellite. All the stations within the footprint of the satellite can pick up the signals transmitted by the satellite towards the ground. The cost of transmission via the satellite is the same whether one station or all the stations pick up the signals. Hence, satellites are ideally suited for one-way broadcast applications like television broadcast. Data network applications are inherently two-way, interactive and are individual oriented. Many applications in data networks call for privacy and secrecy, an aspect that is very expensive to implement in satellite networks. Consequently, using satellites for data networking calls for innovative solutions. Satellite data networks differ radically from terrestrial data networks in many ways. Table 6.2 summarises these differences.

Table 6.2: Comparison of Satellite and Terrestrial Networks

Satellite Networks	Terrestrial Networks
1 Star topology	Meshtopology
2 Satellite is a critical component; its failure results in total failure of the network	Node failures do not affect the entire network. The network is fault tolerant.
3 Broadcast in nature	Point-to-point in nature
4 Large distances and hence large time delays are involved, typically 270 ms for one way communication	Distances are relatively small and hence the time delays are small, typically 40 ms for intercontinental connections
5 Microwave communication only	Microwave, cable or optical fibre links
6 Susceptible to weather conditions at higher frequency bands	Cable or optical fibre links are insensitive to weather conditions
7 Satellite life span is typically 7-10 years	Life span is typically 30 years or more

As discussed in Unit 4, communication satellites are placed at an altitude of 36,000 km above the equator. Ground terminals are not necessarily at the equator and are usually on both sides of the equator depending on the geographical location of the country. As a result, the average distance between a ground terminal and the satellite transponder works out to be around 40,000 km. This distance introduces a propagation delay of around 135 ms each in the uplink and the downlink. A **hop**, i.e., signal moving from one station to another, consists of one uplink and one downlink involving a propagation delay of 270 ms. Signal traversing a complete loop is termed *turnaround* and involves a delay of 540 ms (4×135). This delay gets added to the response time that can be critical for many interactive and real-time applications.

The cost of a satellite communication system is appropriated between the space and ground segments. The space segment comprises spacecraft and the equipment on board the spacecraft. The ground segment consists of the earth stations, antennas, ground control systems, user terminals, etc. If the system is to serve a small population of ground terminals as in the case of intercontinental links, less investment can be made in the space segment and more in building large antennas and powerful ground stations. On the other hand, if the population of the terminals to be served is large as in the case of direct broadcast television systems, more investment is to be made in the space segment so that the cost of terminals can be kept low. The present trend is to design more and more of **direct-to-user (DTU)** systems. This requires that we have satellite terminals that are affordable by every home. **Very small aperture terminals (VSATs)** are a step in this direction. Here, the nomenclature 'aperture' is used synonymously with dish diameter. VSATs have dishes with apertures in the range of 1-3 m and are designed to be DTU terminals.

Small DTU terminals can, however, communicate only with large hubs and not directly among them. Direct VSAT-to-VSAT connectivity is not practically feasible. Figure 6.10 shows the communication scheme using a large hub and DTU VSATs. There is a

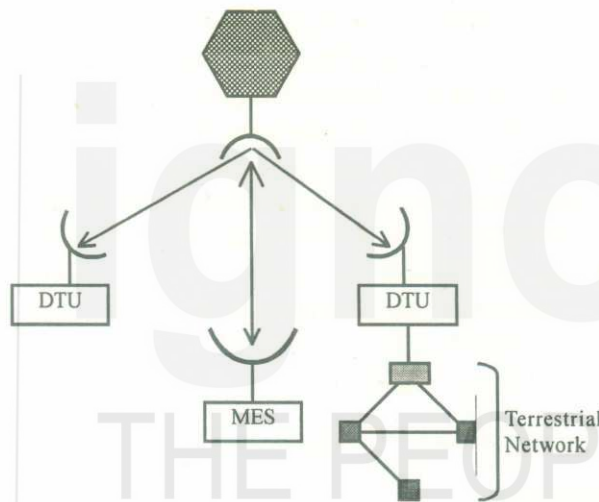


Fig 6.10: Two-hop Satellite Network

master earth station (MES) in the configuration which forms a vital communication link. The antenna of the MES acts as the large hub (typical size 10-30 m) with a high gain. A DTU terminal can communicate only with the MES antenna. Hence, information from one user to another is sent via the MES. The information hops twice before it reaches the destination, once from the source user to MES and the second time from MES to the destination user. Hence, this scheme is known as a **two-hop scheme**. In India, two-hop schemes are adopted in NICNET and RABMN. The main applications of VSAT systems are personal computer connectivity, low speed data communication, voice transmission, and teleconferencing.

In the context of networks and OSI reference model, there are some important aspects of satellite communication systems, which require special consideration:

- 1) Aspects related to the physical layer functions of the OSI reference model: satellite network topology and configuration, modulation schemes and bandwidth utilisation.
- 2) Aspects related to the data link layer: satellite being a common communication medium, simultaneous accessibility by all or a group of earth stations becomes a non-trivial function.
- 3) Aspects related to the network layer: satellite communication being broadcast in nature, routing becomes a trivial function; however, organising point-to-point or point-to-multipoint connections in a broadcast environment needs special consideration.
- 4) Aspects related to the session and transport layers: Hop and turnaround delays have to be accounted for.

A satellite-based network has a 'star' topology with the satellite being the central hub. If the hub fails, the entire communication network comes to standstill. In view of this, reliability and redundancy considerations are important when designing satellites.

A satellite transponder is a common resource to be shared by a large number of ground stations for multiple, simultaneous communications. Techniques are required to access and share the common transponder capacity. Many techniques have been devised and all of them are based on the frequency or time division concepts. When the number of user stations runs into thousands, much larger than the satellite channels available for data transmission, a class of multiple access protocols, known as ALOHA protocols are used for accessing the transponder. These protocols also work well when the number of stations is small. ALOHA protocols originated with radio networks. The scheme was proposed and studied in the early 1970s by Norman Abramson and his colleagues at the University of Hawaii. They were concerned with the problem of giving access to a central computer situated in one of the Hawaiian Islands from a number of terminals located in other islands using a single common radio communication channel. Satellites also present a similar scenario. ALOHA means 'hello' in Hawaiian language.

Satellite data networks have also been attempted by placing a set of satellites in Low Earth Orbits (LEO). These satellites move much faster than the earth's rotational speed. They need to be tracked and are visible only for a short duration from any point on the earth. The idea in using LEO is to place the set of satellites one after another in the same orbit such that when one goes out of sight another appears on the horizon. All the satellites put together provide continuous visibility. Examples of projects with this central idea are Motorola's *Iridium*, *Globalstar* and *Teledesic*. These projects have not been very successful so far.

Self Check Exercise

- 3) A client accesses a server via a two-hop satellite network. The master earth station takes 30 ms to forward a transmission. The server takes 500 ms for processing a particular query from the client. Calculate the response time for this query.

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

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

.....

.....

.....

.....

6.10 LOCAL AREA NETWORKS

A local area network (LAN) typifies a distributed environment and finds applications in a number of areas. Some examples are:

- i) Office automation
- ii) Factory automation
- iii) Distributed computing
- iv) Fire alarm and security systems
- v) Process control
- vi) Document distribution

Of these, the first two applications have developed fully. Architectures for these applications have been standardised. Office automation standard is known as Technical and Office Protocol (TOP), and the factory automation standard is known as Manufacturing Automation Protocol (MAP).

There are three major aspects in a LAN:

- i) Transmission medium
- ii) Topology
- iii) Access method

Four important transmission media are in use: unshielded twisted pair wire (UTP), co-axial cable, fibre optic cable and wireless. Twisted pairs are used in low speed LANs using baseband transmission. In this mode of transmission, data is transmitted as simple electrical levels often without any modulation. There is no multiplexing and the entire bandwidth of the medium is used for transmitting signals from one station. Baseband transmission links are sometimes termed as 'wire only' links as there are no other equipments used for transmission. In contrast, broadband transmission uses modulation techniques and is suitable for transmitting high speed and multiplexed data. Co-axial or UTP cables are used for broadband transmission at speeds of 10, 100 155 and 1000 Mbps. Fibre optic cables carry data at rates in the range of 100 Mbps to a few gigabits per second. Wireless, as a medium is a recent addition to LANs. Wireless LANs have evolved during 1990s.

Three topologies are widely used in LANs: star, bus and ring. These topologies are shown in Fig 6.11. In star topology, the LAN systems are connected to a central switch that establishes connection between pairs of systems. Such LANs are generally designed around electronic PABXs or computerised branch exchanges (CBX). A bus in the bus topology shown in Fig 6.11(b), is a single pair or a bunch of wires that carry the electrical signals. The LAN systems are connected to the bus using a passive tap such that all systems are able to monitor the signals on the bus simultaneously. The bus acts as a

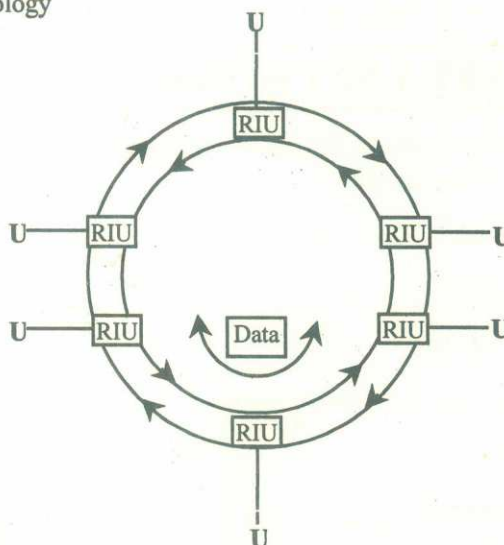
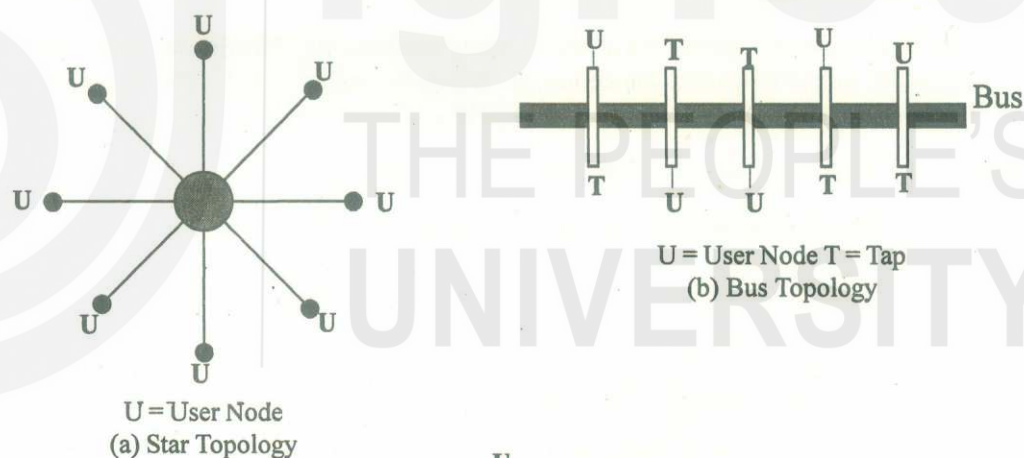


Fig. 6.11: LAN Topology

broadcast medium. In a ring topology, the LAN systems or their interfaces take an active part in the transmission. A ring is formed by a connection of a number of point-to-point

links between pairs of Ring Interface Units (RIU) as shown in Fig 6.11(c). RIU being active device, its failure puts normal ring operation out of order. It becomes necessary to consider fault tolerant techniques such as dual rings or folded ring operation for handling RIU failures in a ring topology. In contrast, being a passive unit, the failure of an interface tap in a bus topology does not affect the operation of the LAN. There are a number of other topologies that have been studied in the context of LANs. These include tree, hypercube, multistage store and forward networks. However, LANs based on these topologies are not widely used at present. Both bus and ring topologies are difficult to maintain in their original form, as the cable will have to be run from room to room and from computer to computer. To make maintenance easier, all computers are connected using UTP cables in a star topology to a device called **hub**. The hub implements a logical ring or bus as required while the physical topology looks like a star.

