EEX3373

Block 02

COMMUNICATIONS AND COMPUTER TECHNOLOGY

Published by
The Open University of Sri Lanka

Course Team

Course Team Chair
HUW Rathnayake

Course Team Manager
GSN Meedin

Author
LSK Udugama
CJ Basnayakege
NR Prematilake
NAR Priyanka

Content Editors
HUW Rathnayake
LSK Udugama
DN Balasuriya
US Premarathne

Desktop Publishing
HR Jayetileke

Graphic Artist
HR Jayetileke

The Open University of Sri Lanka
Nawala, Nugegoda, Sri Lanka

First published 2019

Introduction to the course

In the modern era, Information is a vital component in our day to day life. Any Information at our finger-tips had been a dream over many decade and is gradually becoming a reality today. Many new technologies such as Internet of Things have been clear indications of this success.

With the advent of the human activities, the need for the information has also grown and the types of information required had been diverse. On the other hand, sources of information and the location where information is required has become more and more distanced. Consequently, the information transfer from source to destination had become a substantial task. Amidst this challenge, many new technologies have been proposed and implemented to handle the ever-growing information communication need.

As an engineering student embarking in a degree program specializing in either electrical engineering, electronics engineering or computer engineering one should have a sound background knowledge of information communication as it will be the foundation for many of the new trends in all three above disciplines.

In this Unit, first the information communication is introduced where more emphasis is given for the computer networking standards, hardware and various other configurations. Next the underlining techniques of telecommunication which provides the infrastructure for the information communication is presented. More specifically the latter half of this unit describes the physical media employed in today’s communication, some of the additional processing techniques used during information transfer and also gives an insight to the communication system design.

At the completion of this Block 02 a student will be able to describe the information communication networks focusing on standards, hardware and also the security aspects. Furthermore, the student would be able to compare different media types used as infrastructure for networking, describe and compere different modulation techniques and to perform simple design calculations for communication networks.

Wish you all the very best.

EEX 3373 course team

Contents

Session 9
Introduction to networking

Contents

INTRODUCTION

This session provides you with an overview of computer networking. We discuss how the concept of computer networking evolved in history. Also, we learn why we connect computers, what data communication is and how a simple network is physically built for data communication. After that we discuss the internationally agreed standard model for data communication (ISO-OSI reference model) and its layered architecture. Also, we learn how the ISO-OSI model is deployed in communication protocol stack of the Internet. Knowledge of these areas is essential for you to proceed to the next sessions in this course unit.

9.1 CONCEPT OF COMPUTER NETWORKING

Although the early micro-computers (personal computers) were isolated (stand-alone) machines used by individuals. However, it didn't take long to realize the need to share the computer resources. The concept of networking emerged during the 4th generation of computers in 1970s and 1980s. It started because of tedious work experience with dedicated printers connected to computers in organizations. Many computers wanted to share the printers without keeping them idle most of the time. Also, it was necessary to share data storages without maintaining duplicates of same data on different desktops. Some of the reasons that motivated networking of computers and devices includes the following.

- To share expensive resources such as Printers, Disk/Tape drives, Computational Power, Data Sets, application programs etc.
- To permit the distribution of workload
- Concurrent/parallel computing, Client-server computing, Fault tolerance
- To facilitate the transfer of information
- Database transactions, Electronic mail and Networked distribution of data archives

Growth of communication traffic, emergence of new services and technological advancements have consistently driven the evolution of data communications and networks. The growth of network traffic attributes to various applications (e.g. factory automation, voice and data services by telecommunication service providers) demands to expand the network capacities and efficient transmission techniques. Therefore, the technologies evolved to connect computers of different types forming local area networks (LANs). And later the LANs were expanded to connect geographically dispersed computers forming metropolitan area networks (MAN) and wide area networks (WANs). Incidentally, personal area networks (PAN) were also built to cover much smaller ranges. With the development of intercontinental communication lines, these individual networks were connected resulting in a massive network that we today call the Internet. The Internet has developed to offer hundreds of services in a scenario of a global e-society.

DATA COMMUNICATIONS

Data communication is the process of transferring digital data in a reliable and efficient manner. We need data communications among the devices connected in a computer network to facilitate the intra-communications within the network and inter-communications between different networks.

To understand the process of data communications (see Figure 9.1), let us consider a simple example and formulate a model. Consider a scenario where a user "X" intend to send a text message (e.g. via electronic mail) to another user "Y". The term "user" here refers to the computer which resides in the network rather than the human being using that device. So, the information originates, and data get transmitted at "X". At "Y", the data is received, and the information is retrieved. The information (m) is typed using a keyboard and stored as a sequence of bits (Bs(t)) in "X". Then, the binary data is then transmitted over a wired or a wireless transmission medium as an analog signal (s(t)). At "Y", the transmitted signal is received (r(t)) and stored as a binary sequence (Br(t)). Then, the information is retrieved (m' = m + e(t), where e(t) is the transmission error) from the binary data at "Y".

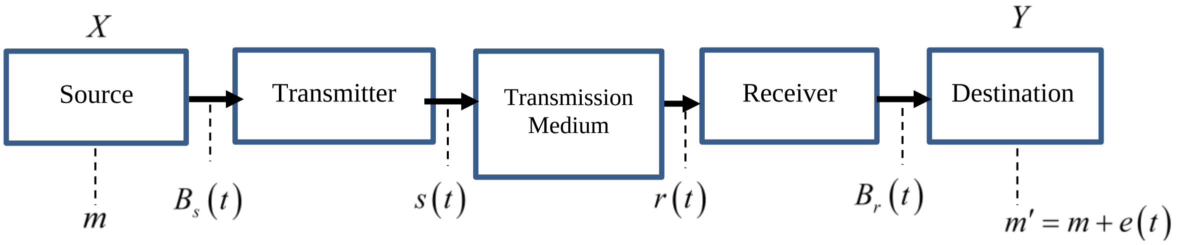


Figure 9.1 Data Communication Model

COMPUTER NETWORKS

A computer consists of a lot of resources such as memory, processor and disk-space that is not being used even when somebody is using it. Further, there are external resources which can be connected to a computer to do specific tasks. For an example, a printer is connected to a computer or several such computers to print documents. These kinds of external resources can be shared by different computers in a network. On the other hand, if the computers are connected or if other people can access the resources of a computer, idling time of resources can be minimized. Computers are connected to network communications either by using a ***modem*** (external or internal)

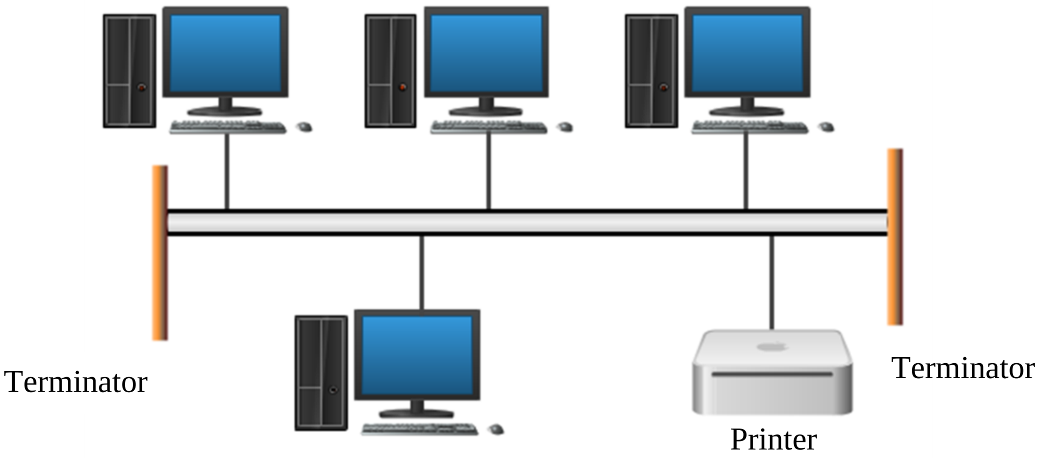


Figure 9.2 A set of computers networked to use a printer.

or a ***network interface card (NIC)***. Hubs, switches and routers are useful to connect different computers in a network as well as different networks to expand further.

Two or more computer systems connected together using data communication links is called a network. The concept of networking came primarily for sharing resources such as processor power, hard disk space and external resources like printers etc.

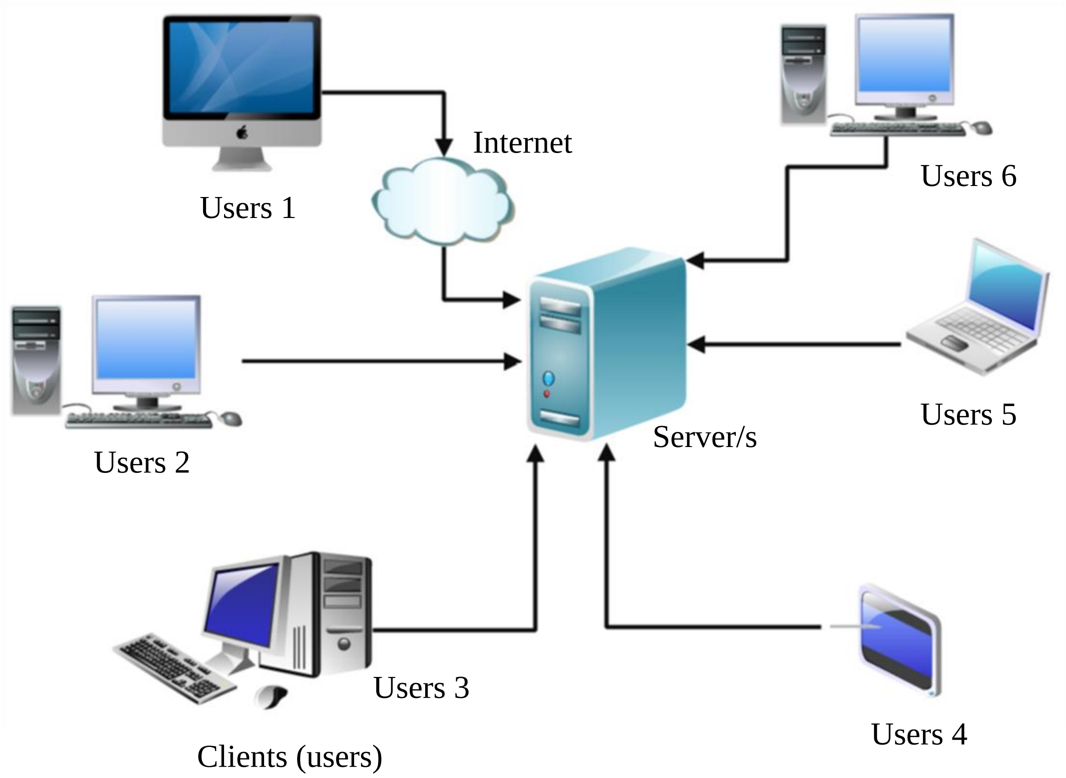


Figure 9.2.a Client-server network architecture

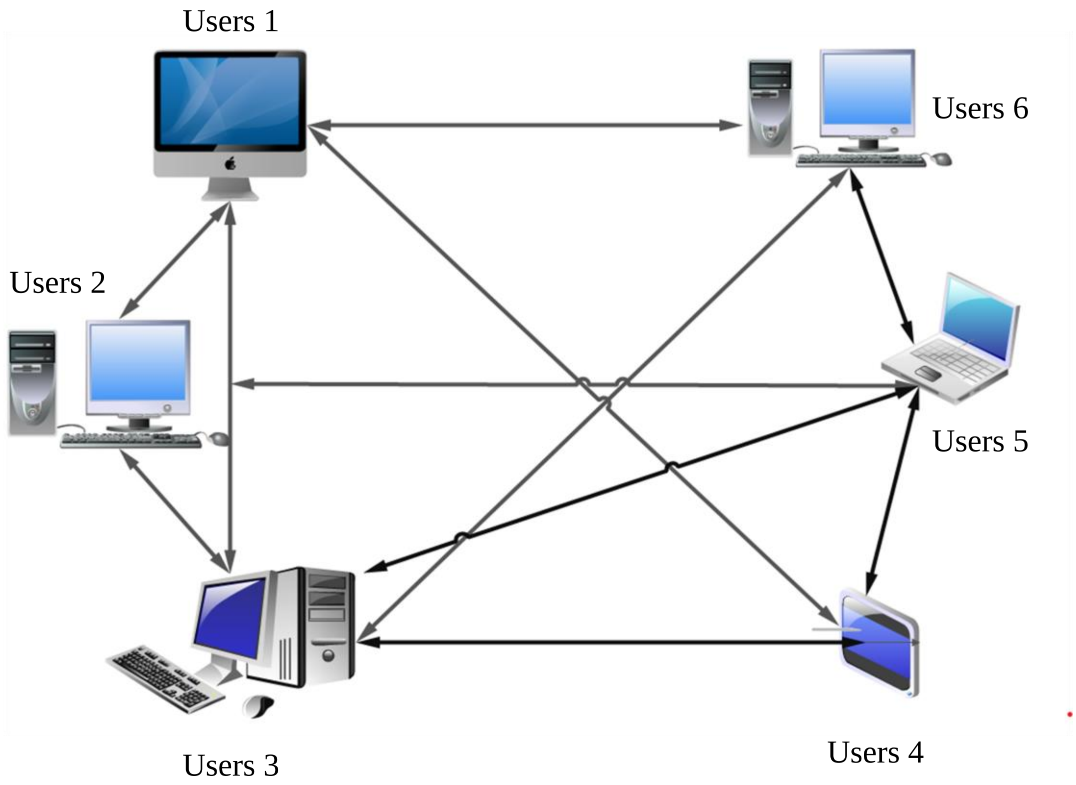


Figure 9.2.b Peer-to-peer architectures

BUILDING A COMPUTER NETWORK

Depending on the technology, structure and area of coverage, there are different types of networks available. As shown in Figure 9.2.a and Figure 9.2.b, among the different approaches to build a network (in an office or a work place for resource sharing) includes are client server and peer-peer networking architectures. As already discussed, the networks are classified according to their geographical size, ranging from personal area networks (PAN), local area networks (LANs) to wide area networks (WANs). Moreover, there are multiple ways of connecting computers to facilitate inter-communications in a LAN which are known as topologies. Bus topology, ring topology, star topology and tree topology are some of the example topologies. In the next few sessions, we will discuss all these different aspects of networking in detail.

SAQ: Discuss the advantages developing standards (e.g. IEEE 802.11a) for computer networks.

9.2 OSI REFERENCE MODEL

The OSI Reference Model is based on a proposal developed by the International Organization for Standardization (ISO) along with the ITU-T. The model is called ISO OSI (Open Systems Interconnection) Reference Model because it is concerned with connecting open systems. Open systems means the systems that are open for communication with other systems – that is not proprietary. Prior to OSI, networking protocols were proprietary. Though large networks support multiple network protocol suites, many devices were unable to talk to other devices due to lack of common protocols between them. With protocol standards such as SNA, AppleTalk, and NetWare, OSI attempts to get everyone to agree to use common network standards and protocols.

Let us understand the following terminology related to the OSI model in data communication;

1. **Protocol** – is a set of agreed rules between communicating parties on how communication is to proceed.
2. **Peer entities** are the entities in the corresponding layers on different computer systems.
3. **Interface** – defines which operations and services the lower layer makes available to the adjacent upper layer.
4. **Flow Control** – Speed of the sender may differ from the speed of the receiver. In such a case both sender and receiver should agree on how to send and receive data without them being lost. Receiver should acknowledge receipt of data indicating receiver's current situation.
5. **Error Control** – It is concerned with error-detection and/or error correcting of the passing messages. Receiver should tell the sender which messages were correctly received or not so that sender may resend them or discard the whole message.
6. **Routing** – routing is the function of facilitating a path from a source node on one network to a destination node on another network.

In the OSI reference model, the process of communication is partitioned onto a set of **seven layers**. Table 9.1 summarizes the functions and the protocols corresponding to each layer. To understand the usefulness of the OSI model, let us consider an example. In Figure 9.3, it is illustrated how a data packet is passing through from the application layer of one system to the application layer of another system. Based on this example, we can observe the following,

- each layer performs a related subset of functions required for communication
- each layer provides services to the next higher layer while depending on the next lower layer to perform more primitive functions
- communication is achieved by having corresponding (peer) entities in the same layer in two different systems communicate via a protocol
- each protocol entity sends data down to the next lower layer to get the data across to its peer entity
- each entity communicates with entities in the layers above it and below it, across an interface

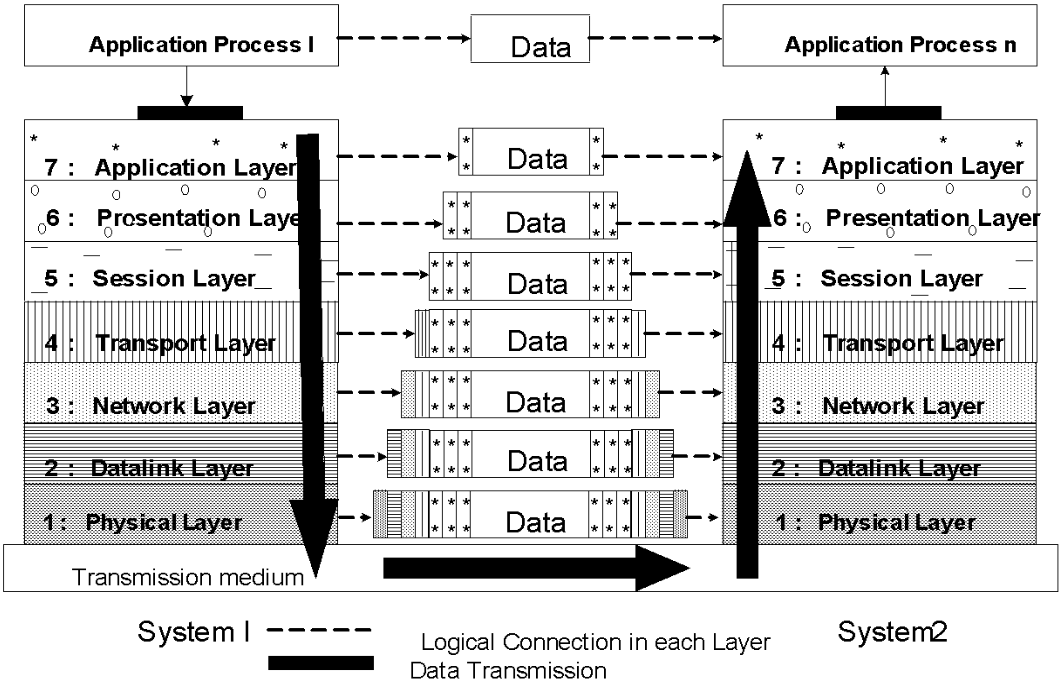


Figure 9.3 Data packet transmitted from system1 to system2

Table 9.1 Summary of the Seven (07) Layers of OSI Reference Model

Layer	What the layer is responsible for	Information maintained	Protocols	Devices
Applica tion Layer	Provides common interface for different user applications	User data	Telnet, FTP, HTTP, SNMP POP3, SOAP	Gateways
Present ation Layer	Concerned with the syntax and semantics of the information transmitted	Messages	ASCII, EBCIDIC, GIF, JPEG	Gateways

Session Layer	Establishes and manages end-user connections i.e. Allows users on different machines to establish sessions between them E.g. a session might be used to allow a user to log into a remote time-sharing system or to transfer a file between two machines.		RPC	Gateways
Transport Layer	Provides the functions necessary to guarantee a reliable network link. Provides error recovery and flow control between two end points of the network connection. Insulates upper layers (5-7) from having to deal with the complexities of layers 1-3.	Datagrams	TCP, UDP	Gateways
Network Layer	Responsible for establishing, maintaining, and terminating network connections Does routing and forwarding of information in the network	Packet	IPV4/IPV6, X.25	Routers and “layer three” switches

Data Link Layer	Responsible for reliability of the physical link established at layer 1. Create and recognize frame boundaries (framing) Provides error control between adjacent nodes	Frame	Token Ring Ethernet SLIP, PPP HDLC ATM, FDDI, X.25 (LAPB)	Switches, which forward data packets based on the MAC address, bridges
Physical Layer	Concerned with transmitting raw bits over a communication channel.	Bit	10BaseT, Bluetooth, ISDN, DSL, RS232, X.25, Token Ring, Ethernet	NIC's repeaters, hubs

TCP/IP (TRANSMISSION CONTROL PROTOCOL/ INTERNET PROTOCOL)

TCP/IP is another reference model. This model is different to the OSI model that we discussed previously. The main difference is the number of layers. In OSI model there are seven layers while in the TCP/IP model there are only four layers. Figure 9.4 shows the corresponding layers of the two models. We discuss these in detail at a later session.

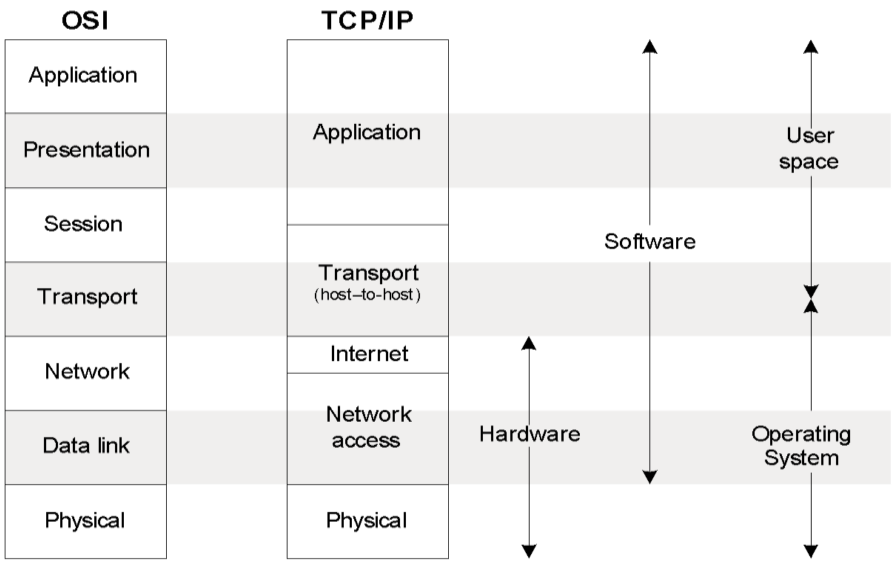


Figure 9.4 Comparison of TCP/IP and OSI reference models.

SUMMARY

In this Session, we discussed the concept of computer networking as well as the different architectures of networking. Finally, we discussed the OSI reference model for data communication in terms of its layers and introduced its deployment in the TCP/IP protocol stack.

LEARNING OUTCOMES

Now you will be able to,

- Describe the concepts of data communications and networking.
- Describe the client server and peer-to-peer architectures.
- Describe the ISO-OSI reference model for data communication.

Session 10
Connecting Devices and Network Topologies

INTRODUCTION

In this session we will study about different connecting devices and different topologies that can be used to arrange the computers in a network to facilitate the communications. When we form a network, it is required to use suitable devices to connect the computers. In different types of networks, the computers are arranged in various ways to facilitate communications.

10.1 CONNECTING DEVICES

A wide variety of connecting devices are used to form networks, varying from NICs, different cables, switches, routers etc. These devices along with wired or wireless transmission media allow the communication among the networked computers.

10.1.1 NETWORK INTERFACE CARDS (NICS)

NIC is installed in an expansion slot of the computer. This card (see *Figure.10.1*) connects the computer to a network, and contains information on the computer's location and also instructions for sending and receiving data over the network.

The NIC has an Ethernet port which connects the computer to the network via an Ethernet cable. NIC can convert the computers' low power signals to high power signals that can be transmitted over the network. Networks transmit data in a serial data format (1 bit at a time), and the data bus of the PC moves data in a parallel format (8 bits at a time). The NIC converts the signal from serial to parallel format or from parallel to serial format, depending on its direction of communication. The speed at which the NIC transmits (or receives) are measured in megabits per second (Mbps).

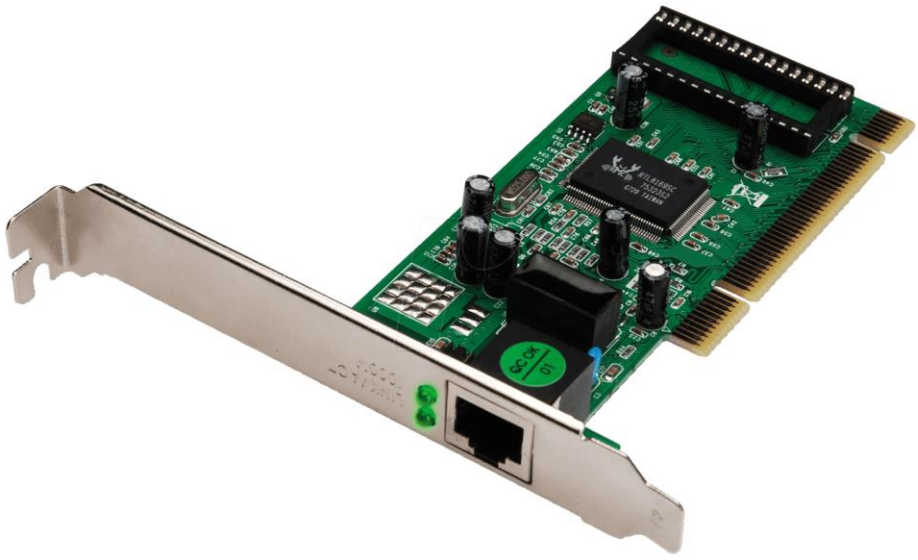


Figure 10.1 Network Interface Card (NIC)

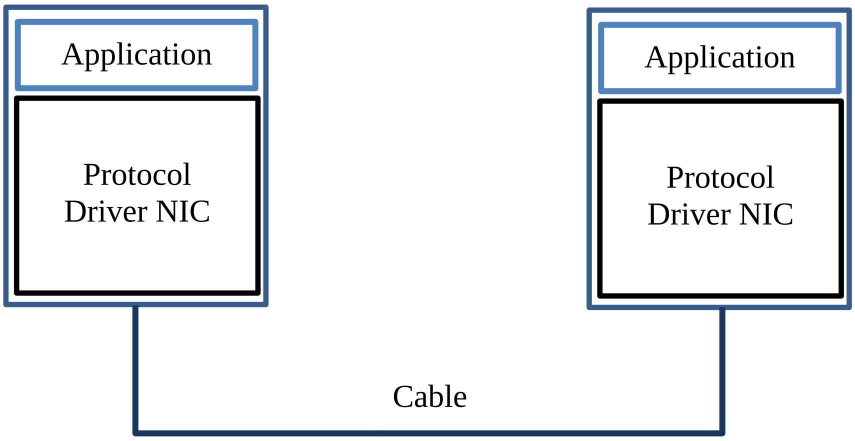


Figure 10.2 Two computers connected with NICs

As shown in Figure10.2, two computers relate to network interface cards and network cables. A device driver (or the "driver") is a specific type of software developed to allow interaction with the NIC to facilitate the communications. It also provides commands for transmitting and/or receiving data from the NIC. Networks connectors provide computers with the ability to transfer data from one computer to another. In addition to the connecting devices, we need a set of rules, called "protocols", which allows all the network's members agree on to perform the communications. Communication Protocols specify how computers interact and exchange messages. A Protocol usually specifies:

- format for data representation
- how to handle errors
- signaling
- authentication procedures.

10.1.2 MODEMS

Modem is the shorten form for Modulator/Demodulator. It connects a computer's serial port to a phone line. A modem converts digital signals (a stream of bits) to analogue signals (a continuous wave) which is the format used in telephone lines. Then the analogue signal is transmitted through the telephone lines.

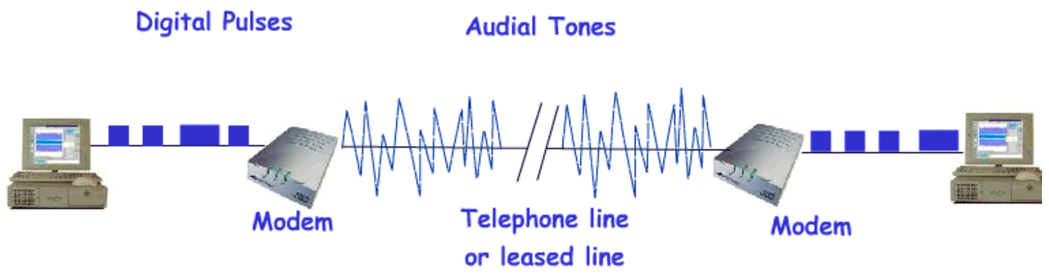


Figure 10.3 How modems function in converting digital bit streams to analogue signals and vice versa

A modem or a telephone line-based interconnection is an alternative to network interface card-based connection. Figure 10.3 shows how external modems are used to connect computers to send and receive data over the telephone network. There are internal modems as well that can be fitted to an expansion slot on the motherboard. Modems have different transmission speeds, measured in bits per second (bps), or in baud.

10.1.3 REPEATERS

A repeater can be used to connect two (or more) network segments to overcome the segment length limitation due to attenuation and pulse distortion (Figure 10.4). A repeater cleans the noise in the signal and amplifies the signal that it transmits from one segment to another. Repeaters simply allow you to extend distance limitations in the network. It does not allow any more bandwidth or faster data transmission. In the event of failures, by disconnecting one side of the repeater it is possible to isolate the associated segment from the network.

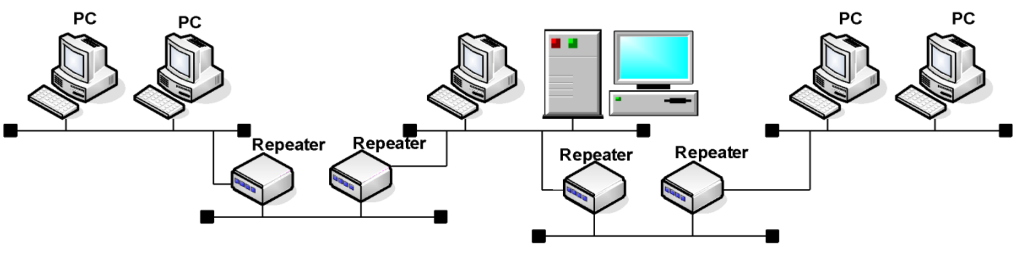


Figure 10.4 Networking with repeaters

10.1.4 BRIDGES

Bridges were designed to interconnect network segments together. Most bridges today support filtering and forwarding functions. The bridge learns about the network and the routes during the initialization process. IEEE 802.1 D specification is the standard for bridges.

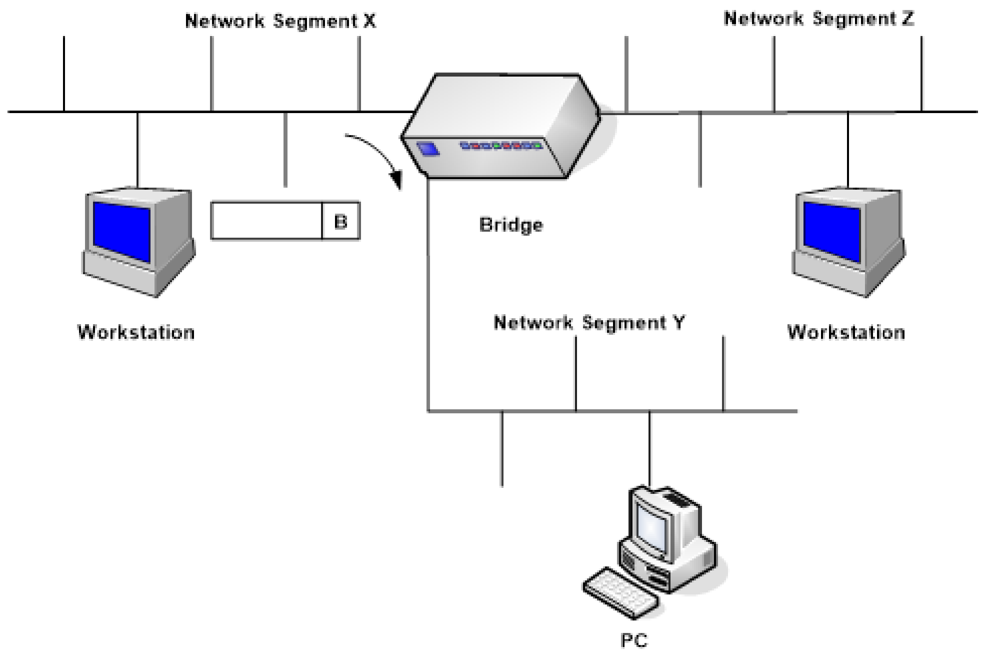


Figure 10.5 A few network segments connected by a bridge

Figure 10.5 shows three separate network segments connected via a bridge. Each segment must have a unique network address so that the bridge can forward packets from one segment to the other.

10.1.5 HUBS

A hub provides several ports. Computers can be attached to the hub using UTP cables as shown in Figure 10.6. Hubs dedicate the entire bandwidth to each port or the device attached.

A few years ago, hubs were widely used to make networks. Ethernet over twisted pair cable (10baseT) was the most common Ethernet technology in the recent past. Twisted pair cables were connected to hubs with the maximum length of cable being 100m. The connectors were ordinary RJ45 jacks.

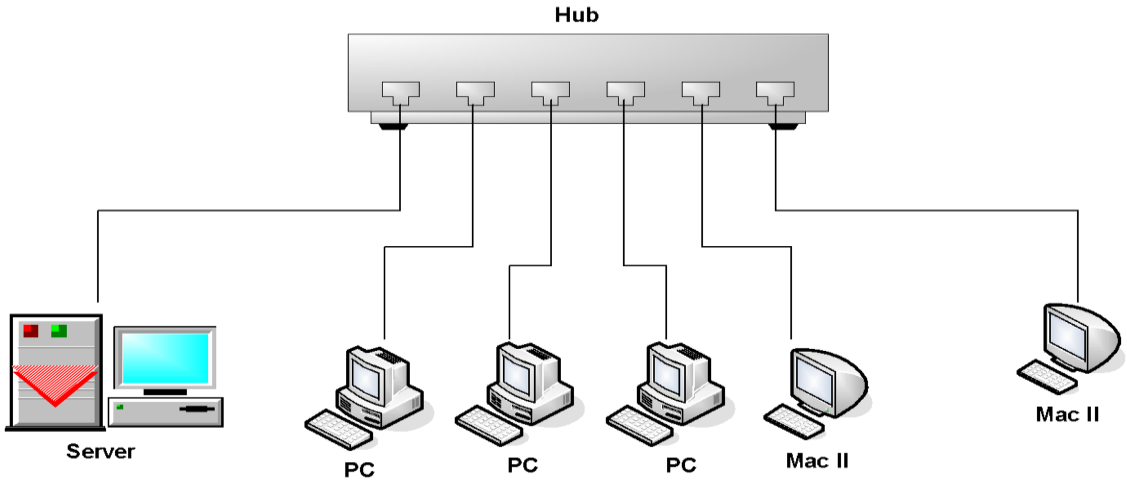


Figure 10.6 Making a network using a hub

Figure 10.7 A picture of a stackable hub

Stackable hubs are inter-connected with short lengths of special cable (Figure 10.7). When they are connected, they act like a modular hub, because each one can be managed as a single unit. Rules for connecting devices on Thin-Ethernet are given below;

- Minimum distance should be 0.5 m between T-connectors (Figure 10.8)
- Maximum cable length should be 185 m (Figure 10.8)
- Maximum number of nodes can be connected is 30

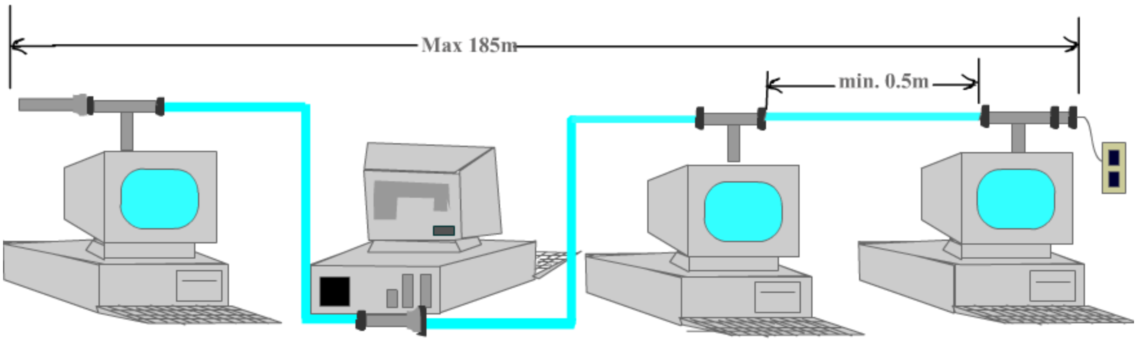


Figure 10.8 Thin Ethernet cable length limits

A cable can be extended by installing repeaters as shown in Figure 10.9, which amplify the signal. A repeater counts on each segment as a node and can be connected at any location in the Thin-Ethernet cable.

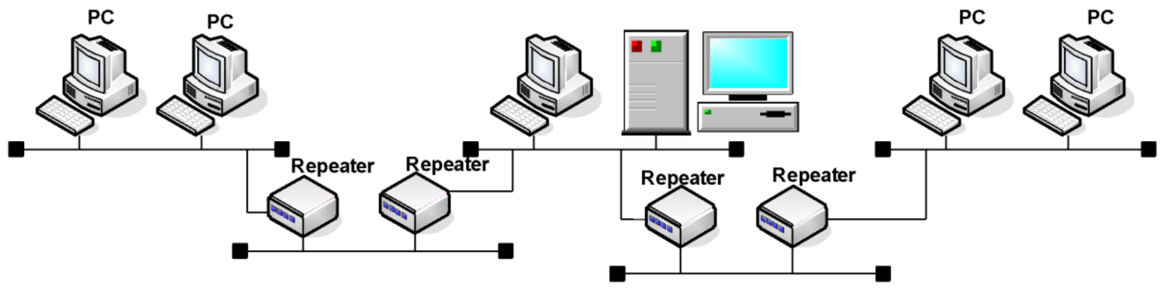


Figure 10.9 Extending a thin Ethernet cable

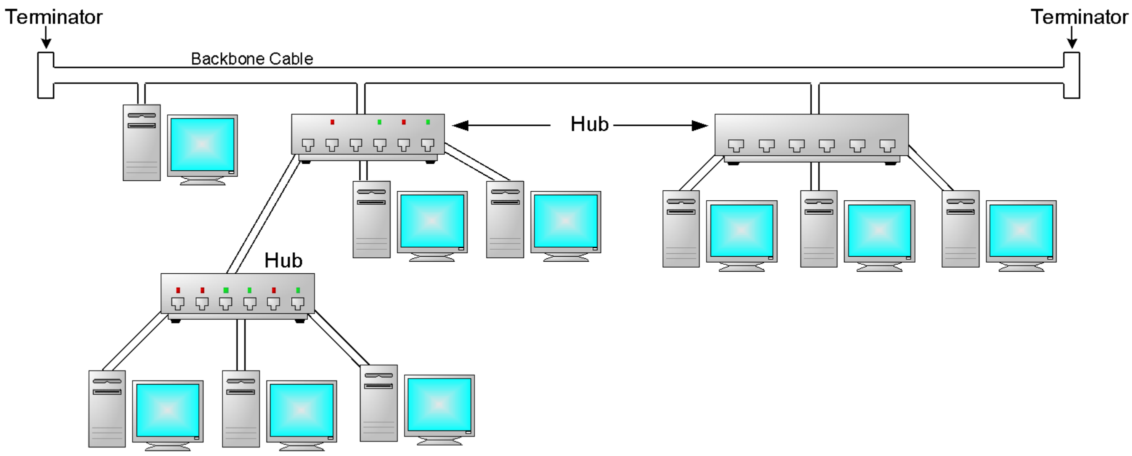


Figure 10.10 Networking with hubs

However, if a network needs more than 2 repeaters the following limitations apply. When an Ethernet signal travels from its source to destination station, it can travel through:

- maximum of 5 segments
- maximum of 4 Repeaters or hubs
- maximum of 3 populated segments (Populated segments have more than 2 nodes connected, un-populated segments have only a node at each end)

Note: For this discussion, a 10BaseT-Hub is considered as a repeater

There can be more repeaters/hubs in the complete network, and an Ethernet signal can pass-by more than 4 Repeaters/hubs, as long it does not have to go through more than 4 Repeaters/hubs. The network becomes unreliable if these rules are violated. A network is also limited in its maximum size, when Repeaters and/or hubs are used to extend a Thin Ethernet (10base2) or Twisted Pair Ethernet (10baseT/UTP) network. For large configurations, a switch may become necessary to optimize network utilisation.

10.1.6 SWITCHES

Switches are similar to bridges except that switches have more than two ports (Figure 10.10-10.12). Switches are totally different from hubs in the way frames are forwarded. A hub (or a repeater) forwards a received frame to all attached equipment i.e. actually broadcasting. A switch, on the other hand, learns the association between the systems MAC addresses and the interface ports in the same way as a bridge, and forwards the frame to only the required segment. Thus, a switch reduces the number of packets on the other LAN segments (and hence the load on these segments), increasing the overall performance of the connected LANs. The switch also improves security, since frames only travel where they are intended to.

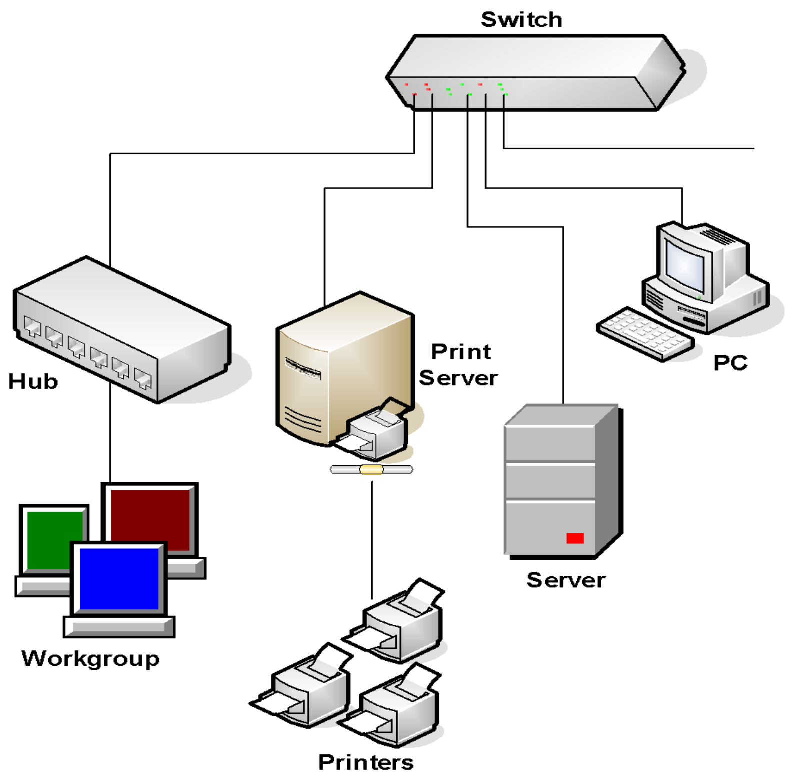


Figure 10.11 Making networks with a switch

Switches work at the MAC layer (2nd layer). There are modern switches which work in Network Layer (3rd Layer) and Transmission Layer (4th Layer). They allow better traffic management and modern networking concepts such as virtual LANs (VLAN). Some layer 2 switches also allow VLANs and traffic management facilities. Modern switches come with an operating system. So, the network administrators can manage the traffic and configure basic settings for a better throughput.

Connecting Bridges and Switches Together

A rule to remember in interconnecting bridges and switches is that, a bridge / switch / hub LAN must form a tree, and not a ring. That is, there must be only one path between any two computers. If more than one parallel path, a loop would be formed, resulting in endless circulation of frames over the loop. This would soon result in overload of the network. To prevent this happening, the IEEE (in IEEE 802.1D) has defined the Spanning Tree Algorithm (STA) which automatically detects loops and disables one of the parallel paths. The Spanning Tree Algorithm may also be used to build fault-tolerant networks, since if the chosen path becomes invalid (e.g. due to a cable / bridge / switch fault), and an alternate path exists, the alternate path is enabled automatically.

10.1.7 ROUTERS

Routers are used to connect two or more networks (Figure 10.12). These networks can be of different types.

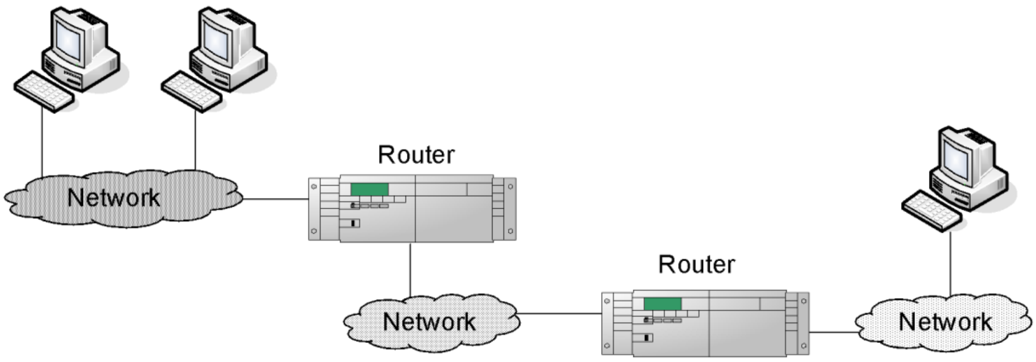


Figure 10.12 Connecting different networks using routers

Routers pass packets to the network segment they are intended for. They work similar to bridges and switches in that they filter out unnecessary network traffic and remove it from network segments.

A router consists of a computer with at least two network interface cards supporting the IP protocol and an operating system. A router can be used separate networks logically based on IP subnets. A router is also used as the main gateway link between an organization and the outside world.

Instead of specifying link hardware addresses (MAC addresses), the router table specifies the network (IP addresses). The routing table has a list of known IP destination addresses with the appropriate network interface to be used to reach that destination. A default entry may be specified to be used by all the addresses that are not explicitly defined in the table. A filter table can be used to ensure that unwanted packets are discarded. The filter may be used to deny access to particular protocols or to prevent unauthorised access from remote computers by discarding packets to specified destination addresses.

A gateway is used to connect networks of different types; For example, a Wireless LAN to the Internet or another Wide Area Network. Gateways are also used to provide access to special services such as email and fax.

Gateways that connect two IP-based networks, like TCP/IP with IPX/SPX, have two IP addresses, one on each network. One is a Local Area Network address, to which traffic is sent from the LAN. The other IP address is the Wide Area Network address, to which traffic is sent coming from the WAN. Instead of a WAN, when it is the Internet, that address is usually assigned by an Internet Service Provider.

Usually a computer can be configured to perform the tasks of a gateway. Gateways are also called protocol converters and can operate at any layer of the OSI model. A protocol translation/mapping gateway interconnects networks with different network protocol technologies by performing the required protocol conversions.

A gateway may contain devices such as protocol translators, impedance matching devices, rate converters, fault isolators, or signal translators as necessary to provide system interoperability. It also requires the establishment of mutually acceptable administrative procedures between the two networks. An example is a system for transferring electronic mail between two dissimilar (and incompatible) networks.

So far, we learnt about the devices that we can use to connect the computers in a network. A formed network can be classified by considering many characteristics. One way of categorizing the network is by considering the size of the area that the network is distributed among. Let’s look at such a classification of network.

10.2 NETWORK TYPES

There are different types of networks:

- Local Area Networks (LAN) is a network typically contained wholly within a single site and entirely owned and controlled by a single organization.
- Wide Area Network (WAN) is a network that spread over a geographically large area. It is characterized by its dependence on *Public Telecommunications Providers* such as Sri Lanka Telecom for either *dial-up* or *dedicated* services.
- Metropolitan Area Networks (MAN) is large network usually spanning a University or a city. They typically use wireless infrastructure or optical fiber connections to link their sites

Features of Local Area Networks (LANs)

- Computers in a LAN are usually close to each other, usually in the same building
- Each computer and peripheral is an individual node on the LAN
- Nodes in regular networks are connected to the LAN via cables
- Nodes in wireless networks connect via radio waves or infrared transmitters
- Computers on a LAN do not have to be the same brand or use the same operating systems, but they have to use the same network standards, such as Ethernet or Token Ring
- LANs in a work area, such as a classroom or a building, are connected to each other with high speed networking connections called the network backbone

Features of Wide Area Networks (WANs)

- Most WANs are private and are owned by the businesses that operate them

- WANs are the network infrastructure connections that are used to connect remote LANs to each other
- A LAN is a node on a WAN
- Typically, WAN connections are slower than LAN connections
- There are several different types of WAN designs. The design of the WAN, also called the topology of the WAN determines the cost, reliability and speed of the network
- WANs can extend over great distances and can span the globe

Network Classification

Table 10.1 shows a summary of how networks are classified according to the distance they cover.