Three access methods are predominantly used in LANs: switched access, contention or multiple access and token passing access. Switched access is used in LANs that are designed around CBXs. Both contention and token passing access techniques are used in bus topologies and token passing technique in ring topology. There are also other access techniques like reservation access, load adaptive access, and tree structure based access. Discussion of such schemes is beyond the scope of this Unit. The most popularly used LAN is a logical bus with a hub-based physical star topology and contention based multiple access. There are many contention-based protocols. The one most popularly used is known as *carrier sense multiple access with collision detection* (CSMA/CD). The LAN itself is called **Ethernet** LAN.

Wireless LANs have evolved in two directions driven by two societal factors. First, professionals owning notebook computers desired to have access to Internet from different places like airport lounges, railway platforms, restaurants, hotel rooms, etc. As you know, notebook computers are portable and professionals carry these with them wherever they go. Second, mobile owners wanted to connect to Internet and home computer owners desired to connect different devices to their computers in an effortless manner. While the first category meant access over a distance of about 100 meters, the second one was concerned with access over distances not exceeding 10 meters. The technology associated with the second category of LANs is called **Bluetooth**, a name given by the original proposers.

LANs tend to use different address formats for addressing the machines connected to them. For example, Ethernet uses 48-bit addresses for identifying the machines. These addresses are factory set on the **Network Interface Card** (NIC) that connects computers to the hubs in Ethernet or token ring LANs. If an IP packet arriving on a WAN is to be distributed to a computer on a LAN, then we need to go through another level of address translation. The IP addresses have to be converted to LAN machine addresses. This is done by the **Address Resolution Protocol** (ARP) that runs on the Internet layer. ARP accepts IP address as input and returns machine address as output. It broadcasts the IP address on the LAN asking the machine with that IP address to respond with its machine address. A reverse process is carried out by **Reverse ARP** (RARP) protocol to obtain IP address when machine address is known.

Asynchronous transfer mode (ATM) and cell switching discussed in Unit 5 are the only world standards that are applicable to both WANs and LANs. All other standards are either applicable to WAN or LAN. In view of this, great interest has come about in ATM LANs. ATM speeds are unmatched in the conventional LAN environment. The lowest operating speed of ATM is 155 Mbps with ability to transport High Definition Television (HDTV) in real time. The next upgrade in ATM takes a user to 622 Mbps with potential for moving to speeds up to 10 Gbps. No other LAN is capable providing such a support. ATM is poised to become the most preferred telecommunication infrastructure of the future for both WANs and LANs. In order to ensure that existing LAN protocols run on ATM infrastructure, many studies have been undertaken to emulate Ethernet or token ring on ATM. **LAN emulation on ATM** is a topic of current interest.

The responsibility of bringing out standards for LANs is given to Institute of Electrical and Electronics Engineers (IEEE) by ISO. The standards are brought out under series IEEE 802. Some of the important LAN standards are listed in Table 6.3. As mentioned earlier, IEEE 802.3 Ethernet LANs are the most popular ones. There are many variants of Ethernet. IEEE has introduced a concise notation to represent the features of the variants. The notation has three parts: first part specifies the speed of operation, the second the transmission technique and the last one some physical aspect of the medium used. Examples of Ethernet variants in concise notation are 10BaseT, 100BaseT and 100BaseF. The first of these examples means 10 Mbps LAN with baseband transmission using twisted pair wires connected in star configuration to a hub. In the last example, the letter 'F' stands for optical fibre. Unfortunately, the notation itself has undergone many changes leading to confusion in interpretation. Currently, a variant of Ethernet LAN operates at a maximum speed of 1000 Mbps. This is usually referred to as **Gigabit Ethernet**.

Table 6.3: Some LAN Standards

Number	Description
802.1	Overview and Architecture of LAN
802.3	CSMA/CD Ethernet
802.4	Token Bus LAN
802.5	Token Ring LAN
802.6	Metropolitan Area Network (MAN)
802.11	Wireless LANs
802.15	Personal Area Networks (Bluetooth)
802.16	Broadband Wireless

IEEE 802.4 LANs, promoted by General Motors are used in manufacturing automation. MAP protocol suite is built around token passing bus architecture. IEEE 802.5 LANs, promoted by IBM are mainly used in IBM installations. Both IEEE 802.4 and 802.5 have not found popular appeal. However, token ring ideas are applied in evolving a fibre optic LAN/MAN standard known as Fibre Distributed Data Interface (FDDI) which is described as part of IEEE 802.6. Currently, wireless LANs are being pursued with great interest. High-speed wireless LANs have also been receiving attention. Issues related to high-speed wireless LANs are studied under IEEE 802.16 that is titled as broadband wireless.

Self Check Exercise

4) Discuss the features of a 1000BaseT Ethernet LAN.

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

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

.....

.....

.....

.....

.....

.....

6.11 PROTOCOLS FOR LIBRARIES

Library functions are a potential application area for data networks. Library functions differ from office functions or factory requirements in many ways. Application level protocols are required to support network-based library applications. Two protocol and service specifications have been finalised by ISO for libraries as shown in Table 6.4.

Table 6.4: Library Protocol Standards

IS 10160	Inter-Library Loan (ILL) Services Definition
IS 10161	Inter-Library Loan (ILL) Protocol Specifications
IS 10162	Search and Retrieve (SR) Services Specifications
IS 10163	Search and Retrieve (SR) Protocol Specifications

Standards for other library functions need to be worked out. A standard protocol stack for library automation is highly desired. This is even more important in the context of emerging **digital and virtual libraries**. Currently, libraries are being automated without following any standard architecture. Such an approach may pose serious problems in the long run. Library functions are non-time-critical in nature. Hence, a protocol stack may be modelled along the lines of TCP.

Self Check Exercise

- 5) Recommend a protocol architecture for library automation along the line of Internet protocol suite. Assume gigabit Ethernet as the underlying network.

Note: i) Write your answer in the space given below.
ii) Check your answer with the answers given at the end of this Unit.

.....

.....

.....

.....

.....

.....

6.12 INTERNETWORKING

The internetworking problem in data networks can be solved elegantly by considering the protocol architectures of the networks to be interconnected. We now discuss the principles involved in internetworking. Protocol suites for specific applications may be formed by using protocols from different standards. For example, an application may use Internet protocols in layers 3-5 with IEEE 802.3 LAN or X.25 network as the underlying network. A LAN may run SNA protocols in layers 4-7. It is the commonality of protocols in the upper layers, which decides on the schemes to be used for internetworking. An internetworking unit resolves the protocol differences in lower layers. An internetworking unit is termed as **a repeater, a bridge, a router or a gateway**, depending on the level at which internetworking is achieved, as shown in the Figure 6.12.

First let us consider two networks which use identical protocols in all the layers. These may be interconnected using repeaters at the physical layer level as shown in Figure 6.12(a). As you know, repeater boosts the signal level and transmits bits from one network to another making the two networks look like one single logical network. This kind of connection is useful in large campuses where a logical LAN is realised using multiple physical LANs. All the physical LANs must use the same protocol structure and the same physical media. Two networks which differ in physical layers and the corresponding

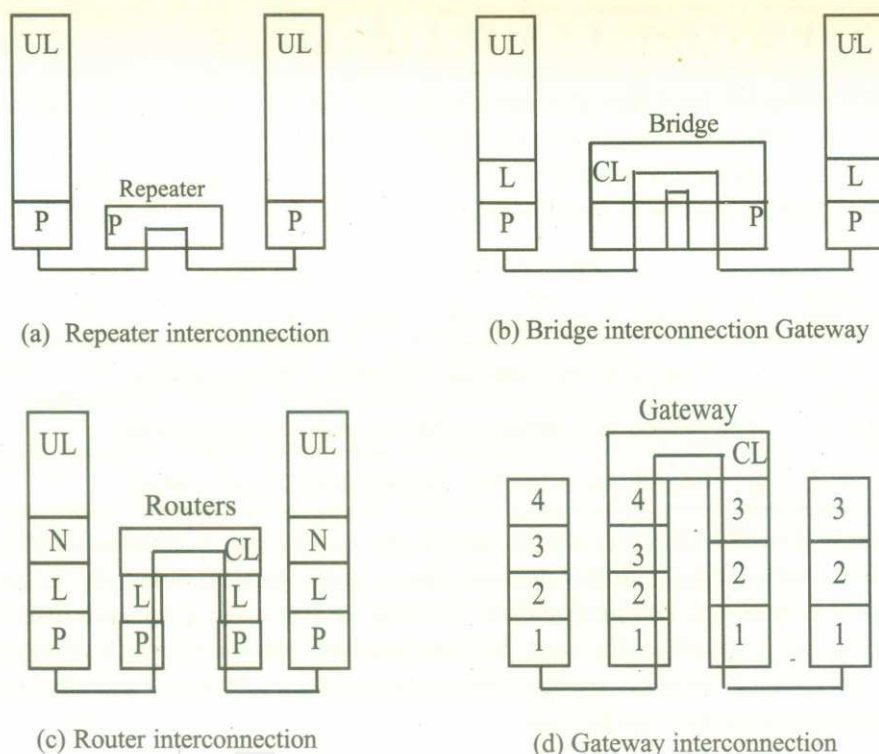


Fig. 6.12: Internetworking structures

media access but not in logical aspects of data link layer and have common protocols in the upper layers are interconnected using a bridge as shown in Figure 6.12(b). A bridge is an intelligent device capable of recognising addresses and transmitting only the internetworking frames across the networks. Bridges may also be used in interconnecting identical networks from the point of view of traffic management. Routers are used as shown in Figure 6.12(c) when the two networks to be interconnected use the same protocols in layer 3 and above. Example of router-based interconnection is a WAN with a LAN. The whole of Internet is based on the use of routers. An internetworking unit which implements a combination of bridge and router functions, sometimes called *brouter*, may be used to interconnect two LANs via a WAN. Networks with different communication architectures use a gateway as shown in Figure 6.12(d). Networks that differ in some nonadjacent layers may be interconnected using suitable gateway structures. In the context of internetworking, some routing and gateway protocols are useful. They are *Border Gateway Protocol* (BGP) and *Exterior Gateway Protocol* (EGP) and *Open Shortest Path First* (OSPF) protocol.