Table 10.1 How networks are classified according to the distance they cover

Location	Distance	Example
Circuit-board	0.1m	Data flow machine multiprocessors
System	1m	
Desk	10m	
Building	100m	LAN
Campus	1km	LAN
City	10km	MAN
Country	10km	WAN
Continent	1000km	WAN
Planet	10,000km	WAN

In a network data is transmitted wired or wireless. Therefore, we should get an understanding of how different transmission media are used for transmission. Next, we will discuss about different wired transmission media such as, twisted pair, coaxial and fiber optics cable. Subsequently, we will briefly discuss about the wireless transmission links.

10.3 TRANSMISSION MEDIA

10.3.1 TWISTED PAIR

Copper wire pairs are the most basic of the transmission media.

In the two wire untwisted pair, the insulated wire conductors run in parallel, often in a moulded, flat cable (Figure 10.13). Normally used over short distances or at low bit rates, due to problems with crosstalk and spurious noise pickup. Performance in multiple conductor cables is enhanced by dedicating every second cable as a ground (zero volt reference), and by the use of electrically balanced signals.

Figure 10.13 Two wire untwisted pair

In the twisted pair cables, the insulated conductors are twisted together (Figure 10.14 and Figure 10.15), leading to better electrical performance and significantly higher bit rates than untwisted pairs. UTP is unshielded, like telephone cable, whilst STP is shielded and capable of higher bit rates. Data rates in unshielded or shielded can go up to 100Mbps now, with a maximum distance up to 100m.

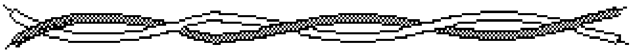


Figure 10.14 Two wire twisted pair

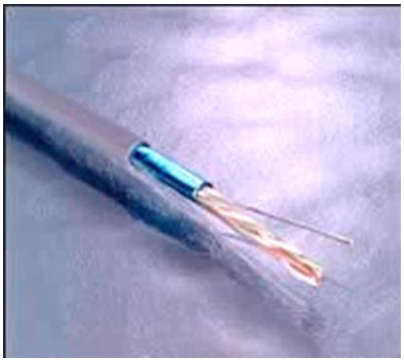


Figure 10.15 Pictures of twisted pair cables

10.3.2 COAXIAL CABLE

The conductors are arranged so that a center conductor is carried inside, and it is insulated from an electrical shielding, with the braided outer conductor (Figure 10.16 and Figure 10.17). Because the electrical field associated with conduction is entirely carried inside the cable, problems with signal radiation are minimized. Here very little energy escapes, even at high frequency. In addition, there is little noise pick up from external sources. Thus, higher bit rates can be used over longer distances than with twisted pairs. Both 50 ohm and 75-ohm characteristic impedance coaxial cables have been used, in a variety of sizes.

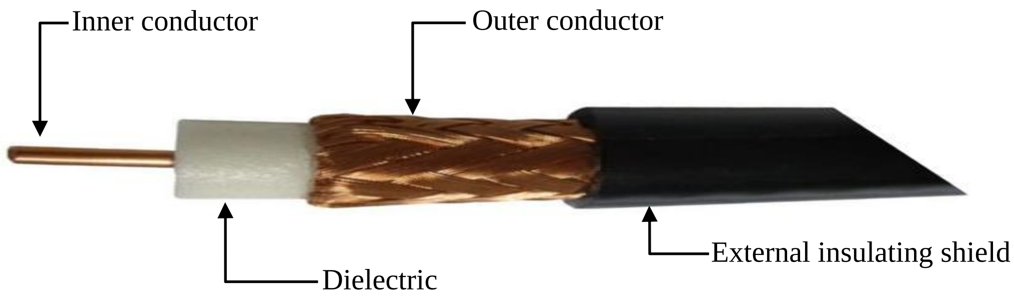


Figure 10.16 Coaxial cable

Baseband

- Thin-wire - requires direct connection to computer
- Thick-wire - typically uses twisted pair drop from transceiver to workstation

Broadband

- Single-cable system: receive/transmit on different bands over same cable
- Dual-cable system: separate cables for receive and transmit
- Headend - acts as an end unit that will rebroadcast in the other direction, typically used for dual cable system

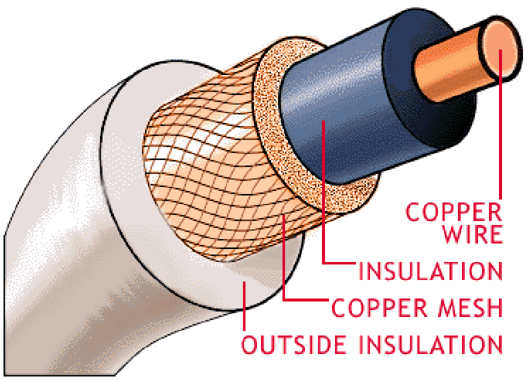


Figure 10.17 Real coaxial cable

Table 10.2 shows EIA/TIA[1] classification system for specifying the performance of UTP, STP and coaxial cables.

Table 10.2 EIA/TIA classification system for specifying the performance of UTP, STP and coaxial cables.

Category	Description
Coaxial Cable	Although widely used for video and RF distribution, in networking this cabling is primarily used for ARCNET and 10BASE-2 Ethernet. Also referred to as thinnet.
Cat 1	Currently unrecognized by TIA/EIA. Voice grade, UTP telephone cable.
Cat 2	Currently unrecognized by TIA/EIA. Data grade UTP. Previously frequently used on 4Mbit/s token ring networks by IBM.
Cat 3	Currently defined in TIA/EIA-568-B, data grade UTP utilizing frequencies up to 16MHz. Historically popular for 10 Mbps Ethernet networks.
Cat 4	Currently unrecognized by TIA/EIA. Data grade UTP. Provided performance of up to 20 MHz and was frequently used on 16Mbps token ring networks.
Cat 5	Currently unrecognized by TIA/EIA. Data grade UTP. Provided performance of up to 100MHz and was frequently used on 100Mbps Ethernet networks. Suitable for 1000BASE-T gigabit Ethernet.
Cat 5e	Currently defined in TIA/EIA-568-B. Data grade UTP. Provides performance of up to 125MHz, and is frequently used for both 100Mbit/s and gigabit Ethernet networks
Cat 6	Currently defined in TIA/EIA-568-B. Data grade UTP. It provides performance of up to 250 MHz, more than double the category 5 and 5e.
Cat 7	An informal name applied to ISO/IEC 11801 Class F cabling. This standard specifies four individually-shielded pairs (STP) inside an overall shield. Designed for transmission at frequencies up to 600 MHz

10.3.3 OPTICAL FIBRES

In optical fibre cables the data is carried as pulses of light from a laser or high-power LED.

Characteristics:

- Optical fibre is non-electrical, hence is completely immune from electrical radiation and interference problems. It has the highest bit rate of all media.
- The fibre consists of an inner glass filament, contained inside a glass cladding of lower refractive index, with an outer protective coating. In a step index fibre, there is a sudden transition in refractive index. A graded index fibre has a gradual transition from high to low index, and much higher performance.
- Most common fibres are multimode Figure 10.18, where the inner fibre is larger than the wavelength of the light signal, allowing multiple paths to exist, and some dispersion to limit the obtainable bit rate. In single mode fibres, the inner fibre is very thin, and extremely high bit rates (several Gbps) can be achieved over long distances.

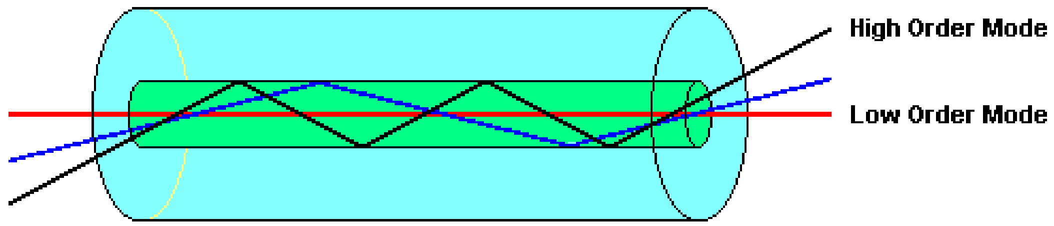


Figure 10.18 Optical fibre cable

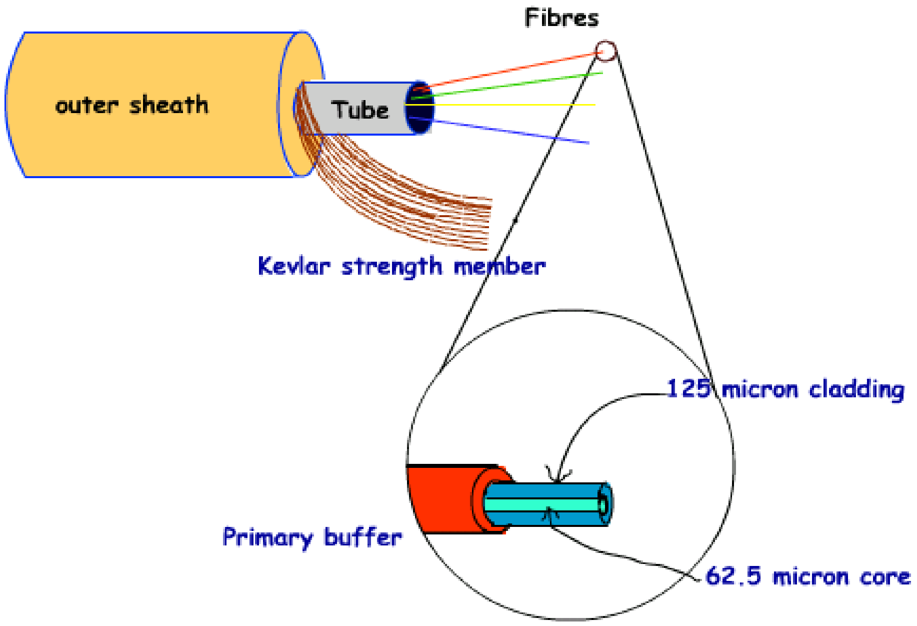


Figure 10.19 How several fibre cables are bundled together

10.3.4 RADIO SIGNALS

Wireless data services use radio signals instead of wires to transmit data back and forth between computers. Use of wireless networks within a localised area, such as within a single floor of an office building or a university library has become very popular.

Wireless LANs use the frequency band 2.4 GHz and 5 GHz and data rates can go from 1 Mbps to 54 Mbps. Compared to pulling fixed cables, wireless systems are cheaper over small areas.

10.3.5 MICROWAVE SYSTEMS

Terrestrial microwave links are high frequency (1 GHz or higher) wireless channels which are point-to-point links with line-of-sight clearance between the transmitter and the receiver. The signal is accurately focused using dish antennas. It can achieve high bit rates over moderately long distances. It is useful for systems in localized areas such as inner-city data systems where cabling would be very expensive.

Satellites links use satellites located in geostationary orbit and fixed receivers on the ground. Bit rates are moderately high.

So far, we have learnt about the connecting devices, different types of network and the transmission media that we can use in different networks. you learnt that LAN as one of the existing network types. Let's learn more about LAN in the next session.

10.4 LAN TOPOLOGIES

As already stated, a topology refers to an arrangement of computers that can be interconnected. Here we will be looking at the three main topologies: Bus, Ring and Star.

Bus topology

In bus topology (Figure 10.20) all networked devices are attached to a single cable or link. Devices are attached to a linear multiport medium where only *half-duplex* operations exist between a device and a bus. Half-duplex is where communication occurs bi-directionally on one cable. This means that a device sending data cannot receive data at the same time. Full-duplex uses two individual cables, one to send and another to receive. This allows a device to send and receive data at the same time.

Ethernet is a LAN standard that has become very popular today for many reasons. It is one of the cheapest and most widely available of the existing LAN technology, and also has the ability to carry high-speed transmission.

Frames travel across the bus topology and a frame provides the address of its destination. If the frame gets to the end of the link and has not found its intended destination, then that frame is lost. In a bus topology, there is no security; every node attached to the line can see the conversations of the other nodes on the link.

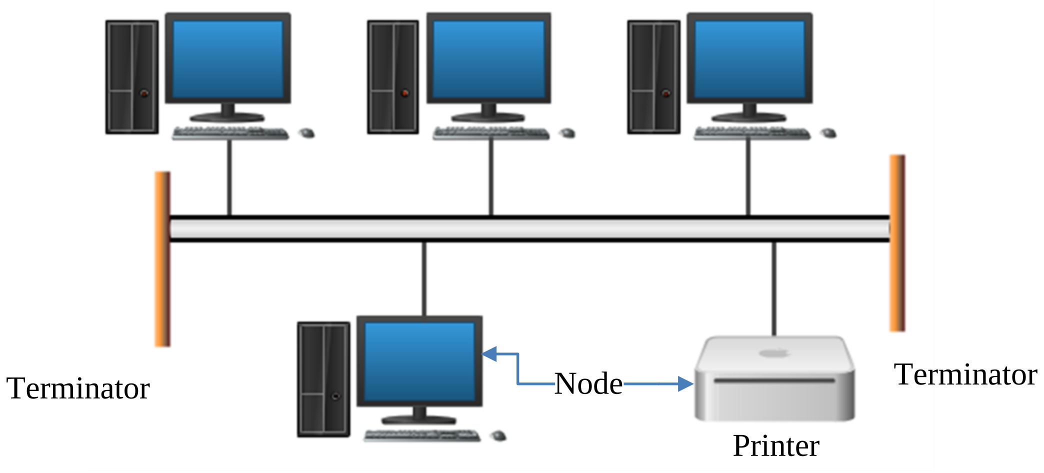


Figure 10.20 Bus topology

Ring Topology

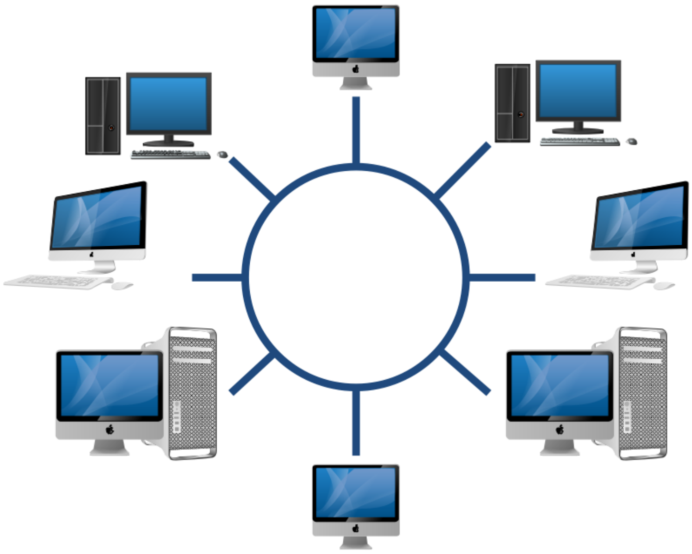


Figure 10.21 Ring Topology

In ring topology (Figure 10.21) LAN (like in a bus topology LAN) all the nodes or devices in the network are attached to the network on the same cable or link. The difference is that ring topology makes a complete circle. Both Token Ring/IEEE and Fibre Distributed Data Interface (FDDI) use ring topology. It can use a single ring for half-duplex operations or a dual–ring architecture for full-duplex operations.

When a break in the ring occurs, such as a cut cable or other cabling problem, it affects all the stations. This means that none of the stations connected can receive or transmit data. In a ring-topology network, centralized access means that faults are easy to detect and isolate. Multiple rings are sometimes used to make a very robust and reliable network.

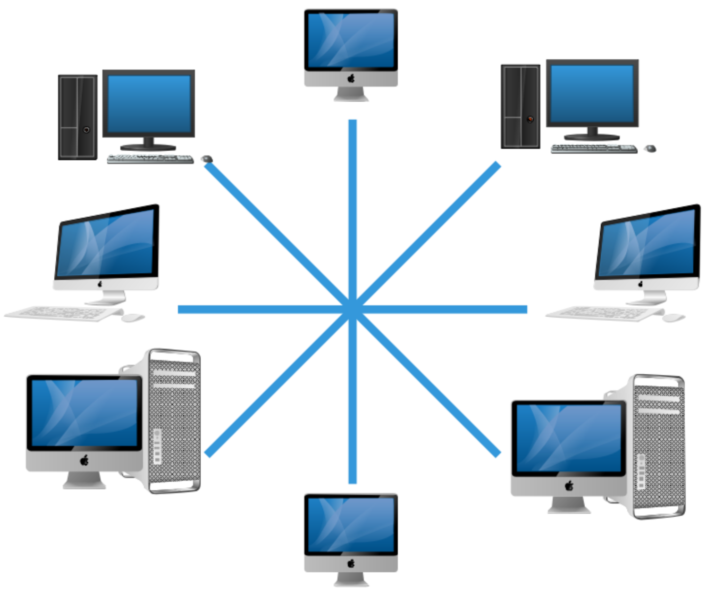


Figure 10.22 Star topology

The star topology (Figure 10.22) is the most common topology in today's networks, and includes Ethernet, Fast Ethernet, and Gigabit Ethernet. Each node in a star topology connects to a dedicated link where the other end connects to a switch or a hub. In the star-topology network, multiple devices are connected to a switch or a hub. One of the main reasons to use a star topology is that the loss of any one node will not disrupt network operations. It is also easy to add or remove a node from the network. From wiring to installation, it is particularly easy to set up a star topology network. So far you learnt about different topologies that can be used in LAN. Let's us now look at the Data link layer protocols that involve in Data communication in the next session.

10.5 MEDIUM ACCESS CONTROL PROTOCOLS

Medium access control protocols are used to provide fair, shared access to the all attached stations of a network.

Three common types are,

- CSMA/CD - Carrier Sense Multiple-Access with Collision Detection. This is basically: listen, if no-one is talking then talk. If interrupted stop and try again later.
- This is the medium access control protocol in Ethernet networks.
- Token - person that has the token is the only one who is allowed to talk. This is the medium access control protocol used in networks having ring topology.
- Slotted - this is similar to time division multiplexing Time slots are assigned in which one can speak.

Carrier Sense Multiple Access with Collision Detection (CSMA/CD)

What are the meanings of these words?

Carrier Sense - consider a device connected to an Ethernet network wants to send data. It first checks to make sure that it has a carrier on which to send its data (usually a copper cable connected to a hub or another machine).

Multiple Access - This means that all machines on the network are free to use the network whenever they like as long as no one else is transmitting.

Collision Detection – when two machines start to transmit data simultaneously, the resultant corrupted data is discarded, and re-transmissions are generated at differing time intervals.

A Shared Medium and Carrier Sense Multiple Access (CSMA)

Consider an Ethernet network that provides shared access by a group of attached computers to the physical medium which connects them. When a computer sends a data frame to another, all computers on the medium receive it physically. However, the frame header contains the destination address which ensures only the specified destination actually accepts the received data. Other computers discard the data frames which are not addressed to them.

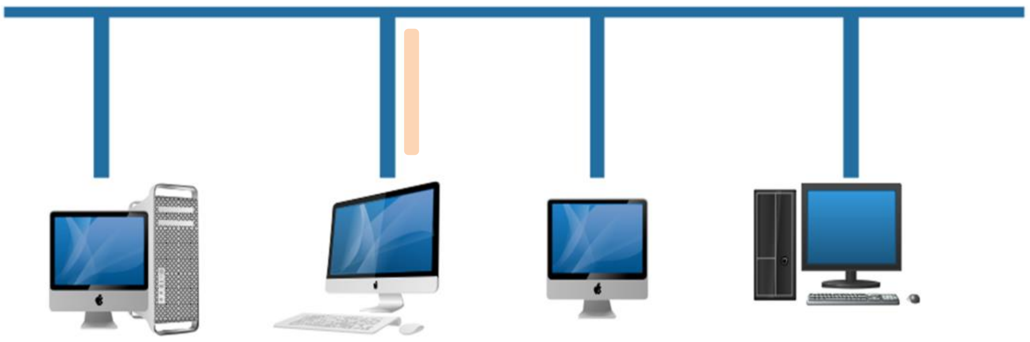


Figure 10.23 A computer sending data packet to the shared medium

Here is a LAN with four computers each with a Network Interface Card (NIC) connected by a common Ethernet cable:

Second computer from the left sends a data frame to the shared medium (Figure 10.23). The data packet has a destination address corresponding to the source address of the NIC in the third computer from the left.

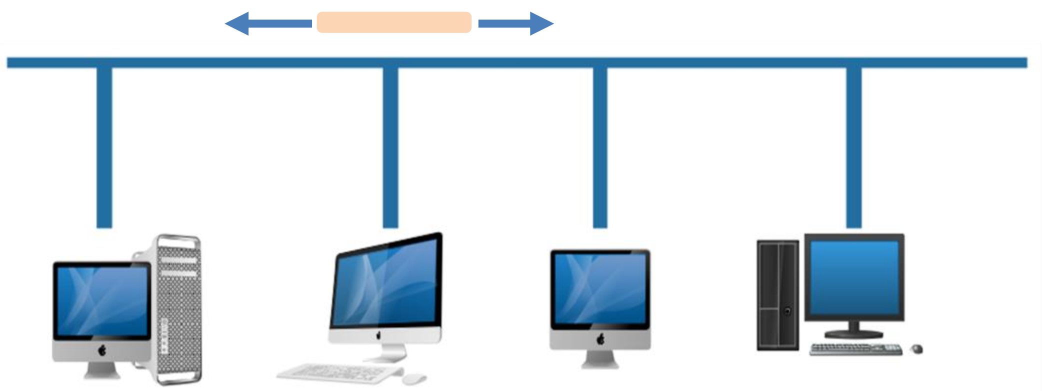


Figure 10.24 The data packet propagate in both ways.

The cable propagates the signal in both directions (Figure 10.24), so that the signal reaches the NICs of all four computers. Termination resistors at the ends of the cable absorb the signal energy, preventing reflection of the signal back along the cable.

All the NICs receive the frame (Figure 10.25). They examine the header destination address to see if the frame should be accepted.

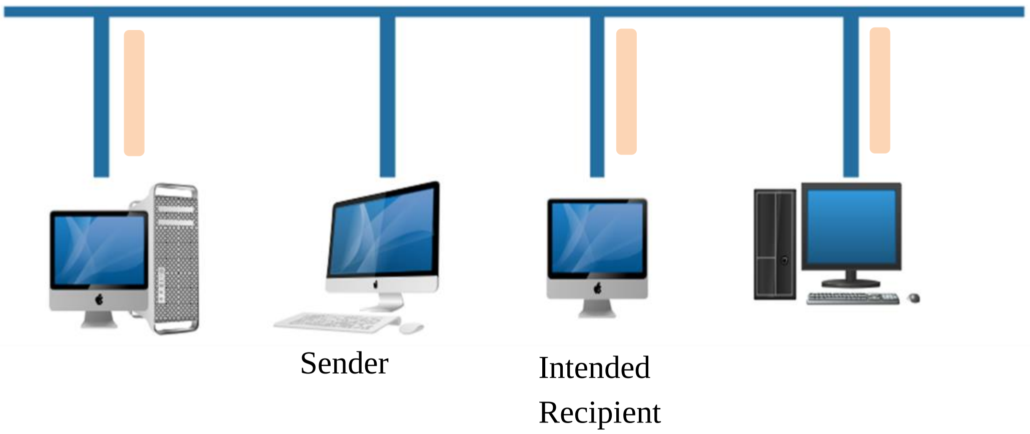


Figure 10.25 The data packet is received by all computers attached to the medium

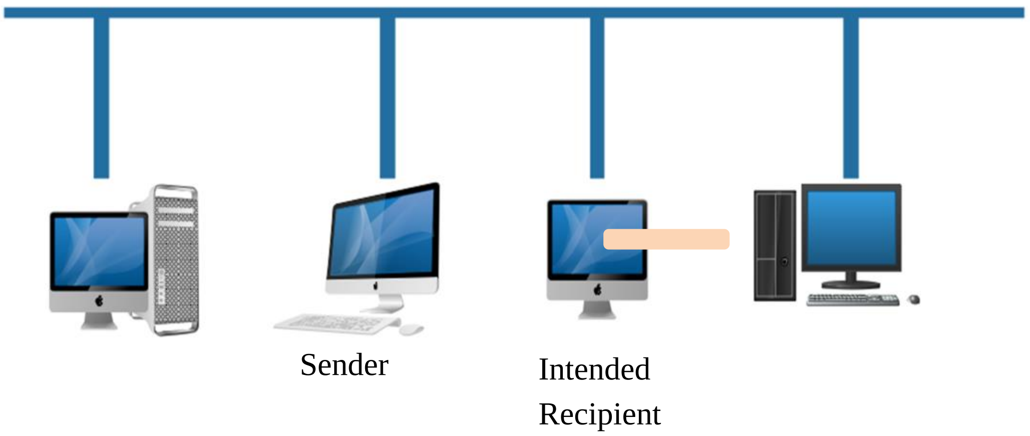


Figure 10.26 Except the intended recipient other computers discard the packet.

Only the NIC in the third computer from the left recognizes the frame destination address as its, and forwarded the frame to the network-layer software in the computer (Figure 10.26). Other computers discard the unwanted data packet.

The shared cable allows any NIC to send frames whenever needed, but if two NICs happen to transmit at the same time, a collision will occur, resulting in the data being corrupted.

Carrier Sense Multiple Access (CSMA)

When a NIC has data to transmit, it checks the cable to see if a signal is being transmitted by another device. If the cable is idle (i.e. no current is present) data transfer can begin. All NICs always listen to see if other NICs have started to transmit data to it.

However, there is a problem in this setup. If two NICs try to transmit at the same time, then both could see an idle physical medium. So both will conclude that no other NIC is currently using the medium and may start transmitting, resulting in a

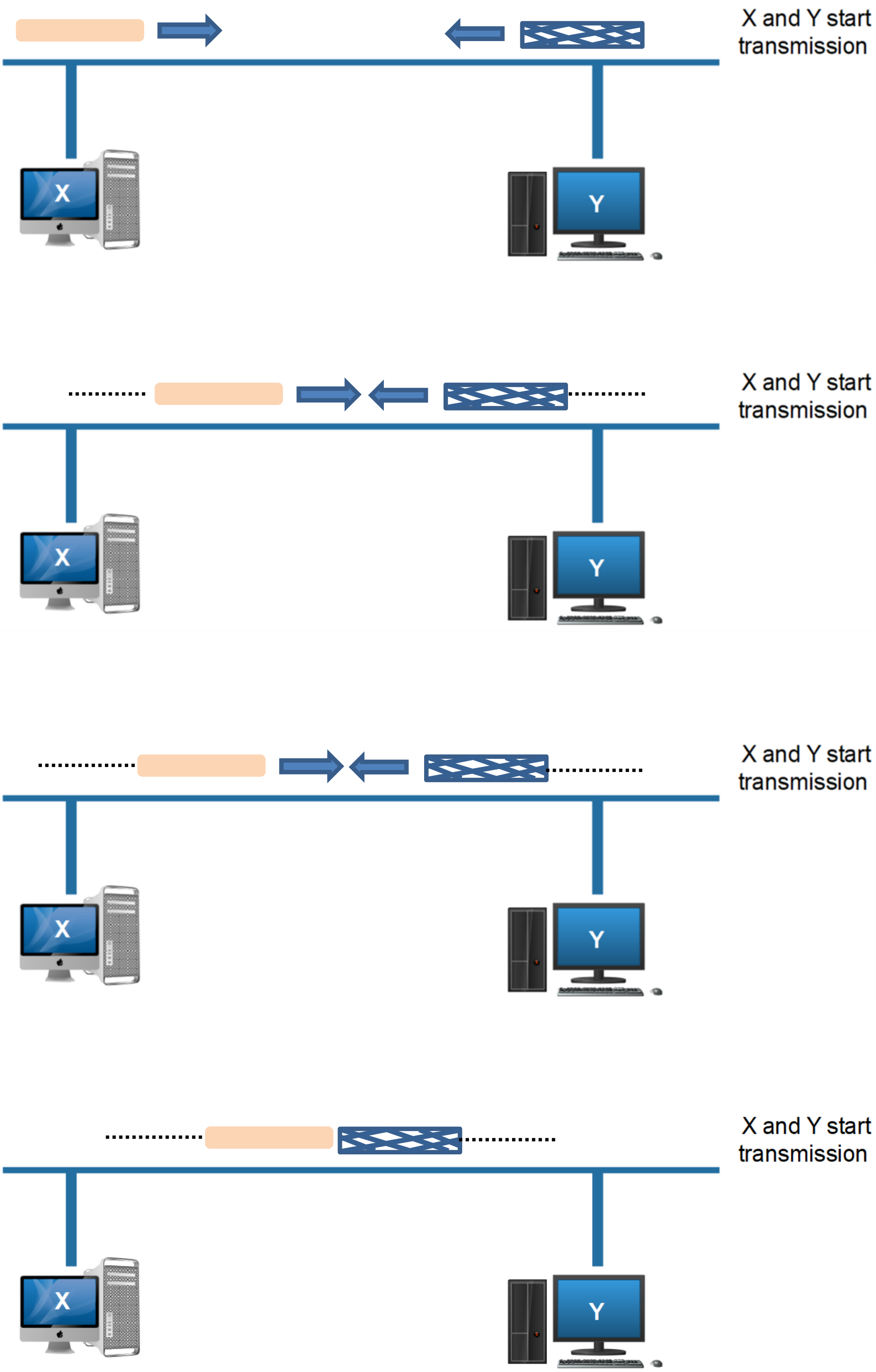


Figure 10.27 Collision detection

collision (Figure 10.27). Collided frames will be corrupted and will subsequently be discarded by the receiver.

Collision Detection (CD)

Collision detection is the second part of the Ethernet access protocol. It is used to detect when a collision occurs. Upon detecting a collision the node waits for a random time period, which is referred to as *random exponential back off time* and then retries to transmit data.

When the NICs listen to the cable and one of them observes a collision (excess current above what it is generating, i.e. > 24 mA for coaxial Ethernet), it stops transmission immediately and instead transmits a 32-bit jam sequence (Figure 10.28). The purpose of the jam sequence is to ensure that any other node which may currently be receiving this data packet will receive the jam signal in place of the corrupted data packet. This causes the other receivers to discard the data packet.

After a period, equal to the propagation delay of the network, the NIC at X detects the other transmission from Y, and is aware of a collision. Y continues to transmit, sending the Ethernet Jam sequence (32 bits).

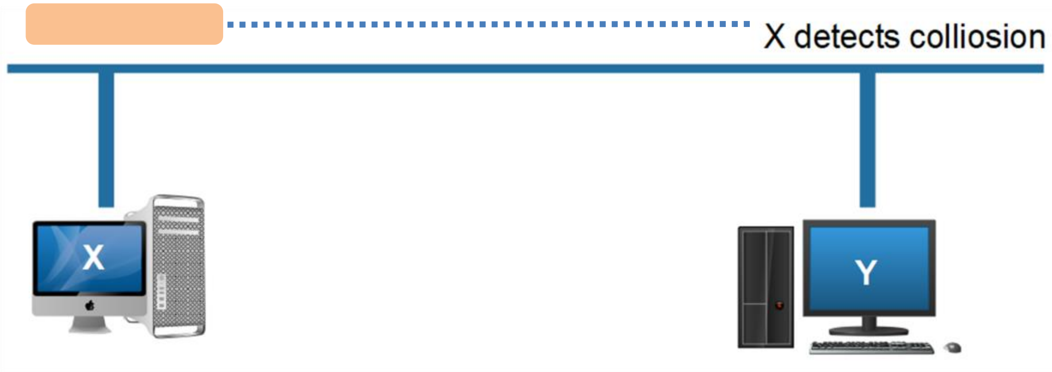


Figure 10.28 Sending the jam signal

After one complete round trip propagation time, both NICs are aware of the collision. When Y ceases the transmission of the Jam Sequence, X also will continue to transmit a complete Jam Sequence. Finally, the cable becomes idle.

To ensure that all NICs start to receive a data packet before the transmitting NIC has finished sending it, Ethernet defines a minimum data packet size. The minimum data packet size is related to the distance which the network spans, the type of media being used and the number of repeaters which the signal may have to pass through to reach the furthest part of the LAN.

Performance of CSMA / CD

It is simple to calculate the performance of a CSMA/CD network where only one node attempts to transmit at any time. In this case, the NIC may saturate the medium and near 100% utilisation of the link may be achieved, providing almost 10 Mbps of throughput on a 10 Mbps LAN.

However, when two or more NICs attempt to transmit at the same time, the performance of Ethernet is less predictable. The fall in utilisation and throughput occurs because some bandwidth is wasted by collisions and back-off delays. In practice, a busy shared 10 Mbps Ethernet network will typically supply 2-4 Mbps of throughput to the NICs connected to it.

Protocol Operation

- DTE listens on the network for a carrier signal (is anyone talking?)
- If no signal found, then start message
- During transmission of message listen for a collision
- If collision found, then stop transmission
- send a jamming signal so that all detect the collision, wait and try again
- amount of wait is random interval with low probability of second collision

If there are many NICs competing to share the bandwidth, an overload condition may occur. In this case, the throughput of Ethernet LANs reduces drastically, and much of the capacity is wasted by the CSMA/CD algorithm. This is the reason why a shared Ethernet LAN should not connect more than 1024 computers.

10.4.3 TOKEN RING

Protocol Operation

At the start, a free Token is circulating on the ring, this is an empty frame. To use the network, a machine first has to capture the free Token and replace the data part with its own message. Figure 10.29 illustrate the Token Ring operation.

Token is a bit sequence - Busy token: 01111111, Free token: 01111110

- When a node wants to transmit
- Wait for free token
- Remove token from ring (replace with busy token)
- Transmit message
- When transmitting finished, replace free token on ring
- Nodes must buffer 1 bit of data so that a free token can be changed to a busy token

Token Ring Self Maintenance

When a Token Ring network starts up one machine will become the 'Active Monitor' and take control of the ring. Usually it is the machine with the highest MAC address who is participating in the ring, and all others become 'Standby Monitors'.

The ring polling allows all machines on the network to find out who is participating in the ring and to learn the address of their Nearest Active Upstream Neighbour (NAUN). Ring purges (reset) the ring after an interruption or loss of data is reported.

In the figure, machine A wants to send some data to machine C. It captures the free Token first. Then it writes its data and the recipient's address onto the Token.

The packet of data is then sent to machine B who reads the address and realizes that it is not its own and passes it on to machine C.

This time it is the correct address so machine C reads the message. It cannot, however, release a free Token on to the ring. It must first send an acknowledgement to machine A to say that it has received the data.

The acknowledgement is then sent to machine D who checks the address, realizes that it is not its own and so forwards it on to the next machine in the ring, the E.

Machine E does the same and forwards the data to F and then to A. Machine A recognizes the address, reads the acknowledgement from number C; and then releases the free Token back on to the ring ready for the next machine to use.

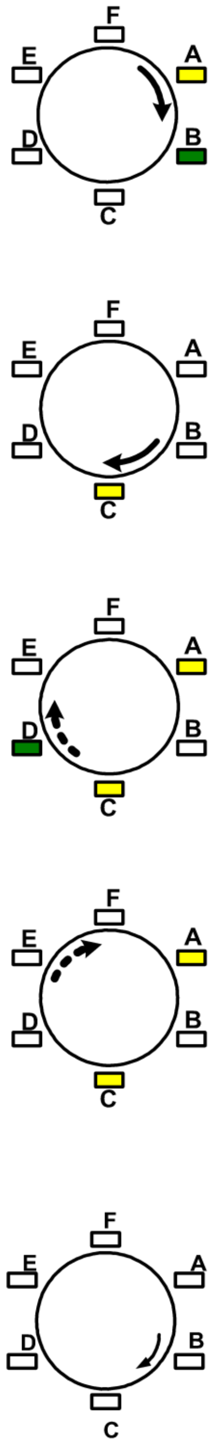


Figure 10.29 How token ring works

Comparison of LAN Technologies

Table 10.3 shows a summary of Ethernet and Token Ring technologies discussed so far.

Table 10.3: Comparison of LAN technologies

Property	Ethernet	Token-ring
Topology	Bus / Star	Ring
Access Method	CSMA/CD	Token Passing
Speed (in Mbps)	10/100/1000	1/4/16/100
Broadcast/Non-Broadcast	Broadcast	Broadcast
Packet Size (Bytes)	64 – 1515	32 – 16K
Self-Recovery	No	No
Data Path Redundancy	No	No
Predictable Response Time	No	Yes
Priority Classes	No	Yes
Cost of Deployment	Cheap	Moderate
Typical Deployment Environment	Small Offices, Educational Institutes, Corporate Offices, e-commerce	Airline, Manufacturing Floor, Banking, Mission Critical Networks

10.6 WIDE AREA NETWORK

A Wide Area Network (*WAN)* is a data communications network that spans a relatively large geographical area. It may consist of number of interconnected LANs and other networks. WANs often use transmission facilities provided by common carriers, such as telephone lines. WANs are also constructed using leased lines and satellites. Internet is the largest WAN and may be the best example.

WAN technologies generally function at the lower three layers of the OSI reference model: the physical layer, the data link layer, and the network layer. WANs are built mainly using leased lines and sometimes with circuit switching or packet switching methods or cell relay.

In this session you will learn the basics of packet switching, circuit switching, X.25 Protocol, Virtual circuits, and relatively new communication technologies like ISDN, Frame Relay and ATM.

LEASED LINES

Leased line is a full-time dedicated link from one point to another that is always available. It has a fixed monthly cost irrespective of usage.

Most are digital 64kbps – 2Mbps

CIRCUIT SWITCHING

Circuit Switching allows data connections that can be initiated when needed and terminated when communication is complete. This works similar to voice communication over a normal telephone line.

An example for circuit switching is Integrated Services Digital Network (ISDN).

PACKET SWITCHING

In packet switching, data is broken into small packets and packets are sent over the network. Each packet may follow a different path over the network. A virtual circuit can be created between customers' sites by which packets of data are delivered from one to the other through the network. If the data packets travel the same path to the destination, this is called virtual circuit, if packets can travel any path, not necessarily the same as each other, this is called datagram. Some examples for packet-switching networks are X.25, Asynchronous Transfer Mode (ATM), and Frame Relay. You will learn about X.25and Frame Relay in detail at a higher-level courses. Next, let us discuss about the Asynchronous Transfer Mode (ATM).

ASYNCHRONOUS TRANSFER MODE (ATM)

ATM (asynchronous transfer mode) is a dedicated-connection switching technology where data is transmitted and switched in small units called cells. Cell size is fixed to 53 bytes (48 bytes of data plus a 5 byte overhead) and it is transmitted over a physical medium using digital signal technology. ATM is designed for handling large amounts of data across long distances using a high-speed backbone approach. Cells are individually processed asynchronously relative to other related cells and are queued before being multiplexed over the transmission path. No virtual circuit is allocated for the duration of the transmission. One problem with other protocols which implement virtual connections is that some time slots are wasted if no data is being transmitted. ATM avoids this by dynamically allocating bandwidth for traffic on demand.

Because ATM is designed to be easily implemented by hardware (rather than software), faster processing and higher switch speeds are possible. Usual bit rates are either 155.520 Mbps or 622.080 Mbps but ATM networks can reach 10 Gbps. ATM is suitable for transmitting voice as well as data and video. It is intended to use optical fibre as the transmission medium. The main difference between ATM and IP protocols is that ATM is connection-oriented while IP is connectionless.

SUMMARY

Different network devices are used to connect nodes or terminals such as computers, printers, servers and multiple network segments together. A brief description of such devices repeater, router, hub and switch were given in this session. When connecting multiple nodes together you must select suitable media and the network should adhere to a topology, a set of rules and regulations. This session briefly discussed the transmission media in use including different cables, twisted pair, fibre optics etc., different network topologies, star, bus, ring etc., and the network types, local area network, wide area network etc. Finally, you were introduced to few key terms, technologies widely used in data communication namely circuit switching, packet switching and asynchronous transfer mode (ATM).

LEARNING OUTCOMES

- Describe different types of networking devices
- Explain different types of transmission media in use
- Describe different types of network topologies
- Describes the operation of different medium access control protocols.

Session 11
INTERNET

INTRODUCTION

Internet is a network of interconnected heterogeneous networks. Internet allows you to access to a whole resource of data and information stored at different **hosts** or **servers** all over the world. Internet began as a U.S. Department of Defense (DoD) funded project to interconnect DoD- research sites (funded by DoD) in the U.S. in 1960's.

The communication links that interconnect these host computers use a common protocol known as TCP/IP. TCP/IP stands for Transmission Control Protocol/Internet Protocol. TCP/IP in use today, was also originated at the US Department of Defense in the late 1960's. Widely used operating systems offer support for TCP/IP.

This session introduces the terms associated with Internet and TCP/IP, describe the Internet, TCP/IP protocols and IP addressing.

11.2 INTERNET ADMINISTRATION

Internet has neither a single owner nor central operator. A few authorities are there to manage things that should be managed centrally, such as addressing, naming, protocol development, and standardization.

- IAB - Internet Architecture Board (formerly Internet Activities Board) oversees standards and development of the internet. IETF and IRTF are the two primary task forces of IAB.
- IETF Internet Engineering Task Force - working groups in IETF have primary responsibility for the technical activities of the Internet, including writing specifications and protocols
- IRTF Internet Research Task Force – Works on long term research projects
- ISOC Internet Society – promote the use of the Internet for communication and collaboration. It provides a forum for the discussion of issues related to the administration and evolution of Internet
- IANA Internet Assigned Numbers Authority - a group in the Internet community that managed IP address space, and all TCP/IP-related numbers historically. IANA and IP addressing, both have undergone many changes since their formulation.

11. 3 DOMAIN NAMES AND ADDRESSING

Host - Any computer connected to the Internet can be a host. A host computer provides information for other people to access and retrieve.

Server - means a program implementing a service (e.g. the Apache WWW server) as well as the host on which the service runs. It is possible to run several server programs on one server host.

Hostname - It is the name assigned to the computer on which you install software.

Domain Name - Domain name is a hierarchical naming system to identify the domain a host belongs. It is usually an abbreviation of Department name, University name or an organization name. Hierarchical structure (Figure 11.1) of domain names is best understood if the domain name is read from right-to-left. The Top-level domain name use the two-letter country codes (defined in ISO standard 3166) unless it has US defined top level domain such as .net, .com, .edu etc.

e.g. Consider the example - OUwebServer.ou.ac.lk.

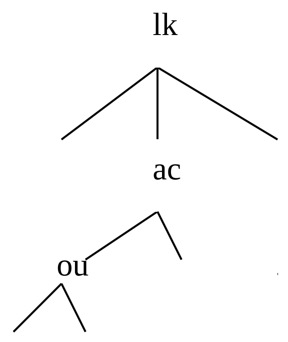


Figure 11.1 Hierarchical naming system of a domain

IP Address – Ipv4 address is a 32 bit number separated into 4, 8-bit. It is a unique number for each device on the network. IPv4 address is given in dotted decimal notation e.g. 192.248.72.3

IPv6 address is 128 bit in length and has a colon hexadecimal notation e.g. 2000:0db8:85a3:0000:0000:8a2e:0370:7334

Domain Name Server (DNS) - Domain Name Server converts the fully-qualified domain names into IP addresses.

Gateway Address is the IP address of the router that connects an internal network to another external network.

Every interface between a host or router and the Internet has an **IP address.** IP address appears in the header of an IP datagram and is used by routers to determine to where a datagram should be sent next. Normally a client starts with a **domain name** like www.slt.lk. This must be mapped to an IP address. The Internet has a **Domain Name Servers** (DNS) to perform this task automatically. To do this task, the DNS server sends a message to the ISP's Domain Name System server asking for a mapping. If this is successful, the client now has an address which can be put in the header of an IP datagram which can be sent towards its destination.

The solution to the name-mapping problem is to maintain a table that maps names to addresses and to make this available across the Internet. Since it makes no sense to have a single server for the whole Internet, this service is distributed among multiple servers. Local DNS servers hold mappings for local machines and answer queries for these directly. Non-local names are passed on to other DNS servers.

ARP (Address Resolution Protocol) – ARP is used to find out the physical address of a network node (node could be a computer, a printer etc) when the IP address is known. A host broadcasts a request for the physical address (i.e. the layer 2 MAC address) of another host on the network. As the IP address (layer 3 address) is given, the target host responds to the requesting host.

11.4 TCP/IP (TRANSMISSION CONTROL PROTOCOL/INTERNET PROTOCOL)

TCP/IP is a layered model for the set of protocols employed in the Internet. Figure 11.2 below is a comparison of TCP/IP and ISO-OSI protocol stacks.

Transmission Control Protocol (Transport layer in TCP/IP) converts a message into smaller packets that are transmitted over the Internet and when it is received by a TCP layer it assembles the packets to form the original message. It facilitates data transmission while shielding upper layers from the details of intervening networks. Internet Protocol (Internet layer in TCP/IP) manages the address part of each packet to make sure it gets to the right destination.

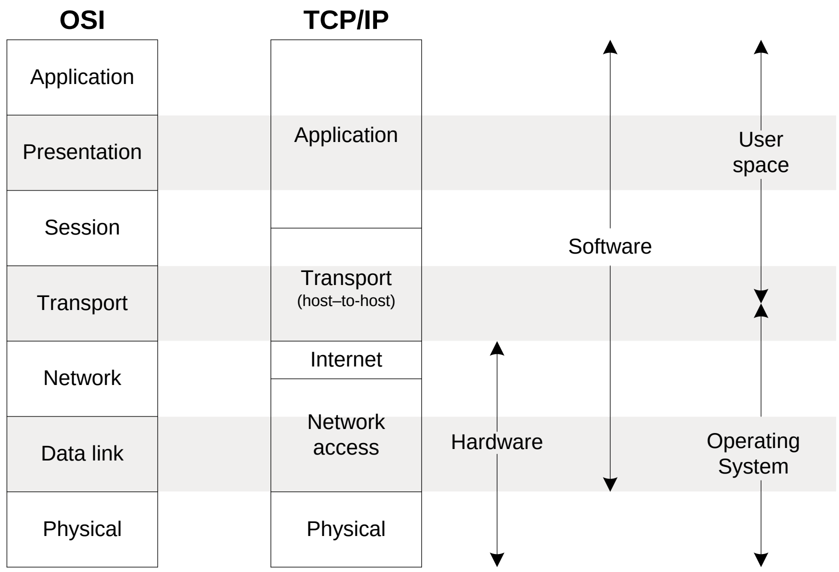


Figure 11.2 Comparison of TCP/IP and OSI protocol stacks

11.5 ADDRESSING WITH TCP/IP

For clarity let us consider IP V4. In IP V4 IP address is a 32-bit address used to uniquely identify a TCP/IP host. The address has two parts called the network ID and the host ID.

- Network ID - identifies all hosts that are on the same logical network
- Host ID - identifies the host. Hosts can be workstations, Servers, Routers, etc.
- Depending on the number of bits assigned for the network id and the host id, there are different IP address classes. A summary f the different IP address classes and the bit allocations are shown in Table 11.1.

Consider an IP address as 24.128.102.7. To determine the network ID we can use a subnet mask. The first step is to convert the IP address to its binary form. Each IP address class is assigned with a subnet mask. Using the subnet mask with an XOR operation on the binary digits of the address, blocks a part of the IP address to reveal the network ID from the Host ID. This will determine if the TCP/IP clients are on the same network or on a remote network. If a packet is determined not to be on the same network, it is sent to the default gateway which is usually a router.

Hosts communicate by Media Access Control (MAC) address. If a MAC address is not known then an ARP broadcast is sent out. The destination hardware will respond with its MAC address and its IP address and these are stored in the ARP cache. The ARP cache is always checked before sending an ARP broadcast.