6.13 NETWORK MANAGEMENT

The complexity of telecommunication 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. In addition, 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.

6.14 SUMMARY

Starting with remote job entry and interactive computing in the 1960s, data networks have grown to the level of network of networks called Internet. Data networks, in whatever form, transfer data in the form of structured data units. This Unit starts with a discussion of the concept of data units. The principles of layered communication architecture are then expounded. This is followed by a discussion on ISO-OSI reference model. Keeping the reference model in view, the Internet architecture is then presented. The protocols of the three important layers of Internet, viz. Internet or IP layer, transmission control or TCP layer and application layer are discussed. The structure of IP addresses and IP datagram are explained and the concept of Protocol Data Units (PDU) is placed in perspective. Most of the applications on the Internet run under a computational model called client-server architecture. This architecture is discussed following the Internet protocols. Conventionally, data networks are classified according to their geographical coverage. This classification scheme is then presented followed by a description of Satellite Based Data Networks (SBDN) and local area networks. Metropolitan Area Networks (MANs) and home networks are touched upon. Unlike in terrestrial networks, data communication via satellites poses some problems requiring special consideration. These

problems and the corresponding solutions are briefly discussed. There has been a significant development in the field of LANs during the 1990s. In particular, wireless LANs and Bluetooth technology have emerged in a significant way. A large number of variants of Ethernet have emerged. Interest has been generated in broadband wireless LANs. ATM being the only standard applicable both to WAN and LAN, realisation of ATM LANs and interworking of existing LANs with ATM LANs are being studied. LAN technologies including the recent developments have been described. Rendering of library services on the network calls for protocols related to library functions. The current status of protocol standardisation in the area of libraries and the future needs are projected. Internetworking is at the heart of Internet. The techniques for internetworking are placed in perspective with the help of the principles of layered architecture. Finally the Unit closes with a brief discussion on network management.

6.15 ANSWERS TO SELF CHECK EXERCISES

- 1) First we write the equivalent binary pattern for the IP address.
The pattern is: 11000000 00011010 00010000 01111000.
The first three bits 110 indicate that this is a Class C address. Class C address has 21 bits as network address which is: 00000 00011010 00010000
The equivalent decimal value is: 6672
There are 8 subnets, which require three bits for addressing them. Last 8 bits are divided as 3 bits for subnet address and 5 bits for host address. Therefore, the subnet address is 3 and the host address is 24.
- 2) We know that there are four modern communication systems using four different media. Accordingly, we may have four types of data networks: Wired data networks, Wireless data networks, Satellite data networks and Fibre optic data networks.
- 3) In a two-hop network turnaround involves 4 hops. This introduces a propagation delay of 1080 ms.
The signal traverses the MES twice introducing an additional delay of 60 ms.
The server takes 500 ms.
Therefore, the response time is $500 + 60 + 1080 = 1640 \text{ ms} = 1.64 \text{ seconds}$
- 4) 1000BaseT. Operating speed 1000 Mbps or 1 Gbps.
Transmission technique is baseband.
Uses unshielded twisted pair (UTP) cable of category 6 that is defined for 1 Gbps operation. (See Unit 4).
Computers are connected in a star fashion to a hub.
Media access technique is carrier sense multiple access with collision detection (CSMA/CD).
- 5) A 5-Layer architecture is suggested.
Layer-1 Cat 6 UTP cables. Star configuration using a bus hub.
Layer-2 CSMA/CD media access control. IEEE 802.3
Layer-3 Internet Protocol (IP)
Layer-4 TCP and UDP
Layer-5 ILL, SR and protocols for union catalogues, electronic document transfer, etc.

Draw up an architectural diagram similar to Figure 6.4.

6.16 KEYWORDS

Address resolution	: Conversion of one address to another like IP address to machine address and vice versa.
Baseband transmission	: A technique in which bits are transmitted without any modulation.
Cable modem	: A modem used for connecting home computer to television cable.
Classed IP address	: Structured IP address with three classes A, B and C.
Classless IP address	: Unstructured IP address with no class definition and of variable length.
Client	: A human being or a computer that accesses the services of a server.
Datagram	: A unit of data usually transmitted in an independent manner.
End-to-end	: A connection between peer entities in two end systems that run applications.
Entities	: Hardware or software modules that implement functions of a layer.
Flow control	: A technique for regulating the speed of despatch of data units matching the capabilities of the source and the destination computers.
Hop	: Traversal of a signal from a ground station to a satellite and back to another ground station.
Host	: Any computer connected to Internet.
Hub	: In satellite communication, hub is the master earth station. In LANs, hub is a device to which computers are connected in a star fashion.
Internet protocol	: The protocol that runs on the Internet layer and is used for delivering datagrams on best of efforts basis.
Internetworking	: Connecting two or more networks.
IP address	: Unique address assigned to a host on Internet.
IP datagram	: Data unit transported by IP protocol.
Layered architecture	: Communication architecture in which functions are performed in a stacked manner.
Link-to-link	: Connection between peer entities in adjacent systems.
Low earth orbit	: Orbits at around 1000-3000 km altitude.
Multiple access	: Access to a shared medium by a number of stations.
Open system	: A system that follows ISO standards.
Peer entities	: Entities at the same level on different systems.
Port	: An interface point between to adjacent layers.
Port binding	: Allocation of a port number to an application.
Protocol	: Procedure for exchange of data units.
Protocol data unit	: Data unit carrying both protocol control information and user data.

Protocol Suite	: A set of protocols in all layers of a communication architecture selected for a particular application area.
Scalability	: Ability to increase the capacity of a system without affecting the operations.
Server	: A system that provides a set of specified services.
Set top-box	: A unit that allows a TV set to function as a low-end computer.
Turn around delay	: Time taken for signal to traverse from one ground station to another and back via satellite.
Transmission control protocol	: A connection oriented protocol used in large data transfers.
User datagram	: A data unit transmitted independently at the user level.

6.17 REFERENCES AND FURTHER READING

Black, U. (1999). *Computer networks: protocols, standards and interfaces*. 2nd Edition. New Delhi: Prentice Hall of India.

Hardy, J. (1999). *Inside networks*. New Delhi: Prentice Hall of India.

Mansfield, K. C. and Antanakos, J. L. (2002). *An introduction to computer networking*. New Delhi: Prentice Hall of India.

Martin, James. (1998). *Telecommunications and the computer*. 3rd Edition. New Delhi: Prentice Hall of India.

Panko, R. R. (2002). *Business data networks and telecommunications*. 4th Edition. New Delhi: Prentice Hall of India.

Sharda, N. K. (2002). *Multimedia information networking*. New Delhi: Prentice Hall of India

Stallings, W. (2000). *Data and computer communications*. 5th Edition. New Delhi: Prentice Hall of India.

Tanenbaum, A. S. (2002). *Computer networks*. 4th Edition. New Delhi: Prentice Hall of India.

Viswanathan, Thiagarajan. (2004). *Telecommunications switching systems and networks*. New Delhi: Prentice Hall of India.

UNIT 7 CONVERGENCE TECHNOLOGY AND APPLICATIONS

Structure

- 7.0 Objectives
- 7.1 Introduction
- 7.2 Goal of Convergence
- 7.3 Convergence Objectives
- 7.4 Basic Requirements of Convergence
- 7.5 Genesis of Convergence
- 7.6 Driving Factors
- 7.7 Convergence Focus
- 7.8 Convergence Architecture
- 7.9 Sectoral Convergence
- 7.10 Technology Convergence
- 7.11 Techniques Convergence
- 7.12 Networks Convergence
- 7.13 Access Convergence
- 7.14 Service Convergence
- 7.15 Applications Convergence
- 7.16 Summary
- 7.17 Answers to Self Check Exercises
- 7.18 Keywords
- 7.19 References and Further Reading

7.0 OBJECTIVES

Different technologies have developed over the years for communication of information. We have discussed them in the previous Units. Such technologies are merged to develop an integrated technology towards achieving the good of networked society.

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

- the meaning of convergence;
- convergence as a phenomenon and its goal;
- important convergence objectives;
- basic requirements for convergence to take place;

- organisations responsible for convergence standardisation;
- the origin and progress of convergence;
- the factors that drive convergence phenomenon;
- connectivity, capacity and content as the main focus of convergence;
- layered convergence architecture;
- convergence in different layers of the architecture;
- sectoral convergence;
- technology and techniques convergence;
- network, access and services convergence; and
- convergence applications like Internet radio and Interactive TV.

7.1 INTRODUCTION

In Unit 4, you learnt about four different types of modern telecommunication systems. Digital communication is the trend in all these communication systems. In Unit 5, We discussed the basics of networking, different switching techniques and the major classes of telecommunication networks. Unit 6 was devoted to computer networks. In this Unit, you will learn about how these communication systems and networks serve the human society. Their usefulness to the society increases significantly when there is a synergy of purpose amongst different technologies and techniques. In other words, the use of technologies, the design of networks and the development of services must all be directed towards a common goal. Such a common goal is the evolution of mankind towards *Networked Electronic Information Society* (NEIS). The process of bringing together different technologies, techniques, networks and services to realise the common objective of evolving towards NEIS is termed as **convergence**.

As you know, a 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. Activities that are carried out in the electronic domain using networks are usually denoted with a prefix *e-* such as e-banking. We now list some of the e-activities that are evolving in the networked society:

- | | |
|---------------|-------------------|
| • e-education | • e-governance |
| • e-library | • e-health |
| • e-banking | • e-commerce |
| • e-mail | • e-entertainment |
| • e-journal | • e-newspaper |
| • e-marketing | • e-procurement |

To illustrate what e-activities mean, we elaborate two of the above activities. E-education means sign up with a university, pay fees, obtain lessons, submit assignments give

examination and obtain results, all via network without having to visit the university actually. In India, IGNOU is planning to offer e-education in a big way. With an e-library, its resources and services can be accessed on the network without having to physically go to the library. For this purpose, it is necessary to store all the library resources in digital form. A library whose resources are available in digital form is known as **digital library**. When accessed over networks from remote places, a digital library becomes an **e-library**. An older terminology for e-activities is *online-activities*. This terminology is still in vogue as in online-ticketing, online-fund-transfer etc. Convergence in different areas is one of the basic requirements to make e-activities possible in a networked society.