11.5.1 UNDERSTANDING THE IP ADDRESS

As we discussed before, the 32-bit IPv4 address consists of four 8-bit fields separated by dots. Each set of 8-bits represents a number between 0 and 255. To understand the addresses you must look at them in binary form.

As an example, let us take 192.248.72.3

In binary form this IP address would translate to:

192 -> 11000000 248 -> 11100100 72 -> 01010000 3-> 00000011

So the IP address in binary form would be 11000000.1100100.01010000.00000011

The Network portion of the IP is on the left side. The host portion of the ID is on the right side. Next, let us determine the number of available networks. This is determined by counting the available bits in the network address portion and raising it to the power of 2. For example, using 7 bits 2^7 networks are available for Class A. By rule the all (0)'s and all (1)'s networks are not used. Therefore, the actual number of networks available for the Class A is 2^7-2.

Similarly, Class B has 2^{14}-2 networks and Class C has 2^{21}-2 networks. Number of Hosts is derived using the same formula as the number of networks. Class A network uses 8 bits for the Network ID leaving 24 bits for the Host ID i.e. by the formula 2^{24}-2 we get 16777214. The number of Hosts for Class B and Class C are calculated the same way (see Table 11.1).

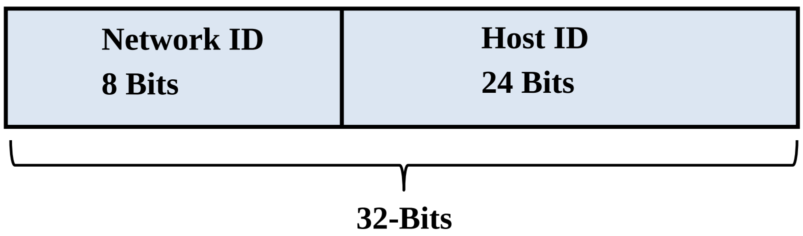


Table 11.1 Subnet mask, number of networks and hosts available for different IP classes.

IP Address Class	Decimal Range	No. networks available 2^x-2	No. of hosts available 2^y-2	Subnet Mask
Class A	1 to 126	126	16777214	255.0.0.0
Class B	128 to 191	16382	65534	255.255.0.0
Class C	192 to 223	2097150	254	255.255.255.0
Class D	224 to 239	Reserved for multi-casting		
Class E	240 to 255	Reserved for research/experimental		

11.5.2 SUBNETTING

Depending on the network architectures if IP addresses were to be given to uniquely identify all the connected devices, the iPv4 address space would have been used up a very long time ago. However, the concept of subnetting was put forward to efficiently reuse the addresses without compromising the unique identification of the devices in a network. Subnetting is the process of borrowing bits from the Host ID to the network ID. This process allowed organizations to gain extra subnets(networks) without requesting additional network numbers from the Internet.

Example: One class A subnet is 24.x.x.x. This means you have 126 networks and 16777214 hosts available. If you need more networks what can you do? There are two solutions. You can either get another IP or you can subnet yours.

Network ID	Host ID

Having only network ID and host ID without subnetting

Network ID	Subnet ID	Host ID

network ID and host ID with subnetting

Default Subnet Mask for Class A networks

11111111 . 00000000 . 00000000 . 00000000 = 255.0.0.0

A subnetted Class A ID. 3 bits were borrowed from the host ID to use as the subnet ID.

11111111 . 11100000 . 00000000 . 00000000 = 255.224.0.0

Using the 3 bits that was borrowed, the possible combination of bits is 000,001,010,011,100,101,110,111. The 000 and 111 are excluded by rule which leaves 6 subnets.

This follows the formula 2^x-2. Borrowing 3 bits gives you 2^3-2 or 6 subnets.

Table 11.2 shows the different subnet masks for class A and the number of networks possible.

Table 11.2 Subnet masks are derived for class A

Subnet Mask(decimal)	**Subnet Mask(binary)**	**No. of Subnets**
192	11000000	2
224	11100000	6
240	11110000	14
248	11111000	30
252	11111100	62
254	11111110	126
255	11111111	254

How do you know whether two IP addresses are on the same subnet? You can use logical "AND" with the subnet masks and the required IP addresss.

Example: 192.248.112.3 mask 255.255.254.0 and 192.248.113.3 mask 255.255.254.0

11000000. 11100100.01110000.00000011 (.112.)
AND 11111111.11111111.11111110.00000000
11000000.11100100.01110000.00000000

```
        11000000. 11100100.01110001.00000011 (.113.)
AND     11111111.11111111.11111110.00000000
        11000000.11100100.01110000.00000000
```

If the results match, the IP addresses are on the same subnet.

Classless Addressing

Classless routing is known as CIDR (Classless Inter-Domain Routing) or supernetting is the addressing scheme currently used in the Internet. Since there are no address classes in CIDR, the size of the network ID cannot be seen from the address alone. In CIDR, the length of the network ID is indicated by placing it following a slash after the address. This is called *CIDR notation* or *slash notation*.

e.g. 192.248.112.3/22 means this network has 22 bits for the network ID and 10 bits for the host ID.

CIDR was based on the already successful practice of subnetting. By *supernetting*, or allowing the subnet boundary to move to the left, into the network portion, groups of neighboring classful networks could be combined into single routing table entries. CIDR also eliminated most of subnetting's restrictions.

11.6 ACCESSING THE INTERNET AND INTERNET SERVICES

Who is the service provider to access the Internet? We gain access to the Internet through an Internet Service Provider (ISP). ISP is a business organization that provides us access to the Internet and related services. Many Internet Service Providers (ISPs) have large numbers of dial-in users. Since IP addresses are a scarce resource and not all the users access the Internet at the same time, IP addresses are dynamically allocated for the clients who are dialing in.

Over the Internet an array of services are provided which have become essentials in our day to day lives. The following section covers three such basic Internet services; e-mail, telnet and FTP. Electronic mail (e-mail) is used to send messages over the internet or any other network. It is the most widely used internet service. It is of very low cost, fast and reliable. Since it is not a real time application, the sender and receiver do not have to be there at the same time. Telnet is used to access and work in a remote computer through Internet. FTP (File Transfer Protocol) is used to upload and down load files over the Internet.

11.7 EMAIL

Messages are sent to an email address and delivered to a mailbox. It can be read, stored or just kept in the mailbox to be read at convenience.

An example for an email address,

mala@ou.ac.lk

Email access is given by different ways. User can login to the server, use web based mail, have LAN/dedicated connection or dialup.

Computers are used to send and receive messages but it can also be done through PDAs, email enables mobile phones etc. One message can be sent to one receiver or to many. Sending multiple copies of the same message is quite easy.

Format of a typical email message,

Envelop contains delivery information

Header contains information about the message

Body is the message itself

Attachment is extra information, a file or picture that is separate from the body of the mail

Header consists of following fields,

From: (mean the sender)

To: (mean the receiver)

Subject: (what is this email is about or topic of the message)

CC: (carbon copies, who else other than given in 'To' filed get a copy)

BCC: (Blind carbon copies, others cannot see that a copy has gone to somebody if that mailbox is given in BCC)

Activity

Send an email to the co-ordinator saying what the most useful feature in the Virtual class is. (for your studies)

You can get the coordinator's email address from the current activity Diary.

11.8 TELNET

Telnet is a part of TCP/IP protocol that lets you log on to another computer and use it. For this you need to have an account in that other computer. Some computers allow new users to log on as guest or new-user. Simply telnet enables networking by telephone since one can log on to a computer even in another country for the price of a local phone call. Incidentally, the name **telnet** comes from **Tel**ephone **Net**work.

What happens here is that using Telnet you type commands into your keyboard that are then sent from your computer to the local Internet service provider. From your provider those commands are sent to the remote computer that you want to access.

One disadvantage of Telnet is that when you Telnet to another computer, you have to use the menus that are set up on that system. They can be unfamiliar, so you'll have to learn by trial and error.

Once you are accepted into their system, follow the menus to look around. Most of these systems do not give access to all the information on these computers. Only a part of their information is made available to Internet users. This is because these computers are part of a larger organization, such as a university. University networks only let teachers and students who have accounts at that university to log on to most of the databases. However they do make their libraries and other services available to Internet users who log on as guests.

To read more on Telnet; http://www.usus.org/elements/telnet.htm

Activity

1. Sign onto the Internet.
2. Launch your telnet program.
3. Enter the telnet address.
4. Write down important information, such as how to log off and what the escape character is (in case you get stuck or your "terminal" freezes).
5. Enter a username and password.
6. Type commands, or choose numbers from text-based menus to complete your task.
7. Exit or log off the host computer.

11.9 FTP

FTP Stands for File Transfer Protocol. FTP allows you to access remote computers and retrieve files from these computers. FTP can be found in the menu of your Internet access software program.

What is available through FTP?

There are hundreds of systems connected to the Internet that have file libraries, or archives, accessible to the public. Most of the material in these file libraries consists of free or low-cost programs for almost every kind of computer. For example, you can find copies of historical documents (e.g. the Magna Carta), song lyrics, poems, or even summaries of TV programs.

You need an account with the remote computer site to log on. However most remote computer sites allow outside user access with the user identification ‘anonymous’. When the name anonymous is used, you inform the FTP site that you are not a regular user of the site, but you would like to access that FTP site and retrieve files.

Activity

1. Launch your FTP software (Free FTP manager, Fetch, FTP CommandPro, and so on).
2. Type the address of the computer you are seeking.
3. Log in to the FTP server (usually as username anonymous).
4. Enter the password when prompted. Most of the time, entering a password is optional for public sites. It is common Net courtesy to enter your entire e-mail address instead of just leaving it blank.
5. Browse the directories for files needed.
6. Get the files needed.
7. Log off (usually by typing quit, exit, or bye).

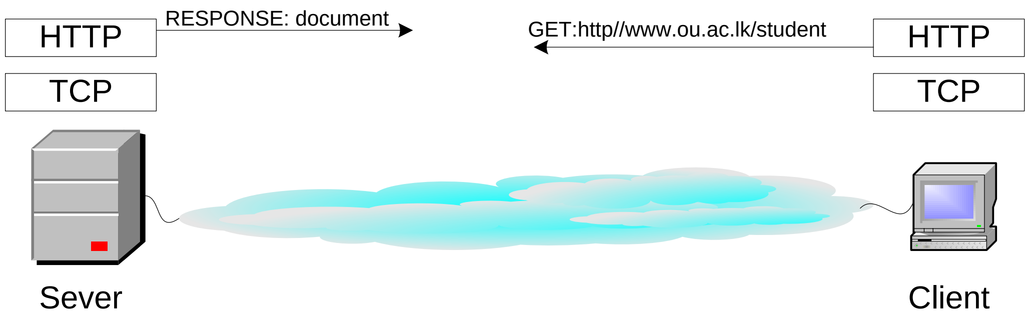
After you log off the server, you can open the files you have downloaded by using a word-processor for a text file, a graphics program for a picture, or a multimedia tool.

11.10 WORLD WIDE WEB (WWW)

The World Wide Web - which is also called the WWW or simply the web, is the most popular service available on Internet. It has a set of interconnected documents which are mainly pictures and hypertext. With hypertext you can jump from one place to another, all over the world, with a single click of the mouse. This section is about how the world wide web functions, how to search something from the web, i.e. use of search engines, authoring a simple web page with HTML, and the use of web authoring software tools.

Other than pictures, web contains text, sound, and even video information where many of the other Internet services are limited to text only. Documents can also be (and often are) linked to other documents by completely different authors. The advantage of using hypertext is that in a hypertext document, if you want more information about the particular subject (and if it has a hyper link), you can just click on it with your mouse to read further details.

To access the Web, you need to run a Web browser program. The Web browser runs in the client machine. The client contacts a web server and get down the information to display to users as shown in the figure below.



When you click on a hypertext link, it can fetch documents from other servers on the Internet as well. There may be an application server that runs the applications necessary to facilitate the user’s request. Web browsers can also download them to your computer's hard disk. Some web servers have search capabilities, so that you can locate documents and databases by searching for specific words and phrases. Some web pages may have text-only versions. Though not so attractive downloading less number of pictures makes the process complete faster.

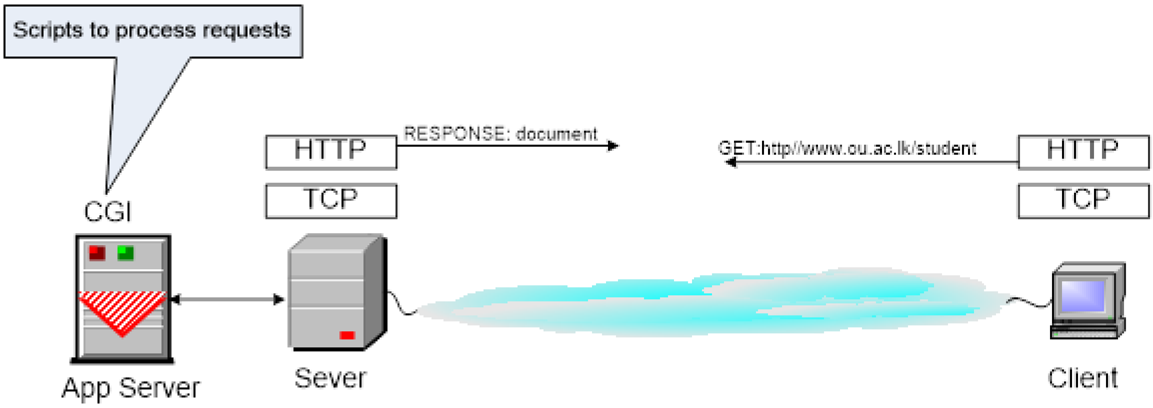


Figure 11.3 Accessing a web from the server

How do you go to a web page?

URL - Universal Resource Locator – URL is a addresses for the location of any Internet resource. URL addresses are case sensitive. URL points to the location where the information you need is held.

e.g. http://www.ou.ac.lk

HTTP - HyperText Transport Protocol is the protocol to transfer information over the World Wide Web. It is a request/response protocol between clients and servers.

11.11 HTML (HYPER TEXT MARKUP LANGUAGE)

HTML is a markup language for creation of web pages. HTML provides syntax to describe the structure of a webpage. It denotes text-based information in a document using tags to show headings, paragraphs, lists, etc. It has syntax to include interactive forms, embedded images, and other objects with HTML. HTML can also describe, the appearance and semantics of a document, and can include embedded scripting language code which can affect the behavior of web browsers.

Elements are the structures that describe parts of an HTML document. For example, the P element represents a *paragraph* while the **H1** element denotes a level 1 heading.

An element has three parts: a start tag, content, and an end tag. A **tag** is special text--"markup"--that is delimited by "<" and ">". An end tag includes a "/" after the "<".

e.g. h1 element has a start tag, **<h1>**, and an end tag, **</h1>**. It is written;

<h> This is My Home Page </h1>

Table below gives a few basic elements. Following page contains a full list of HTML 4 elements-

http://www.htmlhelp.com/reference/html40/olist.html

http://www.w3.org/MarkUp/ This is W3C's home page for the HTML Activity

http://www.w3.org/TR/html401/ HTML 4.01 specification by W3C

Tag	Description
<html> .. </html>	Declares the web page to be written in HTML
<head> .. <head>	Delimits the pages head
<title> .. </title>	Define the title. It is not displayed on the page.
<body>.. </body>	Delimits the page's body
<h*n*> .. </h*n*>	Delimits a level n heading
 .. 	Set .. in boldface
<p>	Starts a paragraph
<hr>	Insert a horizontal rule
<img src = "..”>	Displays an image here
 .. 	Defines a hyperlink

11.12 WEB AUTHORING TOOLS

To create web pages there are authoring software available now. Using these you can create web pages with minimal knowledge of html. A few examples are Dream-weaver, Front Page, Nvu (Open source), and Amaya (Open source). Just like in MicroSoft Word you can type headings, paragraphs, lists etc and HTML tags will be automatically generated. When inserting a hyper-link, or an image it opens a window where you can insert necessary information.

You can create a web site and manage it quite easily using a web Content Management System (CMS). Web content management system is a software system used to manage and control a large, dynamic collection of web material. Two examples for web CMSs are Joomla and Dupal (both are Open source software).

You may find more open sources web CMSs at the page:

http://www.opensourcecms.com/index.php?option=com_frontpage&Itemid=1

11.13 SECURITY CONCERNS

Now that you understand what the internet is and the functions and the concepts regarding the internet and the web services, it is important to have an understanding the security concerns related to privacy, integrity, confidentiality and non-repudiation.

Internet users obtain Internet access through an Internet service provider (ISP). All data transmitted to and from users must pass through the ISP. Thus, an ISP has the potential to observe users' activities on the Internet. Despite these legal and ethical restrictions, some ISPs, such as British Telecom (BT), are planning to use deep packet inspection technology provided by companies such as Phorm in order to examine the contents of the pages that people visit. By doing so, they can build up a profile of a person's web surfing habits, which can then be sold on to advertisers in order to provide targeted advertising. BT's attempt at doing this will be marketed under the name 'Webwise'. From a privacy standpoint, ISPs would ideally collect only as much information as they require in order to provide Internet connectivity (IP address, billing information if applicable, etc). The information collected by an ISP and its utility should be transparent to the users along with an awareness of the potential disclosure risks. ISPs maybe required by the state to provide information for various reasons. In countries such as the United States of America, such a request does not necessarily require a warrant.

To ensure data security, encryption techniques are used. https has become the most popular and best-supported standard for encryption web traffic. However, to further minimize potential risks of disclosure of sensitive information, anonymization techniques can be used. An Anonymizer such as I2P - The Anonymous Network or Tor, can be used for accessing web services without revealing the IP address of the services are that you access. For more information regarding the security concerns, you can refer: http://cs.sru.edu/~mullins/cpsc100book/ module06_internet/module06-05_internet.html

Activity

Home pages are often set up by individual people, organizations or companies who want to enable the interested parties to access the information they put up. For example at the OU website they put up the examination results, important notices, advertisement for student admission, etc.

1. Create your own home page with simple HTML. You may use notepad or any simple editor.

Your home page might include:

your personal details

your hobbies and interests

Links to your school's home pages

Links to favourite reference resources

A picture of yourself

2. Use Macromedia Dreamweaver to create a page (You can do this with Frontpage or any other such tool)

First you'll have to create a 'new' HTML file.

Give it a name and save it.

Then create your home page considering the suggestions given above.

Need help to use Dreamweaver?

Here are some tutorials given in the web;

http://www.adobe.com/support/documentation/en/dreamweaver/ (have to download the .pdf file)

http://www.kellogg.northwestern.edu/kis/docs/howto/software/webauthoring/dreamweaver.htmhttp://www.wellesley.edu/Computing/Dreamweaver/dreamweaverMXbasics.html#text

SUMMARY

It is realized that internet is a network of interconnected heterogeneous networks. It allows you to access to a whole resource of data and information stored at different **hosts** or **servers** all over the world. This session emphasizes the importance of having the communication links and the protocols that are followed. It also gives an overview of the IP address and the use of sub netting. Also gives a description of the how to create a simple web page using HTML.

LEARNING OUTCOMES

Now you will be able to,

- Describe the terms associated with Internet and TCP/IP
- Describe an IP address and the use of subnetting
- Create a simple web page using HTML

Session 12
Introduction to wired and wireless communications

INTRODUCTION

Transfer of information over distances among different users is termed as *telecommunication*. Almost all today's information-based systems such as the Internet are completely based on telecommunication infrastructures. Telecommunication systems are multifold and have evolved dramatically since its inception. In this Session we will be discussing the basics of a telecommunication system. We also discuss the impairments or the limitations in the channel. Furthermore, the physical infrastructure used for implementing the communication channels are also presented. The session concludes with a discussion on the regulatory requirements on sharing the spectrum among different services and users.

12.1 TELECOMMUNICATION BASICS

Communication is defined as the transmission of information from one place to another through use of speech, written messages or coded signals etc. In the past, people used some primitive methods of communication such as smoke signals, sounding bells, sending messages through birds etc. With the advancement of civilization these methods proved to be unsatisfactory and time consuming, as it did not match with the needs of the society. Hence these techniques were improved to suite the modern society. Now, we use a variety of methods for communication. That includes two-way communication methods such as telephony and one-way communication methods such as broadcast systems.

ELEMENTS OF A COMMUNICATION SYSTEM

Any communication system in general consists of a transmitter, communication medium and the receiver as shown in Figure 12.1.

Figure 12.1 Basic communication system

For an example consider that two people standing and talking on the street. Here the type of the information signal is voice and it is a communication system where you are the transmitter and your friend is the receiver. The speech signal travels through air as the communication medium. What will happen when you all keep moving away from each other? After a certain time, you are unable to hear each other. In technical terms, the amplitude of their speech signals dies out as they get mixed with noise from the environment.

Now consider that we use a wire for the communication of the above example. Then, the signal travels through the wire is electric in nature and the two people might be able to communicate for a longer distance with better clarity. Even now the signal (now an electric signal) eventually dies out with the distance.

According to this example we can conclude two things:

1. Since the signal that travels through the wire is in electric form we need some device to convert speech signal into an electrical signal at the sender's end and another device to convert the electrical signal back to a speech signal at the receiver's end. We call this type of devices transducers. Microphone, speaker, etc. are some examples for such transducers.
2. The signal dies out after traversing a certain distance. In the past, since the people get tired to travel in long distances, they used some kind of carriers like birds, horses etc. to send messages. Similarly we send our signals "on" a high frequency carrier wave. (not exactly "superimposed on" but "modulated on"). We call this process as modulation. We then need to demodulate the signal (get the guy down from the horse to deliver the message) at the receiving end. You will learn this process in detail in session 13.

Considering above, now we can form the block diagram of an electrical communication system as given in the Figure 12.2.

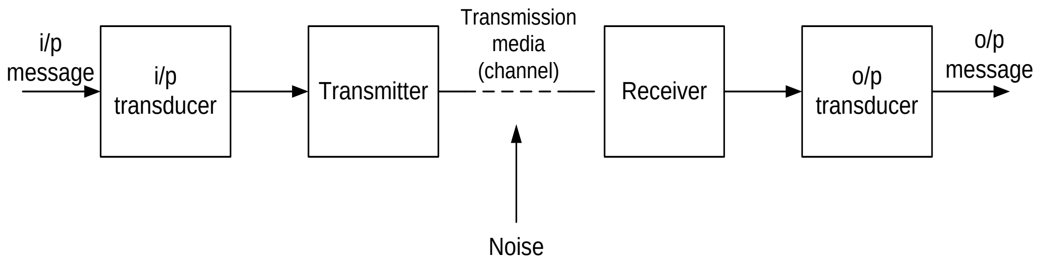


Figure 12.2 Basic block diagram of an electrical communication system

As explained earlier, input transducer converts the original information signal into an electrical signal (or an optical signal).

An electrical signal may be any one or more of the following types which are time variant signals. (i.e. varies with time).

- current in an electric circuit.
- voltage across a resistive component of a circuit.
- magnetic or electric field at a defined point.

Transmitter is a device which is capable of generating electrical signals in the form of varying current, voltage or electromagnetic field. Such devices are identified as voltage sources, current source, depending on the type of signal generated. Similarly, optical signals are having a varying intensity of light.

The modulation process takes place in the transmitter. In addition, processes like amplification, filtering and coupling to the channel are also done by the transmitter.

Receiver is a device which is capable of receiving these signals and reproducing them in the original form. Here again we need to use an output transducer to reconvert the received electrical signal (or optical signal) into the original form.

Between the transmitter and the receiver lies the communication medium, we call it the channel, which is most susceptible to any noise signals or interferences. You will find that all communication system block diagrams are a mirror image around the channel. At the same time the channel may be *wired* or *wireless*.

TYPES OF SIGNALS

We can define a signal as a physical quantity that varies with respect to time, space and carries information from source to destination. Examples in the practical world include audio, video, speech, image, communication, geophysical, sonar, radar, medical and musical signals. Signals can be categorized in various ways. But in general we can divide them into two main types as analog and digital.

Analog signals can be described as a signal that has a continuous nature. When we consider electrical signals, continuously varying voltages, frequencies, or phases, may be used as analog signals (see Figure 12.3).

A digital signal is a physical signal that is a representation of a sequence of discrete values e.g. a bit stream. This signal can have only one of two values at any given time, 1 or 0. A digital system would be more like flipping a light switch on and off.

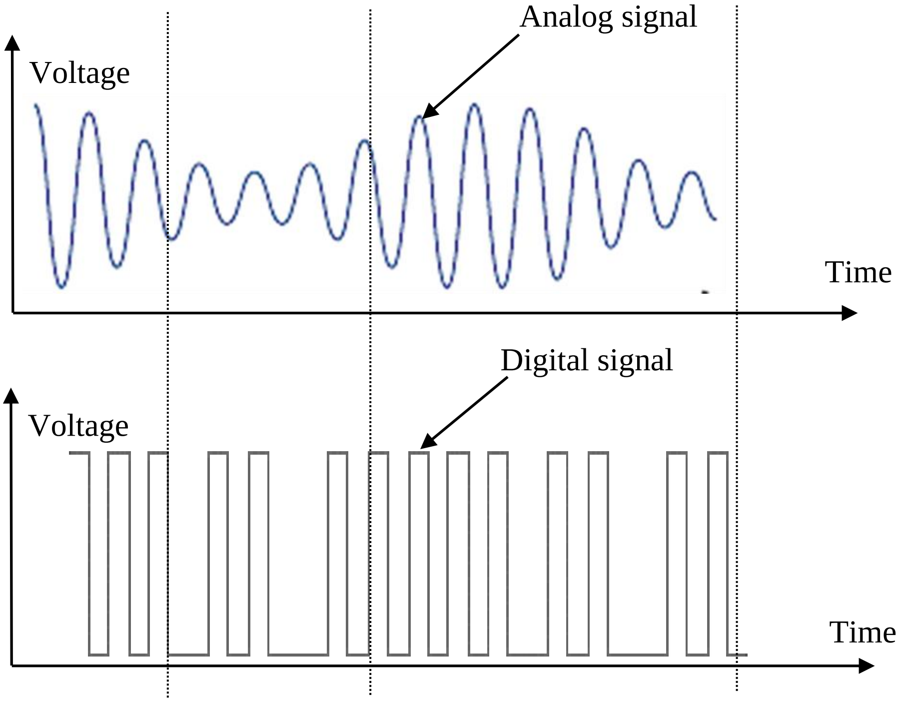


Figure 12.3 Analog and a digital signal respectively

12.2 COMMUNICATION CHANNEL

As already mentioned before, the communication channel is one of the most important elements in a communication system. Channels do have many limitations. Furthermore, signals carried over the channel may undergo different adverse effects. Inherent limitations together with the channel impairments impose an upper limit for the amount of information that can be carried in a particular channel.

12.2.1 NOISE, ATTENUATION AND INTERFERENCE

Noise refers to unwanted signals that disturb communications. Any receiver has to compete with noise as noise deteriorates the quality of the received signal, e.g. the appearance of “snow” on the TV screen, or “static” sounds during an audio transmission.

We can categorize noise as external noise and internal noise. External noise refers to the interference from nearby channels, human-made noise, natural noise etc. Internal noise includes thermal noise, random emission in electronic devices.

In digital communication systems, noise degrades the throughput (throughput refers to how much data can be transferred from one location to another in a given amount of time) because it requires retransmission of data packets or extra coding to recover the data in the presence of errors.

Activity

List different sources of noise and interferences.

12.2.2 BANDWIDTH

Bandwidth is an important parameter related to communication signals. We can define bandwidth in two ways, signal bandwidth and channel bandwidth.

Signal bandwidth is the range of frequencies that makes up a signal. Say, the frequency range for commercial speech is 300 Hz – 3400 Hz. Therefore, the bandwidth of commercial speech signal is 3100 Hz.

Typically, commonly encountered frequency ranges are,

- Human voice 100 Hz – 10 000 Hz
- Human hearing 20 Hz – 15 000 Hz
- Commercial speech 300 Hz – 3400 Hz
- Mains electricity 50 Hz – 60 Hz

When you transmit these signals in a particular communication channel, we can define the channel bandwidth as its information-carrying capacity. The channel may be analog or digital. Analog transmissions such as telephone calls, AM and FM radio, television etc. are measured in cycles per second (hertz or Hz). Digital transmissions are measured in bits per second (bps).

Modern communication systems use a variety of techniques such as coding to increase the channel capacity hence facilitates high speed communications over band limited channels.

12.3 COMMUNICATION MEDIA

In this section we study how the signals are physically transmitted from a transmitter to a receiver. More precisely, we are going to study the physical transmission medium involved in signal transmission.

We already know that to make a call or to send data over a network, a path has to be there between the transmitter and the receiver. In other words there should be a particular transmission medium for all these signals to travel from one place to another, ultimately creating a network. Recall that the transmission medium or communication channels could be either wired or wireless.

- Wired Medium

  Wired transmission medium is also known as guided transmission medium. Here an actual physical tangible pathway is available and the information travels through the channel in the form of an electrical signal or an optical signal.

  Ex: Twisted pair, Coaxial Cable, Fiber optics.

- Wireless medium

  Wireless medium or unguided transmission medium does not have an actual tangible pathway. Usually the transmission and receiving of data is achieved by means of an antenna.

12.3.1 WIRED COMMUNICATION MEDIA

A guided medium provides a conduit between the transmitter and the receiver. The signal in bound by the medium and is guided through the path. In telecommunication we use coaxial cables, twisted pairs and fiber optics as guided transmission medium. Now let us look into each individually.

Coaxial Cable

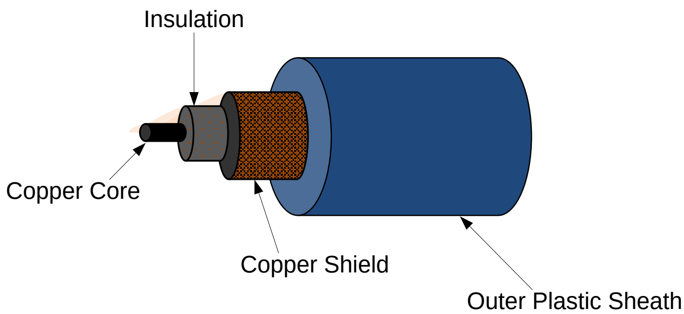


Figure 12.4 Coaxial Cable structure

Coaxial cable consists of two conductors constructed concentrically along the axis. The inner conductor is held in place by spaced insulating rings or a shielded dielectric material. The outer metallic wrapping acts as a shield as well as a second conductor. The diameter of the coaxial cable varies from 1-2.5cm. Primarily these cables were employed for Ethernet computer networks and noisy industrial environments. The shielded concentric structure of the coaxial cable reduces external electromagnetic interference and cross talk. Coaxial cables support both analog and digital signals and allow higher frequency ranges than twisted pair. Nevertheless, coaxial cables are replaced by twisted pair for computer networks and fiber optics for high speed networks.

Twisted Pair

Twisted pair cables are frequently utilized to connect the subscriber to the telephone network. Between the two wires twisted together, one is considered to be positive and the other negative. The noise that appears in one appears on the other as well, but the 180^{o} phase difference nulls the noise at the receiving end due to the opposite polarities. The reason to use twisted pairs instead of parallel is because in parallel electromagnetic interference from devices such as motors create noise. Also, when the two wires are in parallel the one closest to the source of noise gets a higher interference and ends up with a high voltage. As a result, a load imbalance is created between the two wires. Thus, we understand that if the two wires are of same distance to the noise source, it creates a load balance and interference can be avoided. This is what happens in the twisted pair, each wire is closer to the noise source for half the time and further away for the other half. In a cable a number of twisted pairs are bundled together. Twisted pairs are of two types as shielded (STP) and unshielded (UTP). The shielded pairs are covered with a foil or a wire braid shield, further reducing the noise. Twisted pairs accommodate both analog and digital signals and is the cheapest and the most widely used. Out of the two types UTP is the cheapest but supports a relatively low bandwidth.

Fiber Optics

Fiber optics are made out of glass material and use a light beam for transmission of signals. The speed of light through the fiber depends upon the density of the medium. i.e. when the density is high, the speed is low and vice versa. To transmit a light beam through the fiber, the concept of refraction is employed. Refraction is when there is a change of density in the medium the light beam changes its direction such that when moving from a higher density to lower, the signal bends away from the normal.

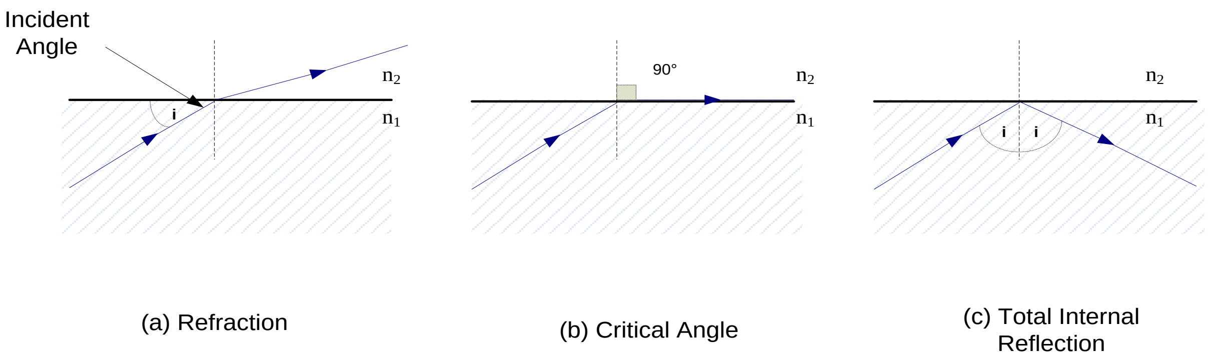


Figure 12.5 Refraction of light

Let us now consider a situation where a light beam travels from a high-density medium to a low density medium. We already know that the angle at which the light falls upon the boundary is known as the *incident angle*. However, when we keep on increasing this angle, at one point the signal exiting glides over the boundary (90º angle). This angle of incidence is known as the *critical angle*. We can further clarify this with the Figure 12.4. When the incident angle exceeds the critical angle, the light wave reflects back into the same medium creating *total internal reflection*. This is the concept used in fiber optics to propagate the light wave along the fiber. As we can see in Figure 12.5 the fiber is made up of two layers with distinct densities namely core and the cladding. The cladding is of lesser density compared to the core. Thus, when a light beam traveling from the core meets the core-cladding boundary, it reflects off the cladding back in to the core creating a total internal reflection.

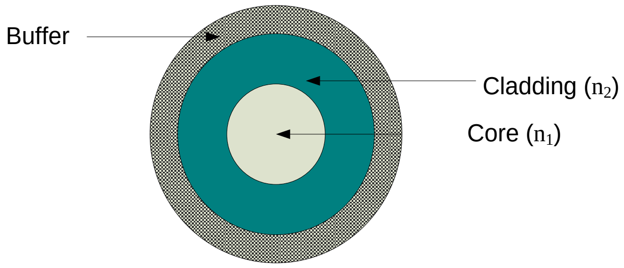


Figure 12.6 Cross section of a fiber

The information that has to be transmitted through a fiber is encoded in ones and zeros and are denoted by serially switching ON and OFF the light source. Fiber optics has the ability to isolate electrical noise and cross talk and no energy radiation takes place, which means there is significantly less attenuation. The higher bandwidth, speed and ability to be employed in long distance transmission lines make it ideal for modern high-speed networks.

12.3.2 WIRELESS COMMUNICATION MEDIA

Recently wireless communication media is preferred over wired communications due to several reasons. Due to the wire-less nature these communication links are easy deployable, costs less and most importantly can provide much appreciated mobility. However, wireless communication have several drawbacks too. They are more vulnerable to adverse effects such as noise, fading, shadowing and also interference. The bandwidth is limited and its security is very low. However, various techniques are being employed to overcome these issues and to harness the aforementioned advantages, hence wireless media has become the future trend in communication. Two basic types of wireless media is popular, namely infrared/visible light and radio waves.

Infrared/ visible light

Use of infrared or visible light based communication has been a popular choice for many short range communications. Many devices such as remote controllers communicate with other devices via infrared media. Meanwhile light fidelity (LiFi) has been an upcoming technology to carryout short range communication over visible light. Both infrared and visible light are electromagnetic waves with poor reflection capabilities, thus requires line of sight (LOS) for their communication. This requirement limits its usable range. On the other hand the security in communications over infrared/visible light systems is low. However, infrared or visible light based systems have several advantages. As their applications are on short range, their use is not regulated or licensed (you will learn this later in this Session). On the other hand their harmful effects to human body is comparatively less compared to other radio and micro waves. Furthermore, the bandwidth they can provide is immense when compared to radio wave communication. An infrared or visible light communication requires a transmitter (usually a laser diode or a light emitting diode) together with a visible light/ infrared sensor (usually photo diode or a photo transistor). They can provide Gigabits of speeds over short-range communications.

Radio waves

First brought in to prominence by Gugliamo Marconi in 1890's radio waves are the most popular type of wireless media as they have very good reflection and scattering properties. As a result, they do not need a LOS path between the transmitter and the receiver. Usually a radio communication link consists of a transmitter and a receiver equipped with an antenna at each. These antennas convert the electrical signals to radio waves and vise-versa.

12.3.3 RADIO WAVE PROPAGATION MODES

Radio waves may travel from a transmitter to a receiver using different propagation modes. Especially due to the air density variation in the atmosphere radio waves can undergo refraction as shown in figure 12.7 which results in longer communication distances using the refracted wave. This subsection presents the different communication modes used in radio wave-based communications.

Ground Wave (Surface Mode)

This mode of propagation interacts with the semiconductor surface of earth and follows the contour of earth. The vertically polarized ground waves bend over hills and mountains. However, as the ground is not a perfect conductor the signals are attenuated depending upon the nature of the terrain they propagate through. Thus, we understand that features like the roughness of the terrain and moisture content of soil affects the quality of the signal. Ground waves generally use Low Frequency (LF) and Very Low Frequencies (VLF). The most common application of ground waves is for military communication where Extremely Low Frequencies (ELF) are employed for communicating with submarines deep under the water. The salty sea water is ideal for ground wave propagation as it provides a good conductive medium.

Line of Sight

Line of sight waves propagate from transmitter to the receiver in a straight line (wave c-c in figure 12.7). It is limited to the distance of visual horizon. Line of sight waves facilitate medium range transmission applications such as cell phones, walkie-talkies, FM Radio, Cordless phones, Radar, Satellite TV, Wireless networks etc. It is the only mode of propagation that is possible at microwave frequencies.

Sky Wave (Ionospheric Mode)

The ionosphere is the atmosphere that extends from 60km to 400km range. It contains layers of charged particles that has the ability to refract the radio waves propagating through back to earth. The ionosphere has three regions namely D region, E region and F region. Level of ionization varies from one region to another and as a result when signals propagate through region D they get attenuated before passing through to region E. Attenuation takes place partially in region E also and some of the sky waves pass through to region F. Some of the signals even manage to reflect back to earth (wave b-b in figure 12.7). Nevertheless, in region F the signals are reflected back to earth. During daytime, this region splits into two regions F1 and F2. For sky waves, high frequency signals are the most suitable as low frequency signals tend to get attenuated more. This mode of propagation facilitates long distance transmission by use of multiple reflections such that a signal can be transmitted from one side of the globe to the other. As a result, Sky waves enabled international broadcasting while managing to provide global links to inaccessible regions.

Tropospheric Mode

We know that the Troposphere is the lowest range of atmosphere extending from the surface of earth to about 6-12km. In tropospheric mode, VHF and UHF signals travel beyond visual horizon and refracts back in to the tropospheric region itself (wave a-a in figure 12.7). However, propagation of radio waves in this mode are affected by changes in the refractive index of air as well as temperature and pressure.
There are other niche modes of radio wave propagation such as Sporadic E propagation, Aurora propagation, Transequatorial propagation etc. Nevertheless, we can conclude that radio wave propagation is affected by many atmospheric conditions like rain, clouds, fog, lightening, seasons etc.

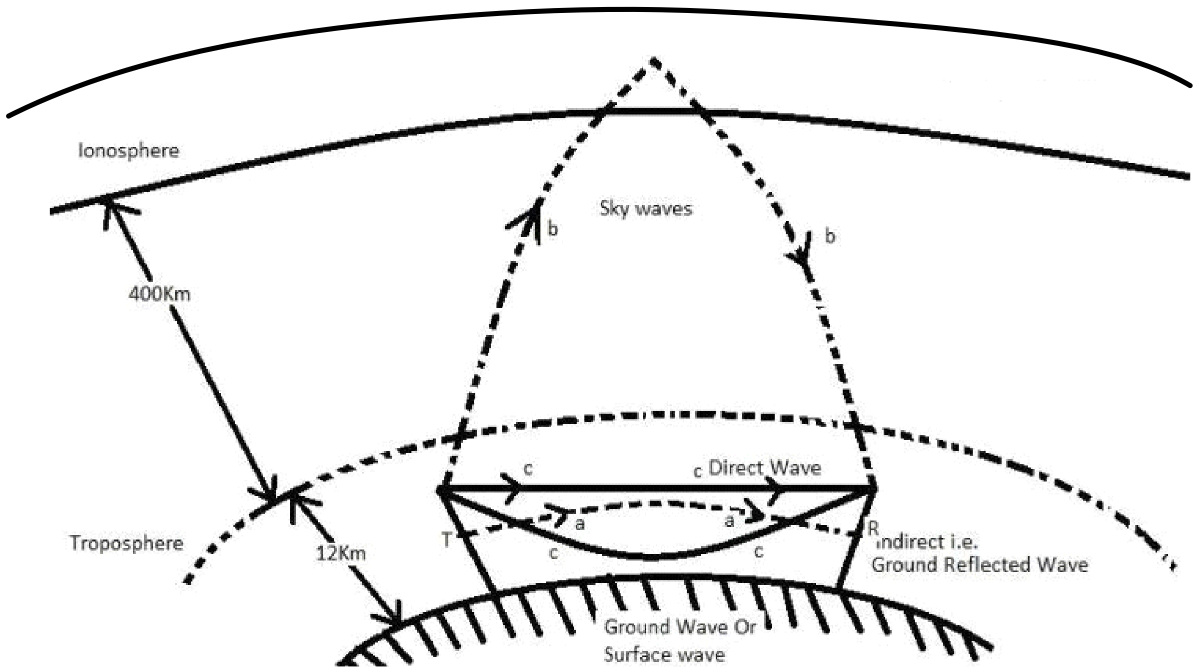


Figure 12.7 Modes of propagation of radio waves

12.3.4 DIFFERENT TYPES OF RADIO COMMUNICATION SYSTEMS

Radio communication systems can be basically categorized into following types:

- Radio broadcast systems
- Cellular radio
- Microwave systems

Radio broadcasting systems

In radio broadcasting service we can broadcast audio signals through the air as radio waves. As you can see in the figure 12.8 we use a transmitter to transmit radio waves to receiving antennas. Radio waves have different frequency segments, and you will be able to pick up an audio signal by tuning the receiver into a specific frequency segment. Radio broadcast systems can transmit signals over long distances at a data rate up to 2 Megabits per second (AM/FM radio).

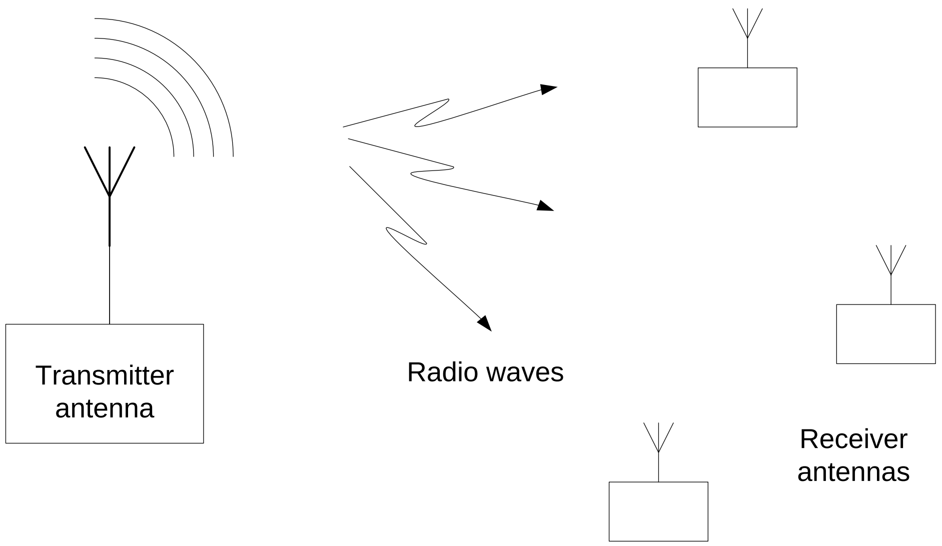


Figure 12.8 Broadcast radio system

Cellular radio systems

This is a form of broadcast radio used for mobile communication. High frequency radio waves are used to transmit voice or data. An area is divided into smaller areas called cells and a specific frequency is assigned for each cell. A frequency corresponds to a specific cell can be reused after a certain distance. Each cell consists of an antenna which is called as a Base Station (BS) and all base stations communicate with the Mobile Switching Centre (MSC). MSC has a connection with the Public Switching Telephone Network (PSTN) to connect calls outside the mobile network. You may clearly understand this cellular structure using the figure 12.9. Based on the multiple access technique used for separating multiple simultaneous calls from each other, the cellular systems have evolved over 5 generation from 1G to 5G.

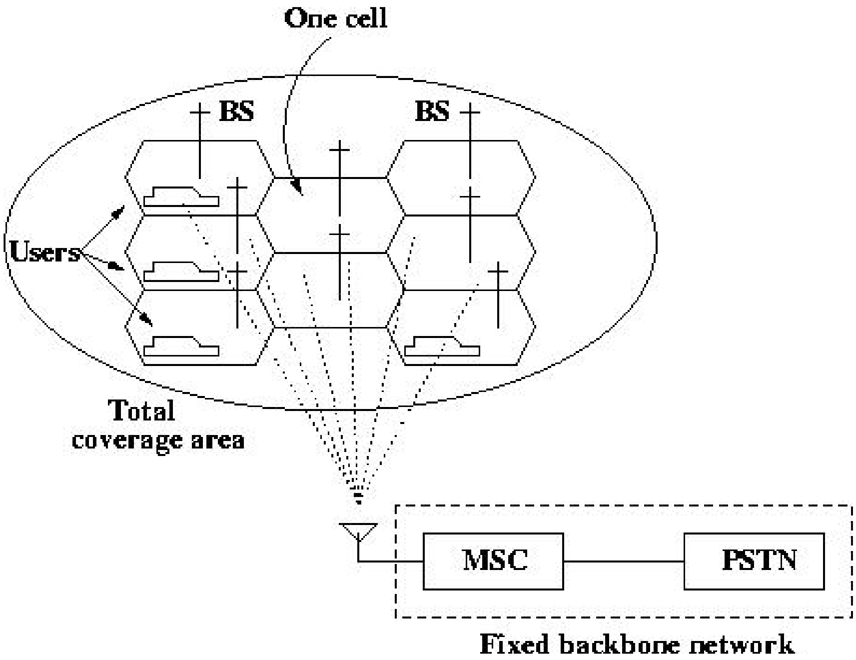


Figure 12.9 Basic cellular structure

Microwave communication systems

Microwave transmission involves the transfer of voice and data through the atmosphere as super high-frequency radio waves called microwaves. They are point-to-point communication systems. Microwave transmission is mainly used to transmit signals between ground-based stations and satellite communication systems.

Microwave transmission mainly uses radio waves whose wavelengths are conveniently measured in small units such as centimeters. Microwaves belong to the radio spectrum ranges of roughly 1.0 gigahertz (GHz) to 30 GHz.

Antennas used in microwave transmissions are of convenient sizes and shapes. Microwave transmission depends on line-of-sight in order to work properly. For two-way communications to take place, two frequencies are used. However, this does not require two antennas because the frequencies can be dealt with one antenna at each end.

The distance covered by microwave signals depends on the height of the antenna. Each antenna is built with a fitted repeater to regenerate the signal before passing it on to the next antenna in line. The main drawback of microwave signals is that they can be affected by bad weather, especially rain.

12.4 RADIO SPECTRUM USAGE AND REGULATIONS

As already mentioned in the previous sub sections, the use of radio waves for communication is very popular. Different bands of frequencies are employed in different communication systems. Table 12.1 below lists some of these application areas.