7.2 GOAL OF CONVERGENCE

Convergence is not limited to technologies. Convergence, *per se*, is a phenomenon covering a wide range of areas such as technology, applications, media, network systems, network access and switching techniques. Convergence is not limited to communications and networks. It is much wider phenomenon covering many technologies and even human resource. For example, convergence is much talked about in automobile industry. There it refers to the use of electronics, communications and computers in automotive controls and operations. For example, use of radar to sense vehicles or other obstacles ahead on the road and automatically adjusting the speed is possible only when the relevant technologies converge. Another interesting example is the use of intelligence in air conditioners. Conventional air conditioners maintain the room ambience according to certain set parameters. Air conditioners with intelligence determine the ideal room ambience and maintain the same. They take into account factors like human biological rhythm, the present heat load, atmospheric temperature etc. in determining the ideal ambience. Artificial intelligence, biological science, instrumentation and computer engineering are some of the disciplines that contribute to producing such an air conditioner. There is little doubt that the world is witnessing a phenomenon of convergence in almost all the fields.

In this Unit, we limit our discussions to convergence in the context of networked society. The degree of success of convergence phenomenon would determine the speed with which mankind would move towards NEIS. In the context of NEIS, the goal of convergence may be stated as:

To achieve the capability wherein any network can deliver any service to any platform, and any user with an appropriate consumer device can access any application that runs on the networked world.

In some sense, the goal as stated above is reflection of the famous idiom "*Unity in Diversity*" in the context of NEIS. Multiplicity of networks and computer platforms would make up the basic infrastructure of the networked world. The networks and platforms would be designed by using a variety of different technologies. There would be a multiplicity of applications and a multiplicity of consumer devices. In the midst of all these multiplicity, there exists the unity of purpose that any user connected to any network is given the capability for accessing any application that runs on any platform anywhere in the networked world. Of course, the user must have the appropriate consumer device that is required by an application. Obviously, a user can not expect to access an application that delivers pictures by using a plain telephone set. Currently, efforts are on to design and develop integrated consumer devices that are capable of transmitting and receiving audio, video and data. Such devices would enable the users to access multimedia applications on the global network. Audio, video and data are considered as three generic forms of information in NEIS. Data includes all computer generated information such as text, computer graphics, computer animation, digitised pictures, etc.

Self Check Exercises

- 1) State and discuss the goal of convergence phenomenon in the context of NEIS.
- 2) What do you understand by an integrated consumer device and how is it useful in NEIS?

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

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

.....

.....

.....

.....

7.3 CONVERGENCE OBJECTIVES

Having discussed the goal of convergence in the context of NEIS, we are now in a position to enumerate a set of design objectives for convergence. The broad design objectives may be considered to be two folds:

- 1) To make every network and system be capable of handling the three generic forms of information, viz. audio, video and data.
- 2) To enable every user access information in one or more of the generic forms.

In designing convergence systems a number specific objectives must be kept in mind. They include:

- | | |
|---------------------|-----------------------------|
| • User-friendliness | • Portability |
| • Seamless | • Reliability |
| • Secure | • Personal utility |
| • Privacy | • Professional productivity |

The first and foremost of design objectives of convergence is user-friendliness. It must be simple and easy for an end user to access the networked world and work with any application. There shall not be a multitude of interfaces at the user premises. Elaborate connection procedures must be avoided. The procedure to establish a connection to an application must be simple and brief. The access must be seamless across networks without the user being aware of the intermediate networks that may be involved in establishing a connection. There shall be no restriction placed on the user on account of network limitations. The user transaction must be secure on the network. User should not lose information or accidentally access information that belongs to someone else. If a user so desires, it must be possible to assure complete privacy for his transactions. People shift homes for a variety of reasons. A family's network-based activities should not be affected because of a shift in the physical location. Portability considerations address such issues. Reliability is a well-known consideration in networks. Failure of network segments or components should not affect user activities. Both personal utility and professional productivity must be enhanced on account of network based life-style.

7.4 BASIC REQUIREMENTS OF CONVERGENCE

Fundamental to the process of convergence is the use of digital technology and digital techniques all over: digital computers, digital communication, digital information and so on. Digital technology offers substantially higher information management capability than analog technology. As explained in Unit 12 of MLI-101, information management involves

seven different functions: *generation, acquisition, storage, retrieval, processing, transmission* and *distribution*. Use of digital technology in every one of these functions results in superior performance. Digital infrastructure allows a variety of services to be integrated and supported on a common platform, as we saw in Unit 5 while discussing ISDN. You may recall that digital information is multimedia in nature, i.e. different forms of information such as text, audio and video are constituents of digital information. You learnt about digital information in Unit 8 of MLI-101. Digital technology ensures that any form of information can be transported and delivered in NEIS. Thus, digital technology is an essential ingredient of convergence.

Interworking and *interoperability* are two basic requirements of convergence. The two terms broadly mean the same thing; i.e. making any two non-identical entities co-operate and work together. The term *interworking* is used at the level of networks and systems. For example, how do we make an ATM network work with an Ethernet? How do the two operating systems, Linux and Windows co-operate with each other? Such are the issues addressed by interworking. The term *interoperability* is used at the level of devices and protocols. For example, how do we make an IP router work with a X.25 router? How do we make IPv4 protocol work with IPv6 protocol? Such are the issues addressed by interoperability.

Standardisation is another basic requirement of convergence. It plays an important role in speeding up the process of convergence. Two international bodies under United Nations set out Standards related to NEIS:

- 1) International Telecommunications Union (ITU)
- 2) International Standards Organisation (ISO)

In addition to these two bodies, Internet related Standards are brought out by *Internet Society*. To a large extent, these bodies complement each other's work by concentrating on certain segments of standardisation. All the three bodies freely adopt one another's standard to enable users to follow any one set of standards. Each of these bodies has its own well laid-out procedure for evolving standards.

ITU is located in Geneva, Switzerland with over 150 members. Any sovereign state that is a member of the United Nations may become a member of ITU. Sovereign states are represented by their respective governments and thus ITU is an inter-governmental organisation. Member governments, mostly represented by their respective telecommunications administration, are constituent members with a right to vote. Others, like network and service providers, manufacturers and scientific and industrial organisations also participate in ITU activities but with a lower legal status.

There are two sectors of ITU that deal with standards related to NEIS and convergence:

- 1) Telecommunications standardisation sector (ITU-T)
- 2) Radio communications sector (ITU-R)

ITU-T is responsible for standardisation of telecommunications including network related radio communication. ITU-R is responsible for the management of common and scarce natural resource, the air waves or the frequency spectrum and regulation of telecommunications. ITU-T is the successor of the Comité Consultatif Internationale de Télégraphique et Téléphonique (CCITT), i.e. International Consultative Committee for Telegraph and Telephone. ITU-R now includes the activities of the former Comité Consultatif de Internationale Radio (CCIR) and the International Frequency Registration Board (IFRB). Also part of ITU-T structure is the Telecommunication Standardisation Bureau (TSB) which is responsible for organising meetings, providing support services and ensuring smooth and efficient functioning of the ITU-T sector. ITU-T is supported by an advisory group called the Telecommunications Standardisation Advisory Group (TSAG).

ISO is a voluntary non-treaty organisation with national standards organisations, like American National Standards Institute (ANSI), British Standards Institution (BSI), Association Francaise de Normalisation (AFNOR), Deutsches Institut für Normalische

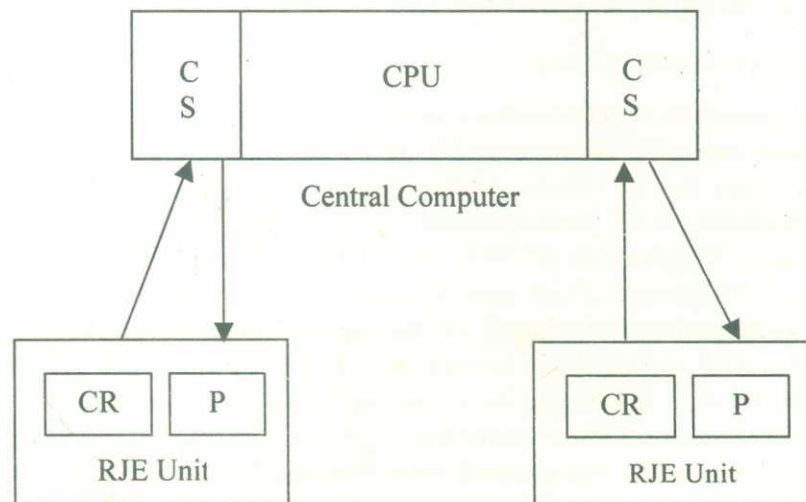
(DIN) and Bureau of Indian Standards (BIS) as its members. The Institute of Electrical and Electronic Engineers (IEEE), the largest professional organisation in the world also plays a major role in evolving local area network and software engineering standards under the aegis of ANSI.

The birth and initial growth of Internet was largely due to the support it received from U.S. government organisations. In 1992, Internet moved away from its U.S. government roots and a society called *Internet Society* was formed to manage and direct Internet activities. At the apex level of the Internet Society, there is the Internet Architecture Board (IAB), originally known as Internet Activities Board. IAB gives focus to and co-ordinates the Internet activities and guides the evolution of the Internet. IAB functions through two steering groups: research steering group (IRSG) and engineering steering group (IESG). Research activities are carried out by a number of research groups under IRSG. Engineering activities include standardisation.

Convergence is not a sudden phenomenon, although it is in the last about 10 years that the issues of convergence are being addressed in a structured manner. In the next section, we see as to how the phenomenon of convergence started about 50 years ago.

7.5 GENESIS OF CONVERGENCE

The genesis of convergence can be traced to mid 1950s. At that time, the computers were expensive resources and required highly controlled environment for their operations. They were housed in big computer centres that were air-conditioned with temperature, humidity and dust control. Entry was restricted to these centres and people were required to remove their shoes and wear aprons to work inside the centres. Computer centres were few in numbers and often located in far away places. Users had to travel long distances to reach the centres. India had only three or four computers in the first half of 1960s located in the major cities like Kanpur, Delhi, Kolkata (the then Calcutta) and Mumbai (the then Bombay). In effect, working with computers was difficult for prospective users. Around the same time, computer manufacturers were keen on increasing the clientele for their computers for economic reasons. Being expensive resources, high usage was essential to justify the cost associated with the establishment of computer centres. There is a famous saying which goes thus: *If Prophet Mohammad can not go to the mountain, the mountain comes to him*. Computer manufacturers took an approach reflected by this saying in solving the problem of low usage of computers. They decided to take the access to computers to the doorsteps of users. The approach is known as **remote computing**, i.e. the users could access a computer and perform computations from a remote location. Implementation of remote computing required a suitable communication link between the remote site and the computer centre as shown in Figure 7.1. In those days, computer programs were used to be prepared on cards by punching



CS = Communication subsystem CR = Card reader P = Printer

Fig. 7.1: Remote computing setup

holes and submitted to a computer centre for execution. A program thus submitted was called a *job*. Accordingly, in remote computing, the equipment at the remote site was called *Remote Job Entry* (RJE) terminal. This usually consisted of a card reader as an input device and a printer as output device suitably interlinked to the main central computer via a dedicated communication line. Both card reader and printer were found to be slow when compared to the communication line and the central computer. Hence, later RJE centres used a tape drive to improve the utilisation of communication links and the central computer. The high-speed tape drive acted as an interface between the central computer and the slow I/O devices in the RJE centre. At that stage, a small CPU was required at the RJE centre to control I/O devices and the tape unit. Remote computing is the first application that brought computer and communication technologies together laying the foundation for technology convergence. Thus, the genesis of convergence lies in remote computing that started in late 1950s.