Table 12.1 Radio communication applications

Band Name	Frequency	Wavelength	Applications
Extremely Low Frequency	3Hz-30Hz	100,000–10,000 km	Submarine Communication
Super Low Frequency	30Hz-300Hz	10,000–1000 km	Submarine Communication
Ultra Low frequency	300Hz-3KHz	1000–100 km	Submarine Communication, Communication in mines
Very Low frequency	3KHz-30KHz	100–10 km	Navigation, time signals, submarine communication, wireless heart rate monitors, geophysics
Low Frequency	30KHz-300KHZ	10–1 km	Navigation, time signals, AM long wave broadcasting (Europe and parts of Asia), RFID, amateur radio
Medium Frequency	300KHz-3MHz	1000–100 m	AM (medium-wave) broadcasts, amateur radio, avalanche beacons
High Frequency	3MHz-30MHz	100–10 m	Shortwave broadcasts, citizens band radio, amateur radio and over-the-horizon aviation communications, RFID, over-the-horizon radar, automatic link establishment (ALE) / near-vertical incidence sky wave (NVIS) radio communications, marine and mobile radio telephony
Very High Frequency	30MHz-300MHz	10–1m	FM, television broadcasts, line-of-sight ground-to-aircraft and aircraft-to-aircraft communications, land mobile and maritime mobile communications, amateur radio, weather radio

Ultra High Frequency	300MHz-3GHz	1000–100 mm	Television broadcasts, microwave devices/communications, radio astronomy, mobile phones, wireless LAN, Bluetooth, ZigBee, GPS and two-way radios such as land mobile, FRS and GMRS radios, amateur radio, satellite radio, Remote control Systems, ADSB
Super High Frequency	3GHz-30GHz	100–10 mm	Radio astronomy, microwave devices/communications, wireless LAN, DSRC, most modern radars, communications satellites, cable and satellite television broadcasting, DBS, amateur radio, satellite radio
Extremely High Frequency	30GHz-300GHz	10–1mm	Radio astronomy, high-frequency microwave radio relay, microwave remote sensing, amateur radio, directed-energy weapon, millimeter wave scanner, wireless LAN (802.11ad)
Tremendously High Frequency	300GHz-1THz	1-0.1mm	Experimental medical imaging to replace X-rays, ultrafast molecular dynamics, condensed-matter physics, terahertz time-domain spectroscopy, terahertz computing/communications, remote sensing, amateur radio

Meanwhile, radio spectrum has to be shared between many users. Hence the regulation is a requirement in radio spectrum usage. Different countries have different telecommunication bodies to perm this regulation. In Sri Lanka the respective authority is Telecommunication Regularity Commission of Sri Lanka (TRCSL) while International Telecommunication Union (ITU) is managing most of the global regularity aspects.

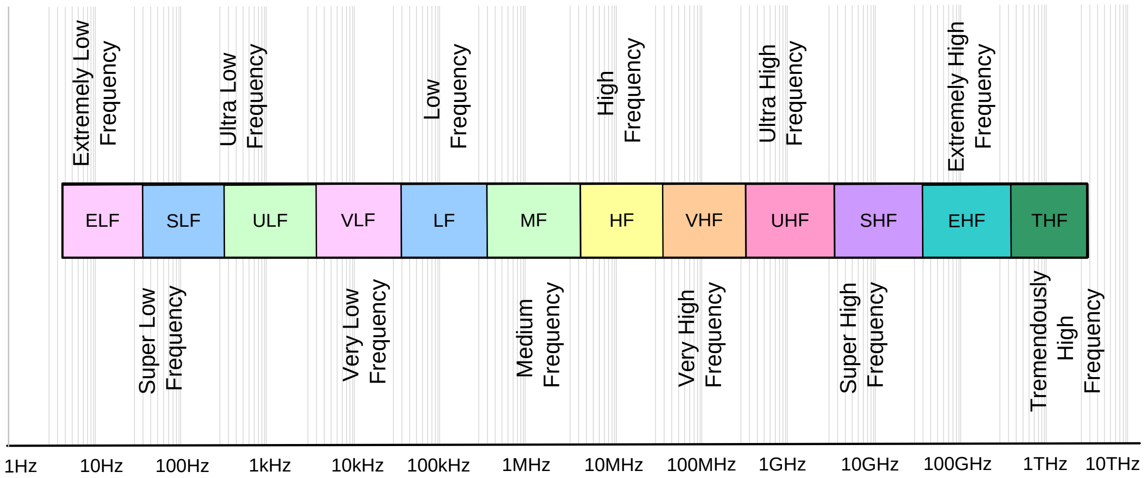


Figure 12.9 Radio Spectrum as defined by ITU

International Telecommunication Union (ITU) divides the radio spectrum into 12 bands from 3Hz to 3000GHz. This categorization is done to prevent any interference caused between users. Different frequency bands are utilized for different technologies and applications where as some bands or parts of bands are sold or licensed to operators of private radio transmission services. The main aim of having a radio regulation is to maintain an appropriately structured system for radio communication to avoid any conflicts. For example, allocation of frequency spectrum has to be done depending upon the region of utilization such that it does not create any interference with another.
With technological advancements, demand for frequency spectrum has increased creating congestion. Thus, there is a need for better methods to utilize the radio spectrum efficiently. Hence, ITU has continuously working on updating regularity policies.

SUMMARY

During the last few decades telecommunication systems have dramatically evolved and have become an essential commodity in today's world.

In this Session we discussed the basic elements of a telecommunication system and then paid our attention to the communication channel. Communication channel is suffering from many impairments and is band limited. Meanwhile, many different types of physical media types are used both wired and wireless to cater for different communication systems. Specially, when wireless communication systems are employed, regularity aspects of spectrum usage is significant and standardization bodies, such as the International Telecommunication Union (ITU), provide the regularity functions.

Based on this introduction, during the next two sessions we will be discussing two of very important aspects in telecommunication, namely modulation and power budget calculation.

LEARNING OUTCOMES

Now you will be able to,

- Describe the basic elements of a telecommunication system.
- Describe the impairments in a communication channel.
- Compare wired and wireless communication media types.
- Explain the need and the job function of the regularity bodies in telecommunication.

Session 13
Modulation Techniques in communication systems

Contents

INTRODUCTION

Now you have learned the basic blocks, functions and components relates with a simple communication system.

In this session we will discuss what is meant by modulation and demodulation. Different types of modulation methods, advantages and disadvantages of each modulation methods.

13.1 MODULATION AND DEMODULATION

The distance that a voice signal travels is very low. Hence, how do we send this signal over a longer distance? This session explains the principles associate with transmission of a low frequency signal in a communication system. The process is named as modulation.

Why modulation is important? As explained earlier the primary requirement of modulation is to make the information signal into a form which is suitable for transmission. Apart from this primary requirement, there are few other reasons for the importance of modulation as given below:

1. Ease of radiation and overcome hardware limitations

We translate signals to high frequencies using modulation. As a result, it becomes relatively easier to design amplifier circuits as well as antenna systems at these higher frequencies.

2. Increases the range of communication

   Low frequency baseband signals suffer from attenuation and hence cannot be transmitted over long distances. So translation to a higher frequency band results in long distance transmission

3. Facilitates multiple access

   By translating the baseband spectrum of signals from various users to different frequency bands, multiple users can be accommodated within a band of the electromagnetic spectrum.

What is modulation?

The transmission of information is similar to transportation of 100 'rice sacks'. The sacks are first loaded to a lorry. Then the lorry 'carries' the sacks to the required destination. In other words the rice sacks are carried by the lorry. In the transmission of information the information is first 'loaded' or 'embedded' to a high frequency, high energy signal known as the carrier signal. This process is known as modulation and the process is shown in Figure 13.1. The information signal is referred to as modulating signal. The carrier signal carries the information to the required destination.

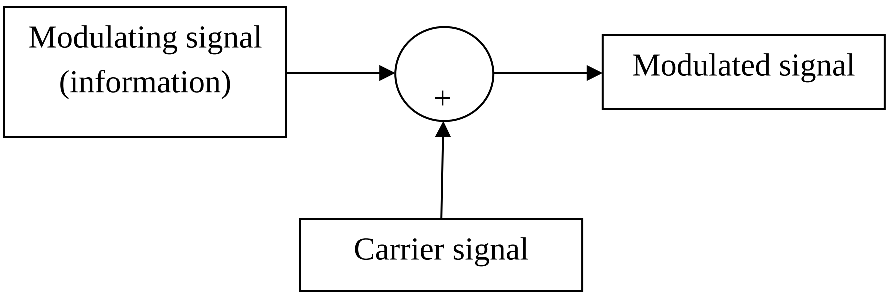


Figure 13.1 Modulation and the process

At the destination the information signal is separated from the carrier signal as shown in Figure 13.2 - this process is known as demodulation.

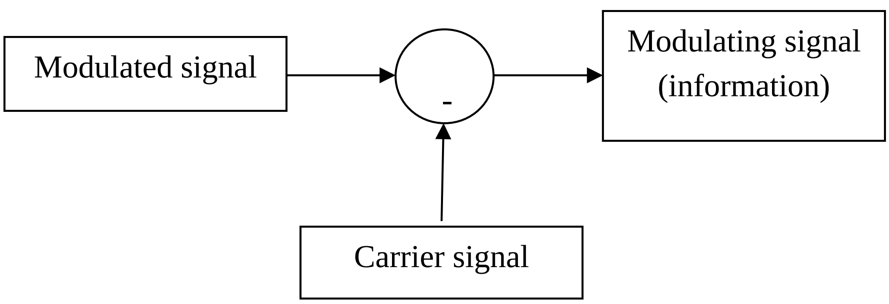


Figure 13.2 Demodulation process

There are two main types of modulation named as analog modulation and digital modulation. High frequency carrier is embedded to the information signal. Where low frequency baseband signal used as information signal in analog modulation and in digital modulation a bit stream is used as the modulating signal.

13.2 ANALOG MODULATION

As previously mentioned, analog communication systems use sinusoidal signals as the carrier. A sinusoidal signal can be described using amplitude and angle (frequency and phase). Therefore, in modulation, these parameters of the high frequency carrier are varied for transmitting information. Accordingly, analog modulation may be divided into Amplitude Modulation and Angle Modulation. Angle modulation can be again divided into frequency and phase modulation.

13.2.1 AMPLITUDE MODULATION

In amplitude modulation the amplitude of the carrier signal is varied according to the information signal (modulating signal). You can see this clearly in the figure 13.3.

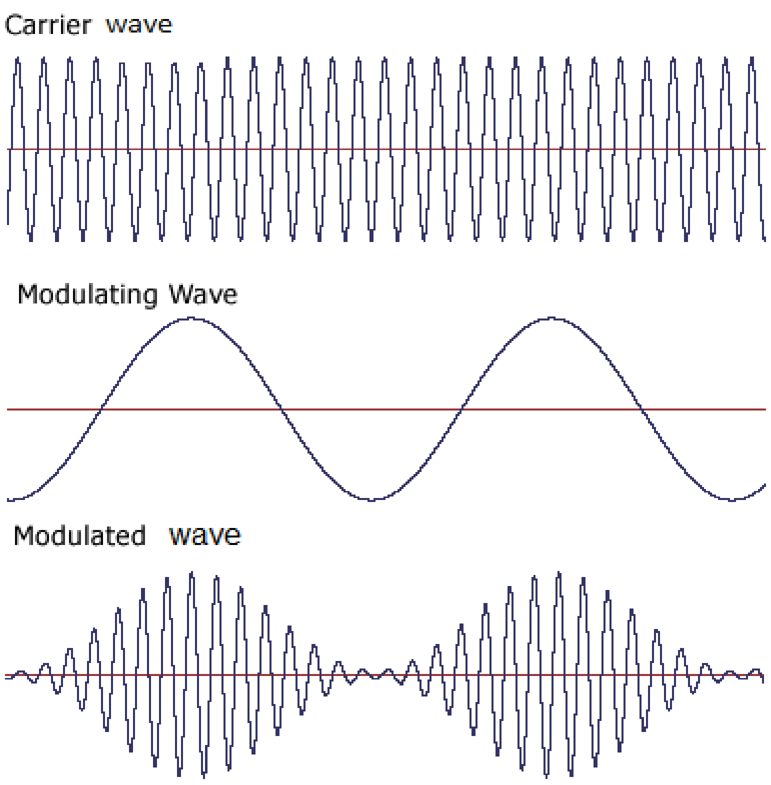


Figure 13.3 Amplitude modulation

Now let's see the theory behind this technique.
Consider the carrier signal, $e_C(t) = E_C \, Cos\, \omega_C t$ and
the modulating signal (information signal), $e_m(t) = E_m \, Cos\, \omega_m t$

Here, $\omega_c = 2\pi f_c$ *and* $\omega_m = 2\pi f_m$ where f_m and f_c are modulating frequency and the carrier frequency respectively.

Then, the amplitude modulated signal, $e(t)$ can be expressed as,

$$\begin{aligned} e(t) &= \left(E_c + e_m(t)\right) Cos\, \omega_c t \\ &= \left(E_c + E_m \, Cos\, \omega_m t\right) Cos\, \omega_c t \end{aligned}$$
------------------------ equation (13.1)

Then, the amplitude of the AM signal is,

$$\begin{aligned} E &= E_c + E_m \, Cos\, \omega_m t \\ &= E_c + m_a E_c \, Cos\, \omega_m t \\ &= E_c (1 + m_a \, Cos\, \omega_m t) \end{aligned}$$
---------------------------- equation (13.2)

Where $m_a = \dfrac{E_m}{E_c}$ = depth of modulation

Here, E_m and E_c are the modulating signal amplitude and the carrier amplitude respectively.

This is also known as modulation index, percent modulation or modulation factor. Multiplying the modulation index by 100 gives the *percentage of modulation.*

From the equation (13.2), you may understand that the carrier amplitude is varied by $\pm m_a$ and this variation of the amplitude takes place at a frequency f_m.

Now the AM signal can be written as

$$\begin{aligned} e(t) &= E_c \left(1 + m_a \, Cos\, \omega_m t\right) Cos\, \omega_c t \\ &= E_c \, Cos\, \omega_c t + m_a E_c \, Cos\, \omega_m t \, Cos\, \omega_c t \end{aligned}$$

$$e(t) = E_c Cos\, \omega_c t + \frac{m_a E_c}{2} Cos\, (\omega_c + \omega_m) t + \frac{m_a E_C}{2} Cos\, (\omega_c - \omega_m) t$$
------ equation (13.3)

How would an amplitude modulated waveform look like if
(i) $m_a < 1$ (ii) $m_a = 1$ and   (iii) $m_a > 1$?

The modulation index should be a number between 0 and 1. Depending on the value of the modulation index following behaviours can occur. Refer figure 13.2 to understand better.

Under modulation ($m_a < 1$)

When $m_a<1$, we call as under modulation. Here, amplitude of the modulating signal is less than the carrier amplitude. Message signal can be comfortably retrieved from the envelope waveform.

Ideal Modulation (when $m_a = 1$)

When $m_a = 1$, the modulation is called as ideal modulation. This is the best modulation and the original transmitted message can be easily retrieved at the receiver.

Since $m_a = 1$

$E_m = E_c$

Over modulation (when $m_a > 1$)

$E_m > E_c$

If the amplitude of the modulating voltage is higher than the carrier voltage, m_a will be greater than 1, causing *distortion* of the modulated waveform.
You can see the effect of this on the AM wave in the figure 13.4.

Broadcast stations in particular take measures to ensure that the carries of their transmissions never become over modulated. Distortion caused by over modulation also produces adjacent channel interference. The transmitters incorporate limiters to prevent more than 100% modulation. However, they also normally incorporate automatic audio gain controls to keep the audio levels such that near 100% modulation levels are achieved for most of the time.

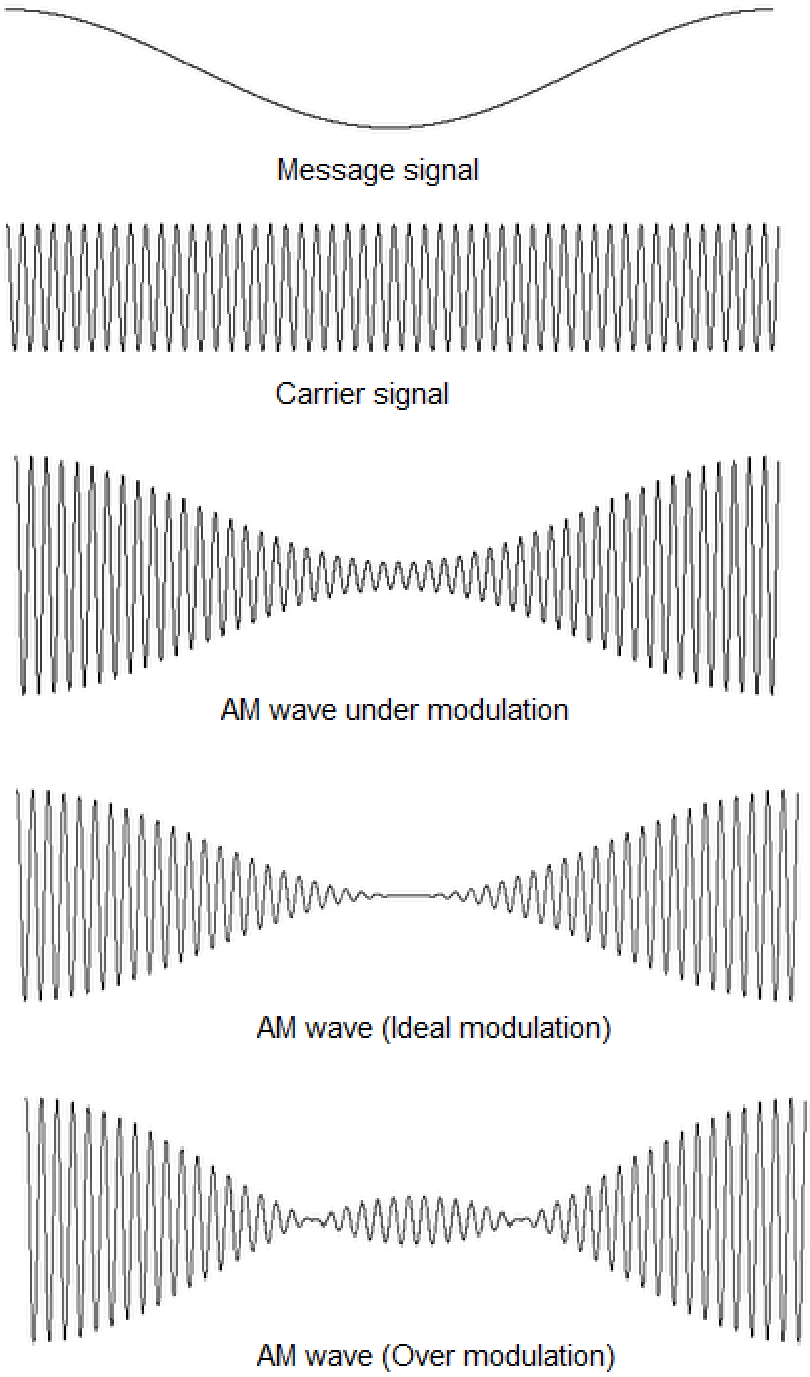


Figure 13.4 Effect of the variation of modulation index on the AM wave

S.A.Q. 13.1

Suppose that on an AM signal, the $V_{max(p-p)}$ value read on the oscilloscope screen is 5.9 divisions and $V_{min(p-p)}$ is 1.2 divisions. What is the modulation index ?

Answer

Let's draw the amplitude modulated waveform first.

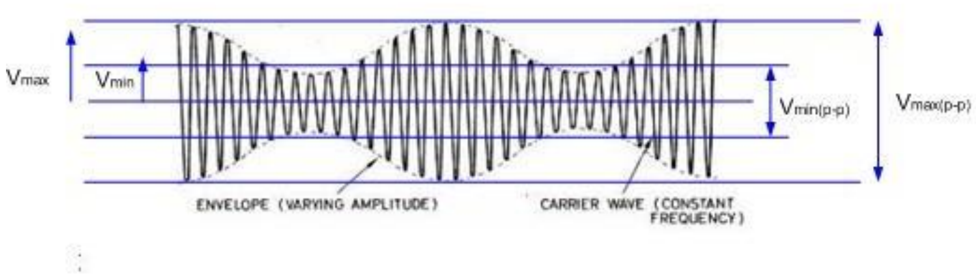


Figure 13.5 Amplitude modulated waveform

We can derive an important equation from this activity.

We know modulation index, $m_a = \dfrac{E_m}{E_c}$

When the AM signal is displayed on an oscilloscope, the modulation index can be computed from V_{max} and V_{min}, as shown in Figure 13.5. The peak value of the modulating signal E_m is one-half the difference of the peak and rough values:

$$E_m = \frac{V_{max} - V_{min}}{2}$$

The peak value of the carrier signal E_c is the average value of V_{max} and V_{min}.

$$E_c = \frac{V_{max} + V_{min}}{2}$$

Now we can write the modulation index as

$$m_a = \frac{V_{max} - V_{min}}{V_{max} + V_{min}} \qquad \text{---------- equation (13.4)}$$

In the given problem,

$$m_a = \frac{V_{max} - V_{min}}{V_{max} + V_{min}} = \frac{5.9 - 1.2}{5.9 + 1.2} = 0.662$$

Activity 13.1

The output signal from an AM modulator is

$$e(t) = 5\cos(1800\pi t) + 20\cos(2000\pi t) + 5\cos(2200\pi t)$$

Determine

1. the modulating signal, $e_m(t)$
2. the carrier signal $e_c(t)$
3. the modulation index.

Answers:

1. the modulating signal, $e_m(t) = \cos(2\pi 100t)$
2. the carrier signal $e_c(t) = 20\cos(2\pi 1000t)$
3. the modulation index = *1/2*

13.2.2 FREQUENCY MODULATION

In frequency modulation the frequency of the carrier signal is varied according to the modulating frequency while the amplitude of the modulated signal is kept constant. Then the carrier frequency will be changed by $\pm\Delta f$. This is called as the frequency deviation of the frequency modulated signal. The carrier's instantaneous frequency deviation from its unmodulated value varies in proportion to the instantaneous amplitude of the modulating signal

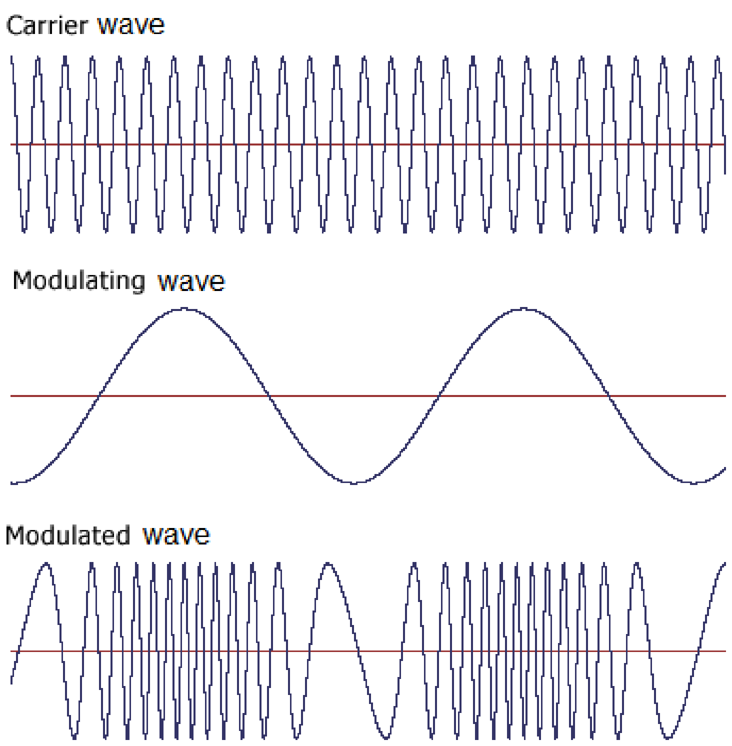


Figure 13.6 FM modulation

We can write the equation of a sine wave in generalised form as

$$e(t) = E \sin\theta$$

where $e(t)$ is instantaneous amplitude, E is peak amplitude and θ is the total angular displacement at time t.