Once remote computing was successful, there was no looking back in the combined use of computer and communication technologies. Based on remote computing concepts, four important developments took place in the 1960s: **multi-programming** and **time-sharing operating systems**, **interactive computing** and **computer networking**.

RJE centres were offline as far as the users were concerned. The users submitted their jobs on cards and their outputs became available many hours later. But, RJE centres were online as far as the main computer centre was concerned. The input data from the card reader or the magnetic tape reached the computer directly and the output from the computer reached the tape or the printer directly. In a sense, the RJE centres were non-interactive online centres. The RJE centres demonstrated the successful remote online connectivity.

Multi-programming operating systems came about to support multiple RJE centres simultaneously. While one program was busy transferring I/O to a RJE centre, another job was allowed to use the CPU. In other words, multiple programs were active simultaneously. The concept of multi-programming combined with online connectivity gave birth to interactive computing under time-sharing operating systems. In interactive computing, the users interact directly with the main computer. I/O is done via visual display units (VDU) with keyboard acting as the input device and the screen as the output device, a configuration that we are very familiar with these days. The VDUs are connected online to the computer. Multi-programming OS was modified to time-sharing OS to support multiple users online. In multi-programming, a job was allowed to execute on the CPU until it is halted for I/O. Other jobs waited until one job was done with the CPU. Such a policy is not suitable for the environment where many users are directly connected to the computer. In time-sharing systems, each user is given a fixed time slice of the CPU time so that all users have equal access to the CPU. Initially, time sharing terminals were placed in the main computer centre. It was just a matter of time that a cluster of VDUs was placed in RJE centres, laying the foundation for network computing.

The idea of sharing data, information and other resources emerged in the late 1960s. The focus shifted from communication between terminal and computer to communication between computers, which marked the emergence of **computer networks**. In fact, the ideas developed in remote computing form the basis of access to Internet even today. The convergence of computer and communication technologies took deep roots by the end of 1960s. This was most aptly summarised in the words of a famous information scientist R. M. Fano as quoted in the following:

The marriage of computers and communications has been celebrated and consummated. The honeymoon is over and the two partners are realising the hard realities of inter-dependence on each other. (R.M. Fano, 1972).

After 33 years of this marriage, the bond is growing from strength to strength and today we are witnessing the phenomenon of convergence in a variety of areas to realise the common goal of *networked society*. We present an overview of the different areas of convergence in Section 7.7. But, prior to this, we shall examine the factors that drive the concepts of convergence in the next section.

Self Check Exercises

- 3) Give a diagrammatic presentation of the linkage between a main computer centre and a RJE centre that consists of a card reader, printer and a tape drive.
- 4) Distinguish between multi-programming and time-sharing operating systems.

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

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

.....

.....

.....

.....

.....

.....

7.6 DRIVING FACTORS

The factors that are responsible for the idea of convergence being pursued are the same as the motivating factors that we discussed in the context of ISDN in Unit 5. In fact, ISDN represents the first major development in service convergence. Combining services of different types on a single network has been the primary goal of ISDN. Since we have discussed the ISDN motivating factors in detail in Unit 5, we only briefly touch upon convergence driving factors in this section. As in the case of ISDN, there are three factors that drive efforts towards convergence:

- Societal needs
- Economic necessities
- Technological advancements

RJE based remote computing gave serious computer users an easy access to the computing resources. Early computer users were predominantly engineers and physicists who needed computers to solve large-scale problems. Interactive computing set-up in university computer centres generated interest in departments like medicine, arts and linguistics to use computers. Remote interactive computing brought common man to computers. The taste of computing power at residences of people generated a lot of interest in the use of computers. The societal demand for networked computing started increasing day by day. In Section 5.6.2, we listed a variety of new applications that have emerged since the introduction of ISDN. In fact, these are services that have emerged on account of societal demands. Music and video downloads via networks (mobile, landline and Internet) are among the most popular applications at present.

With the growing public demand for new services, the service providers are faced with the problem of offering a variety of services at affordable prices. This meant search for economic solutions, better utilisation of infrastructure and adoption of innovative techniques. Convergence offers a solution to the economic provision of multiple services on a common technology platform.

Demand for new services and economic offering of such services can be met only if there is a corresponding technological advancement. In this context, the most significant advancement is the emergence of digital technology that has revolutionised computers and communications. As mentioned earlier, digital technology forms the basis of convergence.

The services envisaged for the NEIS may be placed under two categories: **distributive** and **interactive**. Distributive services are broadcast in nature like television and commercial radio. For the user, the choice is limited to one of the programmes being broadcast. In the case of interactive services, the end user has a facility to interact with the service provider, say a computer, and obtain services of his choice. For example, the user may ask for a specific movie to be shown on his terminal. Such services are termed *on-demand services*. They call for transmission of high fidelity voice, high quality video and other forms of information on interactive basis. For providing on-demand services, we need communication links with high capacity. The current trend is towards more and more interactive applications.

7.7 CONVERGENCE FOCUS

Quite often, a question arises as to what should be the focus of convergence efforts? There are three important aspects towards which convergence efforts may be directed:

- Connectivity
- Capacity
- Content

At present, less than one-sixth of the world population is connected to telecommunication infrastructure for some service or the other. Over five billion people in the world still have no network connectivity. The existing connection may be for a simple service like telephone, dial-up Internet access or a multitude of services offered by ISDN. Should the convergence efforts be directed towards increasing the connectivity level and thereby make a larger percentage of the world population as part of NEIS? Bringing in more people on the network has been considered important by major international organisations. In fact, ITU has suggested that every village in the world be brought on to the connectivity map by the end of year 2015 as part of its global multi-stakeholder initiative called *Connect the World* launched in June 2005.

As mentioned earlier, interactive services are likely to be dominating NEIS. Interactive services, restricted to text transfer like e-mail, information access and simple graphics transfer can be supported by limited bandwidth connections that exist today. The current information transfer rates are in the range of 64–128 kbps. The rates are proving to be inadequate even for the current level of usage. For example, Internet response for Web access at times becomes so slow that it made someone to expand the acronym WWW as *World Wide Wait!* Multimedia interactive services envisaged in NEIS demand transmission and distribution of high fidelity voice, high quality video, 3-D graphics and other forms of information. This calls for a broadband communication pipe supporting data rates of the order of 100 Mbps or more to be extended to customer premises. Broadband services have the potential of increasing the revenue earnings for the network operators as the users may be prepared to pay high charges for the services. Should then the convergence efforts be directed towards enhancing the capacity of the existing communication pipes instead of increasing the connectivity?

A network infrastructure without contents and applications to run on it is like 5-star hotels without guests to occupy them. The importance of content creation has been emphasised by many experts. For example, a CEO of a networking company once remarked something to the following effect “We do not want to build dumb pipes. If we make only racks and servers, that is dumb. What we should do is to meld contents and networks” Norio Ohga, an ex-chairman of Sony once remarked “Without content, network is nothing”. Yet the debate of content versus connectivity is hot. One of the main reasons for this is that the statistics show that consistently connectivity applications earn much more revenue than content delivery systems. In connectivity applications like telephone, the content is created by the end users. The highest revenue is earned from voice conversation in both landline and mobile networks. The other source of major revenue is the fax usage, which is again a connection-oriented service. Another pointer is a comparison between Small Messaging

System (SMS) and Wireless Access Protocol (WAP) on the mobile networks. SMS is connectivity oriented where the content is created by the end user. WAP is a content delivery protocol for wireless devices. Statistics show that SMS is more widely used than WAP and brings in more revenue. On the Internet too, it is the e-mail which is the 'killer' application and not the wide area information system (WAIS). In fact, it appears that even historically the connectivity has been more important for people. For example, in the postal system the main revenue earner is the letter communication rather than the newspaper distribution. It appears that the people prefer to create their own content and all that they need is an efficient connectivity infrastructure that is affordable.

So, the debate goes on. In the opinion of this author, all the three aspects, viz. connectivity, capacity and content are important. Small and medium network operators must concentrate on connectivity. Large operators must focus on capacity. Content creation must be left to end users and specialists like movie producers.

Self Check Exercises

- 5) Distinguish between connectivity oriented service and content delivery service.
- 6) What are the arguments in favour of increasing connectivity instead of offering contents?

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

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

.....

.....

.....

.....

.....

.....

7.8 CONVERGENCE ARCHITECTURE

The phenomenon of convergence in the context of networked society is occurring at many levels. Different levels of convergence may be represented in the form of a layered architecture as shown in Figure 7.2. You are already familiar with layered architecture and the layering principles that were discussed in Unit 6. Same principles apply hereto. At a

Applications Convergence
Services Convergence
Access Convergence
Networks Convergence
Techniques Convergence
Technology Convergence
Sectoral Convergence

Fig. 7.2: Convergence architecture

very broad level, convergence is seen at sectoral level of infrastructure such as telecommunication sector, power sector etc. At the second level is the convergence in technology. Technology convergence is studied at two-levels. First, how different technologies in the same sector exhibit synergy to achieve certain common goals? Second, how technologies in different sectors come together to produce better products? Our main concern is the first question in ICT sector to achieve the goals of NEIS. Technology convergence is fundamental to the entire process of convergence in NEIS. Without technology convergence, other layers of convergence are not possible. Once the technologies are able to interwork, techniques that can be applied across all technologies need to be evolved. In particular, switching and signalling techniques need to converge. Techniques convergence layer deals with this aspect. Once converging techniques are available, the different networks and systems can be made to interwork with each other. Then comes the need for access convergence. From any one interface, the user must be able to access all the services available in NEIS. In the architecture shown in Fig. 7.2, a distinction is made between services and applications. Services are provided by the network operators and the applications are designed and run by end users. Applications make use of the network services.

7.9 SECTORAL CONVERGENCE

At a very broad level, convergence efforts have to start with major infrastructural sectors. Five sectors have been identified as the core in realising a worldwide networked society:

- 1) Telecommunications Sector;
- 2) Power Sector;
- 3) Internet Sector;
- 4) Computer Technology Sector;
- 5) Media Sector.

Each of the above sectors has evolved to deliver certain specific service to the society. The one thing that is common in all these sectors is that they reach out to individual homes. This is where the convergence of sectors becomes relevant. If one sector has reached out a home, can that sector be used to provide audio, video and data services to that home?

Telecommunications sector comprises mobile, landline, and satellite networks. Mobile and landline networks are essentially designed to carry voice signals between two subscribers. The services are connection oriented. You may recall that voice signals are limited to about one-fifth of the full audio spectrum. Hence, the networks are not capable of carrying the full 20-kHz audio spectrum in real time with high fidelity. The network connections are point-to-point in general. The connections are established on dial up basis. The convergence questions therefore are: Can the channel capacity in these networks be enhanced to carry the full spectrum of audio, video and data? How do we organise broadcast and multi-cast services on these networks?

Satellites are usually designed for multi-purpose applications. This is particularly true of Indian communication satellites belonging to the INSAT series. They are capable of carrying voice, video and data signals. The overall bandwidth of satellites is, however, limited. Satellite communication is inherently broadcast in nature. The convergence questions therefore are: How can we organise point-to-point connections efficiently via satellites? How best can the limited bandwidth be utilised?

Power sector offers an infrastructure that has the highest level of connectivity. Normally, the power transmission system covers the entire country and depending upon the extent of electrification, every home has electricity connection. The convergence question

therefore is how best can this large infrastructure be used to meet the goal of NEIS? Integration of telecommunication with power transmission and distribution is emerging as a global phenomenon. The technology is known as *power line carrier communication*. Use of power lines for communication purposes has been studied for sometime. Some of the hurdles in the process include the presence of transformers in the network, interference from household devices like hair dryers and microwave ovens, and lampposts that act as antenna introducing noise in the system. Efforts are on to overcome these difficulties and integrate power infrastructure as part of NEIS.

Internet is considered a content delivery medium. Standardisation in media coding is an important requirement for content delivery services. The Internet is estimated to be hosting about 25 million web sites as of Year 2005. Internet is designed as a data network for supporting data transfer applications. The convergence question therefore is how do we support voice and video applications on the Internet? Internet telephony or voice over IP (VoIP) is an attempt in this direction. The quality of VoIP is yet to reach satisfactory levels. At the current level of load, Internet response has become slow. How do we then add capacity to Internet?

Internet suffers from three distinct disadvantages that do not exist in conventional telecommunication networks. The reliability of telecommunication networks is 99.999% or more whereas that of Internet is about 99.9% or less. Signalling in telecommunication networks is highly structured and sophisticated and is completely hidden from the user. Such a design makes the network highly secure. On the other hand, Internet has no structured signalling and uses control message packets for signalling. Users can generate control message packets and can possibly tamper with signalling. This poses a security threat and makes applications vulnerable for attack by hackers. In other words, Internet is less secure than conventional systems. Conventional networks do not generally suffer from overload that is a major problem on the Internet.

The key to content delivery lies with computer technology sector. The ability to create, maintain and deliver on networks large databases, knowledge bases, data warehouses, digital libraries, electronic movie stores etc. depend on the availability of suitable computing power in terms of hardware and software. It is essential that computer technology sector converges with other sectors to make the dreams of NEIS a reality.

Finally, there is the media sector. Conventionally, media sector represents distributive services like publishing and broadcasting. Publishing includes newspapers, magazines, books, movies, and audio and video recordings. Broadcasting includes conventional radio and TV broadcasting, cable distribution of TV signals and *direct-to-home* satellite television. Currently, the term *media* has come to mean digital audio, digital video and digital data. In other words, media means multimedia information. In the present sense of the term, media services can be both distributive and interactive.

Conventional media sector is a one-way system. They reach out to a large number of homes and individuals. Can the media be made interactive to support the objectives of NEIS? The most promising service of the electronic media that can be converted to interactive mode is cable TV. Cable is the ideal medium for delivering broadband services to home. As of now, cable operators provide high-speed Internet connections using a separate cable. But true convergence envisages a single cable being used for TV transmission, Internet access and a variety of other services. Cable has also the excellent potential for supporting the upcoming *interactive TV* application. But the reach of cable is rather limited. India is estimated to have about 25 million cable connections as of Year 2003. As far as radio and TV broadcasting is concerned, it is more difficult to make them interactive. Attempts have, however, been made with the help of telecommunication networks. We are aware of live programmes on radio and TV where the listeners or viewers phone up the radio or TV station for interaction during the programme.

As you learnt in Unit 4, there are four major communication systems that form part of modern telecommunications:

- i) Electrical communication system
- ii) Optical communication system
- iii) Radio or wireless communication system
- iv) Satellite communication system.

Now, let us see how each one of these technologies converges to support the emergence of NEIS. The four communication technologies complement each other to improve **connectivity** and **capacity**.

Electrical communication uses copper cables of different types that have different information carrying capacities. Copper cables have been in use for a long time for establishing networks worldwide. Information carrying capacity of copper cables is being continuously improved and even today copper cable is a favorite medium in communication systems. Optical fibres are replacing the copper wires in certain segments of telecommunications because of their extremely high information carrying capacity. However, laying copper or optical cables to every village, home and office in the world is a formidable task and it may take many centuries if this were to be achieved. The enormity of the task becomes clear when we realise that we have been able to connect via landline only about 15% of the world population in the last 120 years. Further, laying optical fibres is more difficult than laying copper cables because of special considerations required for optical fibres in bending and routing. With such constraints, large-scale connectivity can not be achieved in a short time frame by using cables. However, only high speed copper cables and optical fibres can offer high bandwidths. The bandwidth capacity of optical fibres is, of course, orders of magnitude higher than that of copper cables.

Our interest in wireless communication lies in short-haul radio technology that may be placed under four categories: mobile communication, wireless LANs, piconet, and *wireless in local loop* (WLL or WILL). Since the introduction of cellular mobile telephony in 1990, over one billion mobile subscriptions have been registered throughout the world as of 2004. There is a discernible trend that indicates that new users prefer mobile connectivity to landline connectivity. Today over 90% of the countries in the world have mobile networks and over 100 countries have more mobile subscriptions than fixed landline subscriptions. Clearly, mobile communication is raising connectivity levels significantly. Data and video carrying capacity of mobile systems are very limited.

Wireless LANs are useful in such cases of office and home networks where wired infrastructure may not be available. They can also provide high capacity. Their main problem is in electromagnetic interference and lack of privacy and security. Wireless LANs may be affected by or cause interference from/to neighbouring networks. Eavesdropping is very easy on wireless LANs.

Piconet is a new convergent technology code named '**bluetooth**'. This 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 which are in the vicinity of each other to communicate in a wireless mode. For example, a laptop computer may connect to a mobile phone and receive electronic mail if the two devices have bluetooth interface incorporated in them. Bluetooth has potential for many interesting applications. Unlike mobile communication that calls for a radio network to be in place, bluetooth devices can communicate directly without any network support.

WLL short-haul radio technology has the excellent potential of making interactive services a reality. Residential premises may be connected to nearby communication centres via WLL radio technology. The centres may be equipped to offer large bandwidth using

optical fibres. This principle may be extended to provide limited interactive facility to individual users via mobile communication technology. However, WLL is capable of providing much higher capacity than mobile technology. The use of short haul radio implies the establishment of radio network, which can be done more easily in urban areas than in rural areas.

In the context of remote area connectivity, satellite communication presents a practical solution for connecting remote areas. Establishing connectivity via satellite links is by far the quickest. Even remotest villages can be brought on to the world network map in a matter of few hours by using satellite links. Local cabling may be used to carry information between the satellite terminal and the houses in the village. Satellite links do not need any landline infrastructure except for local distribution. At present, efforts are on to develop *satellite mobile systems* wherein a hand-held device may connect to a satellite directly without the help of a terminal. This also offers an elegant solution to rural area connectivity even though the capacity of the hand-held terminals may be very limited. As mentioned earlier, the main problem with the satellite is the limited capacity. The bandwidth of a typical present day communication satellite is at least a few hundred times lower than that offered by a single optical fibre. Limited bandwidth of satellites does not permit high-capacity two-way interactive communication to individual homes.

Trained manpower is an important consideration in selecting technologies. Copper cable communication is well known and a large number of trained technicians are available to install and maintain such systems. This is the reason why a large number of cable operators have sprung up in the market. Optical fibres are new and satellite communication is highly specialised. Manpower is limited in both areas. Radio communication is well understood but mobile communication involves new concepts that are not taught extensively at present.

Table 7.1 summarises the strengths and limitations of different technologies that we have discussed so far. The convergence of the four communication technologies, inter-operating with each other, can provide a suitable solution to the problems of connectivity, capacity and trained manpower in the networked information society of tomorrow.

Table 7.1: Technology Convergence

Technology	Connectivity	Capacity	Manpower
Copper cables	Fair	Reasonable	Very Good
Optical fibres	Difficult	Very High	Limited
Satellite	Rural areas	Limited	Limited
Short haul radio	Urban areas	Limited	Low

We now have a paradoxical situation. Short-haul radio and satellite communication enhance connectivity but have limited bandwidth capabilities. High speed copper cables and optical fibres can offer very high bandwidth but have serious connectivity limitations. Hopefully, these technologies are poised to play complementary roles in shaping the networked information society of tomorrow. Clearly, the goal of telecommunications in the years to come is to establish a wide band multi-services network reaching out as large a population of the world as possible.

7.11 TECHNIQUES CONVERGENCE

Switching, signalling and transmission techniques are the main ones in networks. Signalling includes addressing and routing apart from monitoring sessions and connections. In this section, we shall bring out the divergence that exists in these areas and the convergence efforts that are currently being pursued.

As you are aware, the principle of switching is central to all networks. The switching technique adopted by data networks is different from the one used by telephone networks.

Telephone networks use the technique of **circuit switching** whereas data networks use **packet switching**. In circuit switching, a communication path is established between the source and destination before the actual information transfer takes place. This path and the associated resources remain dedicated to the communicating pair until any one of the two entities initiates a disconnect request. The path is used effectively if there is continuous traffic during the service period. However, if the traffic between the source and destination is not continuous, the path remains partly idle during the service, leading to inefficient use of network resources. In packet switching, information transfer takes place in a *store and forward* fashion from node to node in the form of packets. At every node, the destination address is processed before the next forward move is made. The packets move from one location to another in the general direction of the destination. There is no connection between the source station and the destination station.

In the context of multi services network, it is recognised that packet switching is not suitable for real time services like telephony and video distribution. And, circuit switching is not efficient for bursty non-continuous traffic from the point of view of utilising network resources. The realisation that both circuit switching and packet switching are required for different purposes on a common digital network led to the search for new switching techniques. The two techniques have now converged and a new switching technique called **cell switching** is now accepted as the standard switching technique for future broad band infrastructure. Cell is a small-sized packet that moves very fast on a digital infrastructure. Cell switching is capable of supporting both real time and non-real time services efficiently. However, it is foreseen that for sometime to come, all the three switching techniques would co-exist.

The importance of signalling in networks is reflected by statements such as 'If we can not signal, we can not make a call' and 'If we can not signal, we can not transfer a packet'. Attempts to convert existing single-service networks to multi-service ones has led to the emergence of a plethora of signalling techniques. Specific convergence questions are how do we support real time services like telephony on IP-based packet networks and how do we transport packets efficiently on circuit switched networks? Circuit-switched and packet-switched networks use entirely different forms of signalling. The signalling protocols are also very different. Convergence signalling protocols on the IP-based packet networks include *media gateway control protocol* (MGCP), *session initiation protocol* (SIP) and *real time control protocol* (RTCP). These are protocols brought out by the Internet Society. ITU-T has brought out recommendation H.323 that allows interworking of circuit switched networks and IP-based packet networks. H.323 refers to a number other signalling protocols used in circuit switched networks. H.323 is the key convergence protocol that is used at present to offer real time services like IP Telephony on IP-based packet networks.

At the transmission level, two transmission hierarchies are emerging. One is ITU recommendation known as *synchronous digital hierarchy* (SDH) and the other evolved by U.S. in the context of optical networks known as *optical carrier* (OC) hierarchy. OC transmission systems run on *synchronous optical networks* (SONET). SDH is applicable to both copper and optical networks. Luckily, the transmission speeds are identical in both SDH and OC at different levels of hierarchy. The speeds at which SONET runs are specified as OC1, OC2, OC3 etc. The lowest speed of SONET is OC1 at 51.84 Mbps. SDH speeds start at OC3 level that corresponds to 155.52 Mbps. The basic speed of ATM network is also OC3.

7.12 NETWORKS CONVERGENCE

In the early days of modern telecommunication, the underlying technology was analog. Different media generated signals with different characteristics. To transmit such signals over distances, we needed different types of end equipment and networks that were suitable for each type of signal. For example, the voltage, current and frequency range requirements of a telephone instrument, fax machine and a computer terminal are vastly

different. This is what led to the development of a variety of telecommunication networks in the past such as:

- Telegraph networks;
- Telex networks;
- Telephone networks;
- Data networks;
- Alarm networks;
- CATV (cable) networks etc.

Independent networks call for separate administration, maintenance staff and buildings to house the equipment. Independent and duplicate infrastructure facilities lead to high capital cost, low maintenance efficiency and high running cost. In addition, the network facilities are never fully utilised. The net result is that the overheads turn out to be excessive leading to economically unviable network services. To keep the network costs down, the resources need to be shared among different services. This would be possible if the signal characteristics are uniform across the services. Such a solution became possible with the advent of digital technology. With digital technology, a variety of services can be supported on a single digital network. This approach has led to the evolution of the **integrated services digital network (ISDN)** which is the first major multi services network being planned and implemented all over the world. ISDN as envisioned in 1984 provides for a digital pipe at the customer premises with a minimum capacity of 144 kbps and a maximum of 2048 kbps. The realisation of the inadequacy of these rates in the context of NEIS soon led to the definition of **broadband ISDN (B-ISDN)**. Conceived as a blueprint for future telecommunication, B-ISDN would support channels operating at bit rates in the range of 155 Mbps to 622 Mbps at the user premises.

Another important effort in network convergence is the extension of Ethernet over wide area networks. The effort is to create what is called *converged Ethernet transport infrastructure (CETI)*. In CETI, Ethernet capabilities can be accessed seamlessly across ATM and fibre-based networks like SONET. For this purpose, an extended switching capability called *multi-protocol label switching (MPLS)* is being developed. MPLS is expected to extend ISDN and Internet features to LAN level. For example, features such as Ethernet virtual line and Ethernet VPN may become available in future

7.13 ACCESS CONVERGENCE

In Section 7.12, we discussed the need for network convergence for economic reasons. The economic considerations do not apply only to network providers. It is equally important for the users. A user may not invest in different types of access interfaces for different services. A single telecommunication connection should fetch all types of services to a user subject only to the limitations of speed and capacity of his interface. It is this philosophy that has led to the idea of access convergence. The trend of access convergence is being seen in the following:

- 1) Access to Internet from mobile handset and vice versa;
- 2) Access to global satellite services from mobile and vice versa;
- 3) Access to Internet from telephone network and vice versa;
- 4) Access to global satellite services from telephone network and vice versa;
- 5) Automatic access among bluetooth devices;
- 6) Access to Internet via television cables and vice versa.

Access convergence provides tremendous flexibility to end-users. They are able to access a variety of services from whatever device they have on hand. For example, a person with a mobile and a laptop can literally access all Internet services without physically

7.14 SERVICE CONVERGENCE

The idea of convergence in different areas that we have discussed so far is to carry multiple services using one common infrastructure. The trend of service convergence, i.e. offering multiple services in one interface, is seen in the following:

- Real time services like telephony on the Internet;
- E-mail via mobile phones;
- Web access via bluetooth interface and mobile phones;
- E-mail and Web access via television cables;
- Internet services delivered on TV sets;
- Broadcasting radio and TV programmes on the Internet;
- Normal telephone services via cable TV networks.

Broadcasting on the Internet is sometimes called *webcasting*. The above are examples of some of the recently introduced services using the concept of convergence. For the future, ISDN and broadband ISDN are being designed to carry every conceivable service. Let us now see how service convergence issue is being addressed in broadband ISDN. In broadband ISDN, the services are placed under some generic categories as shown under:

- Audio services
 - Telephone quality (speech)
 - Broadcast quality (music)
- Motion video services
 - Studio quality
 - Broadcast quality
 - High definition television (HDTV)
- Still video services
 - Low resolution (fax)
 - High resolution (pictures and photographs)
- Text services
- Computer graphics services
- Computer animation services
- Satellite imageries

Each of the above services, when digitised, demand different communication capacities, i.e. different speeds. To carry a variety of services with different capacity requirements, the architecture of broadband ISDN uses the ATM information transport mechanism. ATM is structured in a layered manner with different functionality stacked one over another. One of the layers is known as **ATM adaptation layer**. It has a sublayer called **convergence sublayer**. This sublayer has been designed to multiplex a variety of services and carry them on the underlying network infrastructure. The underlying infrastructure for ATM is by and large an optical network like SONET.

7.15 APPLICATIONS CONVERGENCE

In this section, we shall discuss some important upcoming convergence applications: *Music on demand*, *Internet radio*, *Internet telephony* and *Interactive Television*.

Music on demand, known as *streaming audio* in technical terms deals with listening to music on Internet. As you know, Internet is a data network capable of transporting 1s and 0s. Once digitised, music is in the form of 1s and 0s. Therefore, theoretically digital music can be transported over Internet. Listening to music happens in real time. But, the Internet is not designed to provide real time services. Therefore, the convergence question is 'how to deliver music in real time on Internet?'

You learnt about digital sound and the associated standards in Unit 8 of MLIS-01. Digital sound is compressed before storage. One of the most popular standards used for digitisation and compression is MP3 standard and format. There are many web sites on the Internet that store music in MP3 digital form. The most popular application architecture on the Internet is *client-server architecture*. You learnt about this architecture in Section 6.7 of MLIS-05. The server runs web server software and the client web browser software. The interaction between the web browser and the web server takes place via the application protocol HTTP that runs on TCP. HTTP is not a real time protocol. One way to listen to music under HTTP is to download the entire music file and play it from the client machine. A typical song in MP3 format takes about 4MB and it would take about 8 minutes to download the entire file at 64 kbps speed. Internet rarely runs at that speed. It could sometimes take as much as 15 minutes to download the entire file. The user will have to sit idle during download of each song. This clearly is an unacceptable solution. Streaming audio overcomes this deficiency and provides a near real time solution.

Streaming audio application is designed under client-server architecture supplemented with a media server on the server side and a media player on the client side. The media server and the media player interact via the *real time protocol* (RTP) that runs on UDP. You learnt about RTP in Section 6.6. Initially, the user interacts with the server via the browser 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 Figure 7.3.

RTP runs on UDP. UDP runs on IP which is a packet protocol. A packet in the case of streaming audio comprises an accumulation of digital samples of music/speech for certain duration, usually 5-8 msec. In other words, each sample is not transmitted in real time but about 40-64 samples are grouped together and transmitted. The grouping of samples and making them available in bursts lead to unnatural audio. However, grouping up to 6 msec is not noticeable by the human ear. As you know, in packet switching, each packet takes a different time to reach the destination depending on the traffic conditions on the network.

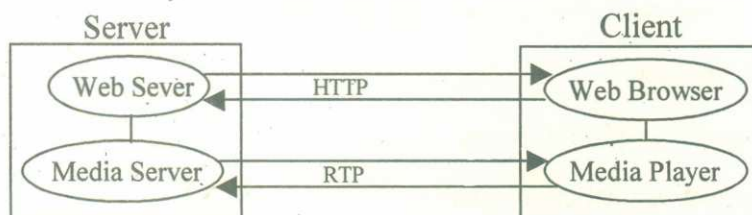


Fig. 7.3: Streaming audio configuration

While an average delay is computable, a packet may arrive much later than the average delay or even be lost on the way. Such delay variation from packet to packet causes jitters in the music affecting the quality. Streaming audio takes care of these problems by providing a music buffer at its input. 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 full 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 and hence the name *streaming audio*. On the one side the buffer is being emptied and on the other it is being filled. The provision of the buffer ensures that continuous music is available even in the presence variable packet delays.

In order to provide flexibility to user in listening, certain control functions as available in a cassette player need to be implemented by the media player-server combination. The control functions include controls like *pause*, *resume*, *stop*, *fast forward* and *fast backward*. Some flow control feature is also required. For example, if the packets arrive too fast from the server, then the buffer would become full and the packets would be discarded at the player end. To prevent this from happening, the media player may ask the media server to pause the flow momentarily. All such control functions 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. The protocols are *real time streaming protocol* (RTSP) and *session initiation protocol* (SIP).

Internet radio is based on the same principles as streaming audio. 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 programme is played out by the station. There is no user interaction for selection. This means that HTTP interaction is absent. User also has no control like *pause*, *resume* etc. Some radio stations play a second channel which 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 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.

The architecture of streaming audio is the building block of **streaming video** too. Video bandwidth is much larger and hence requires higher speeds of transmission on the Internet. The buffer size on the client side is also much larger as the sampling rate for digitisation of video is much higher. For example, 10 seconds of audio requires 80 kB of buffer whereas video for the same duration would require 2 MB of buffer. Another important requirement of video streaming is the synchronisation required between the video signal and the accompanying audio signal. Video streaming is the basis for a variety of applications such as distance learning, digital libraries, home shopping and video-on-demand. If in the future you read your lessons online from IGNOU, remember that video streaming is at work.

Internet telephony is also known as **Voice over Internet Protocol** or simply **Voice over IP** (VoIP). Internet telephony is built around a set of protocols and standards defined by ITU. IT also uses the set of protocols we discussed as part of streaming audio. It was in 1996 that first attempts were made to build Internet telephony gateways when ITU first proposed architecture suitable for voice services in LANs in its standard H 323. This standard was revised in 1998 and forms the basis for Internet telephony. The architecture is depicted in Figure 7.4. Basically, the Internet and the telephone network are interconnected by a gateway. The gateway implements H.323 protocols on the Internet

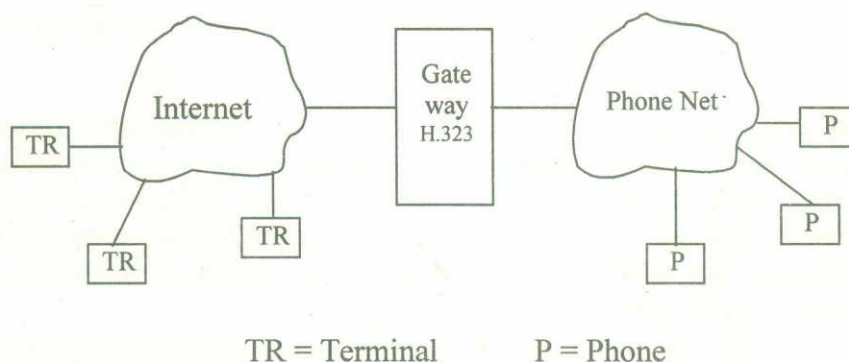


Fig. 7.4: Internet telephony architecture

side and PSTN protocols on the telephone network side. The end devices on the Internet are terminals and on the PSTN they are telephones. H.323 refers to a number of other ITU protocols. The telephone speech is digital and should conform to G.711 ITU recommendation that specifies 64 kbps PCM stream. G.711 conformity is mandatory for VoIP. Some other compressed audio streams that run at lower speeds like 6.4 kbps or 5.3 kbps are permitted optionally. The digital stream is transported over RTP under the supervision of RTCP. Both RTP and RTCP run on UDP. Since multiple compression schemes are permitted optionally, there is a need to negotiate and agree on the scheme before start of the session. This is done by call control protocol specified in H.245. Call establishment and release and call signalling, like sending out ring tone etc. are done according to ITU signalling protocol Q.931. The call control and call signalling protocols run on TCP. On the LANs, ITU suggests the use of a gatekeeper device that organises telephone call for LAN terminals. A protocol, named *registration, admission and status* (RAS) has been defined for the gatekeeper to interact with LAN terminals. The RAS protocol runs on UDP.

Interactive television (ITV), also called two-way television is a convergence application where a one-way video broadcasting system is turned into a two-way system capable of handling media, i.e. voice, video and data. ITV is a television set with a return path to the broadcaster or 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 offer video-on-demand service. For the present, ITV systems are providing broadcast programmes with facility for interaction. ITV is evolving in three stages:

- Programmes are broadcast and the return path is provided via PSTN.
- Programmes are broadcast and the return path is provided via set-top box.
- Video-on-demand programmes and return path via set-top box.

The set-top box 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 network. 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 may be exclusive to that viewer.

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 sport events. This indeed is exciting. Users may pause and resume programmes that they are viewing, provided the programmes are not live. With live programmes, the users may record the programmes for later viewing. And, of course, e-mail can be sent through ITV.

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 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 he/she is interested. The negative aspect is that the service provider is actually treading into the private life of individuals.

ITV systems call for wide band infrastructure to transport individual programmes to different viewers. This infrastructure is an ATM network with *synchronous optical network* (SONET) as the transport infrastructure.

- 7) Why do we need buffers in streaming services?
- 8) A user listens to a radio station via a conventional radio receiver and via Internet simultaneously. What differences would he notice and why?
- 9) Depict the protocol stack for Internet telephony in the form of layered architecture.

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

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

.....

.....

.....

.....

.....

.....

.....

7.16 SUMMARY

In this Unit, the convergence aspects in the context of NEIS have been discussed. Firstly, the Unit places convergence in perspective by explaining what convergence is. The goal and objectives of convergence are then outlined. The basic requirements of convergence including standardisation are then highlighted. The genesis of convergence is traced to remote computing that emerged in late 1950s. The factors that drive convergence efforts are then brought out. Three important issues, viz. connectivity, capacity and content on which convergence efforts must focus are discussed. A layered architecture that depicts the different levels at which the convergence needs to take place is then presented. The convergent phenomena in technology, network, switching, access and service areas are discussed in detail. At the end, how some important convergent applications are realised is discussed.

7.17 ANSWERS TO SELF CHECK EXERCISES

- 1) The goal of convergence may be stated as:

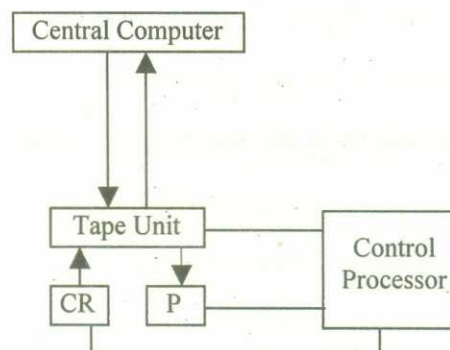
To achieve the capability wherein any network can deliver any service to any platform, and any user with an appropriate consumer device can access any application that runs on the networked world.

The goal reflects the existence of multiplicity in various components of the networked society. Different types of networks, different computer platforms and different access devices would all form part of the networked society. The goal is to enable any user to access any application via any suitable access device in this conglomerate of networks, devices and systems.

- 2) There are three fundamental forms of information: audio, video and text. An integrated consumer device is one that is capable of accessing any form of information. Using an integrated device, one may receive a telephone call or a radio broadcast, receive TV broadcast via cable or directly from an antenna or access textual information from Internet.
- 3) In a RJE setup, a tape unit is used to speed up data transfer between the RJE units and the central computer. It acts as an interface between the slow RJE devices and

the fast central computer. A diagrammatic representation reflects this role of the tape unit and is given below.

In the diagram, the card reader (CR) and the printer (P) are connected to the tape drive. The tape unit is connected to the central computer. The card reader is the



input device and the printer is the output device. There is a control processor that co-ordinates the functioning of all the three units. (Note: The learners may additionally show the communication subsystems in the diagram)

- 4) In both multi-programming and time-sharing systems I/O operations and processing on the CPU proceed simultaneously. In both the systems, multiple programs are active and are in a ready state for execution. However, only one program can use the CPU at a time.

In multi-programming systems, when the CPU is assigned to a program, it uses the same until it is blocked for some I/O operation like seeking data from a disk. At that time, the CPU is assigned to another program that uses the CPU in a similar fashion. In other words, the programs use the CPU continuously as long as they are not blocked for I/O or terminated.

Time-sharing systems are designed for interactive computing with many users sharing the computing resources online at the same time. In these systems, the CPU is assigned for a fixed time called *time slice* for each user in ready state in a round robin fashion. A user may be in blocked state if his program is performing I/O operations. Once an active user uses up his time slice, he is put back at the end of the ready queue and has to wait for the next allotment of time slice.

- 5) Examples of connectivity oriented services are simple telephone calls and fax transmission. In both cases, the network establishes a circuit between the two end points. What information transfer takes place and how it takes place are the concern of the end entities. Network has nothing to do with information transfer. If information transfer fails, the end entities take corrective action.

In content delivery system, the content is provided by the network and the mode of transfer is also decided by the network. The network takes the responsibility for error-free delivery of information. Special protocols are used to achieve efficiency and error control. Examples of content delivery are the *music-on-demand* and information retrieval services.

- 6) The arguments in favour of increasing connectivity instead of offering contents are:

- The traffic on account of people wanting to exchange contents within their own circles is much more than the people wanting to access contents from the network.
- More revenue is generated by connectivity oriented services. For example, the highest revenue is earned from voice conversation in both landline and mobile networks.

- People need connectivity to their near and dear ones more than the contents from the network.
 - If effective, efficient and economical connectivity is available, the people tend to create their own contents.
 - Remaining connected is more important for people in day-to-day life.
- 7) Packets in a packet-switched network experience variable delays in travelling from source to destination. When the packets carry real time information like music or TV programme, the variable delays result in jitters when the information is delivered to the end user directly. To avoid such jitters, the real time information is first accumulated for certain duration in buffers. Once the buffer is full, the delivery starts from the buffer. The delivery is without jitters as the packets come from the buffer and not from the network directly. The variable delays only affect the way the buffer is being filled and not the way in which delivery is effected.
 - 8) The Internet radio will lag in time compared to the conventional radio. This is due to the fact that Internet radio uses a buffer that is filled before delivery starts for the user. The amount of lag is decided by the buffer size.
 - 9) A layered architecture for Internet radio may be depicted as in the following:

Internet Radio
RTP/RTCP SIP/RTSP
UDP/TCP
IP
Internet

7.18 KEYWORDS

- Bluetooth** : Low power radio technology to enable two or more devices that are located in the near vicinity of each other to interact.
- B-ISDN** : Broadband ISDN. See ISDN below. An ISDN with very large information carrying capacity.
- Cell switching** : A convergent switching technique that combines the strengths of circuit and packet switching.
- Convergence** : The process of bringing together different technologies, techniques etc. to achieve a common goal.
- Distributive services** : Services that are broadcast or multicast in nature, i.e. services that are distributed to a large number of users without any request from the users.
- E-Services** : Electronic services accessible from a remote place via a network.
- Interactive services** : Services that allow interaction between the server computer and the user, e.g. searching a database.
- Internet radio** : Multicasting of radio programmes on the Internet.
- ISDN** : Integrated Services Digital Network. A network in which all signals are represented and transported in digital form.
- ITU** : International Telecommunication Union involved in standardisation for convergence.

- Media player** : Software on the client machine for supporting real time services on packet networks.
- Media server** : Software on the server machine for supporting real time services on packet networks.
- On-demand services** : Services that can be obtained by users by placing a demand on the network.
- RJE terminal** : Remote Job Entry terminal. An equipment setup to use computers from remote places.
- Satellite telephony** : Telephone service via satellites.
- Streaming** : Technique of organising real time services via packet switched networks.

7.19 REFERENCES AND FURTHER READING

Convergence of the telecommunications, information technology, media and power sectors. (2001). New Delhi: Centre of Publications

Douskalis, B. (2001) *IP telephony*. New Delhi: Pearson Education Asia.

Hersent, O. Gurle, D. and Petit, J.P. (2001). *IP telephony*. New Delhi: Pearson Education Asia.

Johnston, C.B. *Communication conglomerates*. In: *Encyclopaedia of information systems*, Volume 2. USA: Elsevier Science.

New Telecom Policy. (1999). New Delhi: Government of India

Pavlik, J.V. and Powell III, A.C. (2003). *New media*. In: *Encyclopaedia of information systems*, Volume 2. USA: Elsevier Science.

Rao, K.R.; Bojkovic, Z.S. and Milovanovic, D.A. (2002). *Multimedia communication systems*. New Delhi: Prentice Hall of India.

Tanenbaum, A. S. (2002) *Computer networks*. 4th Edition. New Delhi: Prentice Hall of India.