A frequency modulated wave with sinusoidal modulation has its frequency varied according to the amplitude of the modulating signal. If Δf is the maximum deviation of frequency from average, then instantaneous frequency can be written as,

$$f = f_C + \Delta f \; Cos \; \omega_m t$$

or,

$$\omega = \omega_C + 2\pi\Delta f \; Cos\omega_m t$$

Now $\omega = \dfrac{d\theta}{dt}$

Integrating both sides,

$$\int \omega dt = \theta$$

$$\theta = \omega_c t + \frac{2\pi\Delta f}{\omega_m} Sin\omega_m t$$

$$e(t) = E_c \sin(\omega_c t + \frac{2\pi\Delta f}{\omega_m} \sin\omega_m t)$$

For a sinusoidal carrier and a sinusoidal modulating signal a frequency modulated signal can be given as

$$e(t) = E_c \sin(\omega_c t + \frac{\Delta f}{f_m} \sin\omega_m t) \qquad \text{-------------------- equation (13.5)}$$

Let, $\dfrac{\Delta f}{f_m} = m_f$

$$e(t) = E_c \sin\left(\omega_c t + m_f \sin\omega_m t\right) \qquad \text{-------------------- equation (13.6)}$$

where m_f is the modulation index of the FM wave.

For a given frequency deviation, modulation index varies inversely as the modulating frequency. The ability of the modulating signal to shift the carrier frequency can be understood if *the modulation index* is known.

It is clear from the above equation that the carrier frequency is sinusoidally varied between $f_c - \Delta f$ and $f_c + \Delta f$ when the modulating signal is varied sinusoidally.

When no modulation applied, the carrier is at its nominal frequency, i.e. the carrier frequency. The modulating signal causes the carrier frequency to deviate, i.e. to move above and below the nominal value. With the maximum possible deviations, the carrier frequency is moved up and down by the amount of the frequency deviation, thus, the bandwidth is about twice the frequency deviation. However, this would take up a very large amount of frequency spectrum.

S.A.Q. 13.2

The equation of a frequency modulated signal is given as

$$V_{fm}(t) = 1000 Sin\left[10^9 t + 4 \sin\left(10^4 t\right)\right]$$

Find, the carrier frequency, modulating frequency, modulation index and frequency deviation.

Answer

The frequency modulated wave can be represented by,

$$e = E_C \sin\left(\omega_C t + m_f \sin \omega_m t\right)$$

Carrier frequency, $f_c = \dfrac{\omega_c}{2\pi} = \dfrac{10^9}{2\pi}$ = 159.15 MHz

Modulating frequency, $f_m = \dfrac{\omega_m}{2\pi} = \dfrac{10^4}{2\pi}$ = 1.59 kHz

Modulation index m_f = 4 (by inspection)

$$\frac{\Delta f}{f_m} = m_f \rightarrow \Delta f = m_f f_m$$

Frequency deviation = 4 ×1.59 kHz = 6.36 kHz

Frequency modulation is used in a wide variety of radio communications applications from broadcasting to two-way radio communications links as well as mobile radio communications. It possesses many advantages over amplitude modulation and this is the reason for its widespread use. Some of these advantages are resilient to noise, resilient to signal strength variations, greater efficiency etc.

Activity 13.2

When a Sinusoidal carrier signal is frequency modulated using square wave what would be the shape of the resulting waveform? Sketch the waveform.

13.2.3 PULSE MODULATION

In phase modulation the phase of the carrier signal is varied according to the amplitude of the modulating signal. Thus, if *m(t)* is the message signal and $c(t) = A\cos\omega_C t$ then the modulated signal will be $F(t) = Acos(\omega_c t + \pi)$

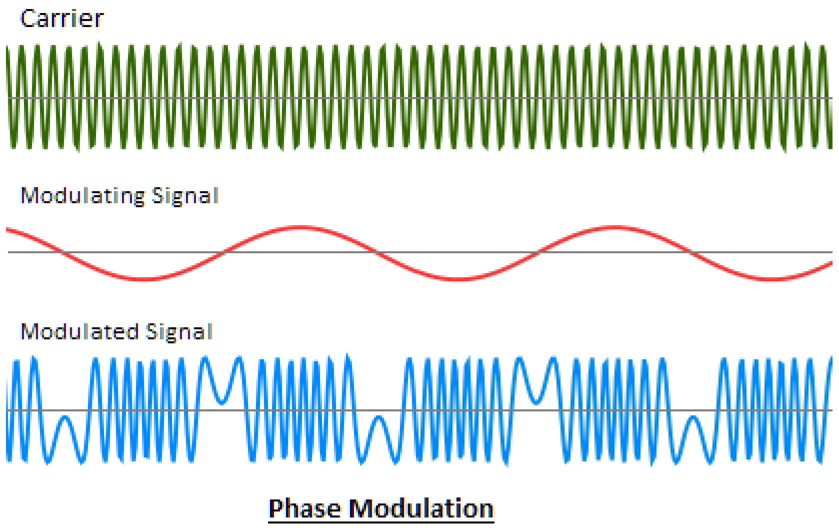


Figure 13.7: Phase modulation

Phase modulation does not affect with noise but the circuitry related is complex compared to AM and FM. This modulation technique is mostly used in satellite communication systems.

13.3 DIGITAL MODULATION

The process of digital modulation is same as the analog modulation except the modulating signal is a bit stream consist with logic '0' and logic '1'. Then it will embed to the high frequency carrier and transmitted. There are three basic types of modulation methods in digital modulation.

Amplitude shift key (ASK)

Frequency shift key (FSK)

Phase shift key (PSK)

These methods are use for different applications according to the available bandwidth and the data rates that needed to transmit via the channel.

13.3.1 AMPLITUDE SHIFT KEY (ASK)

Here carrier frequency is changed according to the amplitude of the digital data. Then the received modulated signal is transmitted via the channel. Figure 13.8 shows the waveforms before and after modulation using amplitude shift key (ASK).

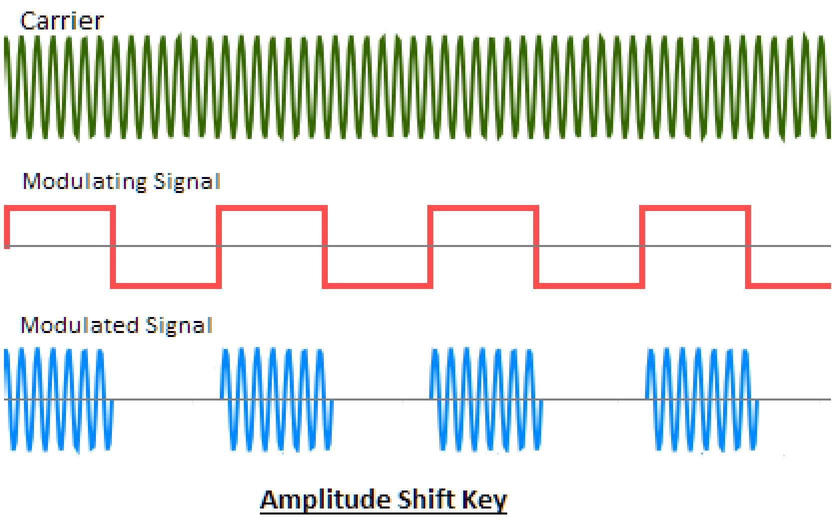


Figure 13.8

According to the figure 13.8, carrier signal is $\cos\omega_c t$ and Modulating signal amplitudes switching between logic '0' and logic '1'. If the modulated carrier $m(t)\cos\omega_c t$, we can see that the carrier amplitude presence and absent according to the signal modulating signal pattern. Because of this switching pattern ASK method is also known as on-off keying (OOK). The applications such as infrared remote controls and fiber optical transmitter and receiver use ASK modulation technique.

13.3.2 FREQUENCY SHIFT KEY (FSK)

Here frequency is changed according to the modulating signal. You can observe the frequency variation in the figure 13.9.

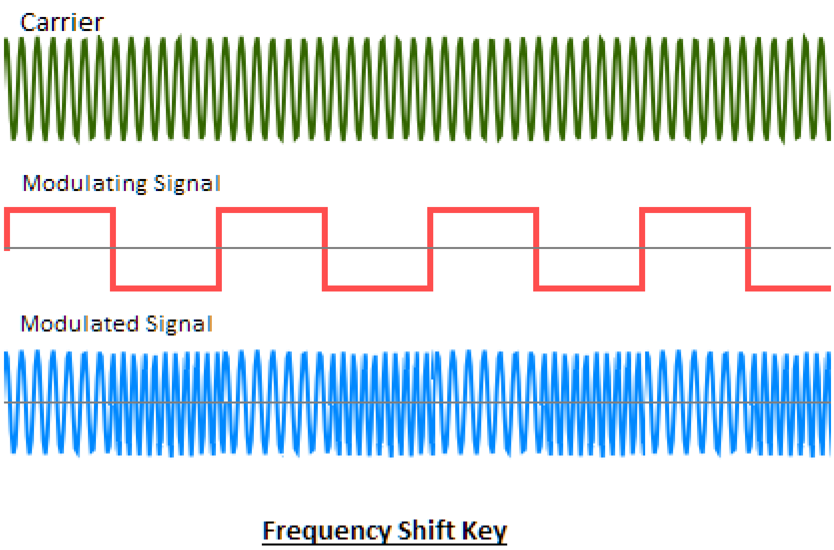


Figure 13.9

According to the figure 13.9 there are two predefined frequency ω_{c0} and ω_{c1}. Modulated signal has the ω_{c0} frequency, when the modulating bit is '0' and modulated signal has ω_{c1} frequency, when the modulating bit is '0'. Then the carrier transmitted

$cos\omega_{c0}$ and $cos\omega_{c1}$ with reference to the information signal (modulating bit). This modulation technique is mostly used in modems for telemetry applications.

13.3.3 PHASE SHIFT KEY (PSK)

This technique change the phase of the carrier depending on the information (modulating) signal and it is shown in figure 13.10.

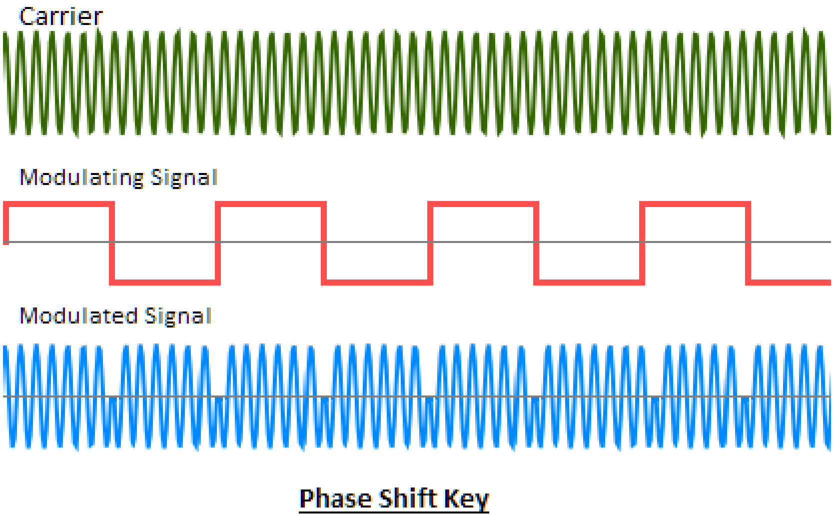


Figure 13.10

This technique uses phase shift under a logic state. If there is a logic '0', the modulated signal represent $cos\omega_c t$ and logic'1' represented by the $cos(\omega_c t + \pi)$. The phase modulation is used in ADSL broadband modems, mobile phones and satellite communication.

S.A.Q. 13.3

What are the advantages of digital modulation than the analog modulation?

Answer

Digital modulation capable of sending higher data rates

Less noise effect

13.3.4 ADVANCED DIGITAL MODULATION TECHNIQUES

Increased requirement of transmitting more and more data in the channels, advanced modulation techniques were developed to accommodate high data rates and high bandwidth requirement to send video. Binary Phase Shift Key (BPSK), Quadrature Phase Shift Key (QPSK) and Quadrature Amplitude Modulation (QAM) are commonly used methods in communication systems. You will be learning these techniques in advanced communication courses.

SUMMARY

In this session we discussed about the concept of modulation and how it is useful in communication systems. Then we studied three types of analog modulation techniques AM, FM and PM. We also learned the basic digital modulation techniques such as ASK, FSK and PSK. It is important to identify the differences between each technique and the relevant applications to be used in various communication systems. The necessity to transmit more data advanced modulation techniques such as BPSK, QPSK and QAM are used according to the requirement.

LEARNING OUTCOMES

Now you will be able to,

- Explain the importance of modulation and the process of modulation and demodulation.
- Explain the amplitude modulation and effect of modulation index.
- Compare the analog and digital modulation
- Explain the different modulation techniques and unique features
- Identify the applications related to each modulation technique.

Session 14
Power calculations in communication systems

Contents

INTRODUCTION

In the previous lesson, we discussed the modulation methods that used to transmit a low frequency base band signal over a distance. Signal power is another essential parameter to transmit a signal from one point to another in line and radio communication. As the signal propagates from the source to the destination the signal either gains or looses power depending on the design of the system.

Very often it is necessary to know the exact power available at a particular point in the communication system. The use of Logarithmic units for expressing power level facilitates speedy calculation of the signal level.

In this lesson we will study how to express a power ratio in decibels. We refer to power ratios when we use amplifiers and attenuators. We also will study different types of logarithmic units used and how corrections are made to measurements of level of power of speech channels to allow for the response of the human ear.

14.1 POWER GAIN EXPRESSED IN DECIBELS

Decibel is defined as ten times the logarithm of the ratio of output power to input power to the base 10.

Consider the network given below.

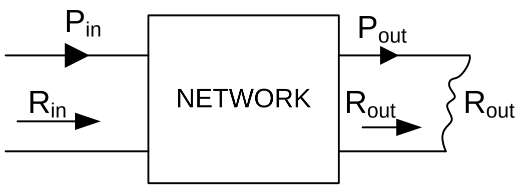


Figure 14.1

The power gain "G" is given by the expression

$$G = 10\log_{10}\left(\frac{P_{out}}{P_{in}}\right) \qquad \text{------------------------------ equation (14.1)}$$

If the input resistance is R_{in} and the output resistance is R_{out}

$$P_{out} = \frac{V_{out}^2}{R_{out}} \ \ and\ \ P_{in} = \frac{V_{in}^2}{R_{in}}$$

Where V_{in} and V_{out} are the input and output voltages respectively.

$$\therefore G = 10\ \log\left[\frac{V_{out}^2}{R_{out}} \div \frac{V_{in}^2}{R_{in}}\right] dB$$

$$= 10\ \log\left[\frac{V_{out}^2}{{V_{in}}^2} \times \frac{R_{in}}{R_{out}}\right] dB$$

$$= \left[10\log\frac{V_{out}^2}{V_{in}^2} + 10\ \log\frac{R_{in}}{R_{out}}\right] dB$$

If match the impedance to obtain the maximum power deliver, hence $R_{in} = R_{out}$.

$$10\log\frac{R_{in}}{R_{out}} = 10\log 1 = 0$$

$$G = 10\log\frac{V_{out}^2}{V_{in}^2}\ dB$$

$$G = 20\log\frac{V_{out}}{V_{in}}\ dB \qquad \text{------------------------------ equation (14.2)}$$

if $G > 0$, then, $V_{out} > V_{in}$

14.1.1 POWER GAIN EXPRESSED IN NEPERS

The logarithm of output voltage to input voltage to the base '*e*' is defined as the Neper. The number '*e*' is an important mathematical constant, approximately equal to 2.71828, that is the base of the natural logarithm. The natural logarithm is generally written as ln x or $\log_e x$.

Therefore, the value in nepers is given by $Np = ln\ (x_1 / x_2)$, where x_1 and x_2 are the values of interest, and ln is the natural logarithm, *i.e.,* logarithm to the base *e*.

i.e. $1\ Np \equiv 20 / ln(10) = 8.685889638\ dB$
$1\ dB \equiv ln(10) / 20\ Np = 0.115129255\ Np$

Now consider the network supplying a load current I_{out} when excited by an input current. Assume that $R_{in} = R_{out}$.

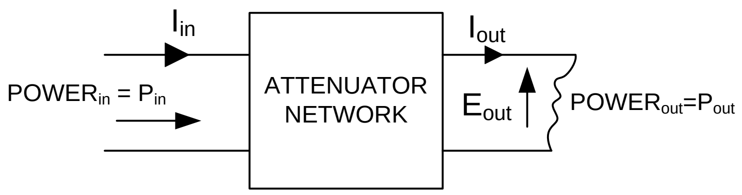


Figure 14.2

Express the gain in decibels and in Nepers.

Gain expressed in decibels

$$G = 10\log\left(\frac{P_{out}}{P_{in}}\right) = 20\log\left(\frac{I_{out}}{I_{in}}\right) dB$$

Gain expressed in Nepers

$$G\,\text{exp}\,ressed\ \ in\ nepers = \log_e\left(\frac{I_{out}}{I_{in}}\right) = \ln\left(\frac{I_{out}}{I_{in}}\right)$$

Now we can convert a decibel value into Nepers or vice versa as follows:

Neper value ≡ dB value × ln (10) / 20 = dB value × 0.115129255
dB value ≡ Np value × 20 / ln (10) = Np value × 8.685889638

So far we have considered the power gain of a given system. We will now consider a lossy system.

In the equation 14.2, what will be the effect if $G < 0$ when $V_{out} < V_{in}$? Then we say that the signal has been attenuated, i.e, the signal has dissipated a part of its energy while traversing through the medium.

Then we express the attenuation $\alpha = \dfrac{1}{G}$

As $\alpha = 10\log\left(\dfrac{P_{in}}{P_{out}}\right) dB \qquad (if\ \ R_L = R_{in})$

$$\alpha = 20\log\left(\frac{V_{in}}{V_{out}}\right) dB$$

or $\alpha = 20\log\left(\dfrac{I_{in}}{I_{out}}\right) dB$

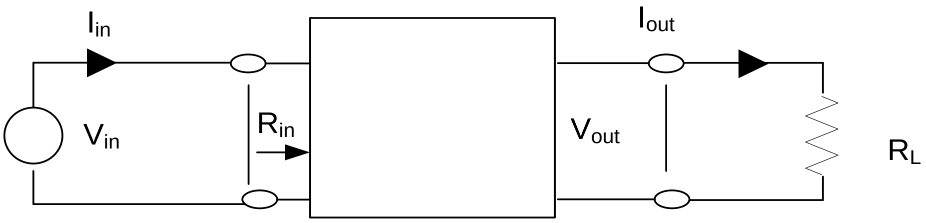


Figure 14.3

Power ratio in decibels = $10\log\left(\frac{P_{out}}{P_{in}}\right) = 20\log\left(\frac{I_{out}}{I_{in}}\right) = 20\log\left(\frac{V_{out}}{V_{in}}\right)$

provided $R_{out} = R_{in}$.

In communication systems we always "match' circuits to ensure maximum power transfer. That is the source impedance is made equal to the load impedance. (Source impedance is similar to the internal resistance of a battery.) A communication system consists of several components such as amplifiers and attenuators. If the gain or loss of each such component is known we can easily calculate the power at the output when the input power is known.

Example 14.1

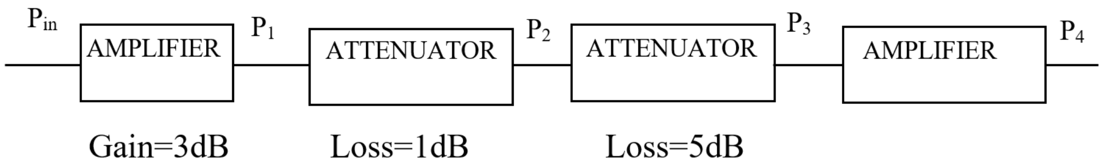


Figure 14.4

Total Gain $= \frac{P_4}{P_{in}} = \frac{P_1}{P_{in}} \times \frac{P_2}{P_1} \times \frac{P_3}{P_2} \times \frac{P_4}{P_3}$

What is the overall gain or loss ?

$$G = 10\log\left(\frac{P_4}{P_{in}}\right) = 10\log\left(\frac{P_1}{P_{in}}\right) + 10\log\left(\frac{P_2}{P_1}\right) + 10\ \log\left(\frac{P_3}{P_2}\right) + 10\ \ \log\left(\frac{P_4}{P_3}\right)$$

$$= 3 - 1 - 5 + 10 = +7\ dB\,\textit{above the input power}$$

As the answer is positive there is a net gain.

Note: If we used the numerical power ratio we would have to multiply the ratios to get this answer. By using the decibel value of the power ratio we are able to add the values and hence compute the value speedily.

If the input power is 1 milliwatts (10^{-3} watt), what will be the output power?

+7 dB can be converted to a ratio,

10 log x = 7

$$x = anti\log 0.7 = \frac{P_{out}}{P_{in}}$$

$$P_{out} = 5 \times 1 = 5\,mW$$

We can also say that the output is 7 dB above 1 mW.

If P_{out} is less than P_{in}, log (P_{out}/P_{in}) will be negative indicating that the signal has undergone a loss during transmission from the input to the output.

Let us consider our earlier example.

M_1	=	0.5,	10 log(0.5)	=	-3.013 dB
M_2	=	2,	10 log(2)	=	+3.013 dB
M_3	=	0.1,	10 log(0.1)	=	-10 dB
M_4	=	5,	10 log(5)	=	+6.9897 dB

(M1, M2…………..M4 are the ratios of P_{out} to P_{in})

$$\frac{P_{out}}{P_{in}}\,in\,decibels = -3.0103 + 3.0103 - 10 + 6.9897$$

$$= -3.0103$$

The answer is negative. What does this indicate? A Power loss

We can see that there is an overall loss.

In the above example we considered the power ratio M_1, M_2, etc. which can be multiplied to obtain the overall power gain.

For example: To obtain the overall power gain, typical values of power ratios of a transmission system are quoted below.

M_1	=	0.463
M_2	=	2.56
M_3	=	0.096
M_4	=	5.136

As stated earlier determination of overall power ratio using such values involves a tedious multiplication. Further we could have a very much more complex system having several networks in series.

How can we simplify this calculation?

The process of addition is much simpler than multiplication. Let us see whether power ratios can be converted to a form suitable for addition. We know from mathematics that

$$\log(M_1 \times M_2 \times M_3 \times M_4) = \log M_1 + \log M_2 + \log M_3 + \log M_4$$

Therefore, if we express the ratio M in logarithmic form we can add individual values. Now you will understand the reason for expressing the overall power gain in terms of the power gains of individual units in the above example.

In practice we convert each value of M into DECIBELS which is the logarithmic value of the ratio M.

Let us now consider several networks connected in series.

This can be represented by

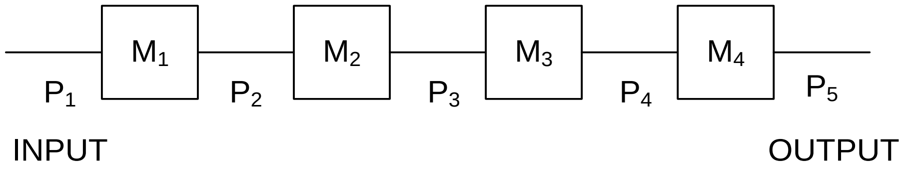


Figure 14.5

What is the output to input power ratio?

$$\frac{P_5}{P_1}$$

Now try to recall what we did in the previous example. If you recalled correctly then you will write the expression as,

$$\frac{P_5}{P_1} = \frac{P_2}{P_1} \times \frac{P_3}{P_2} \times \frac{P_4}{P_3} \times \frac{P_5}{P_4}$$

Thus, you are expressing the overall power ratios in terms of the individual output to input power ratios of each network (represented by M_1, M_2, M_3 and M_4).

$$\frac{P_5}{P_1} = M_1 \times M_2 \times M_3 \times M_4$$

Example 14.2

Let us consider another example.

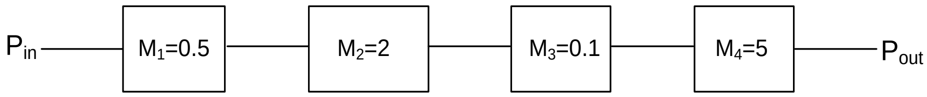


Figure 14.6

If P_{in} = 1 watt, how much is P_{our}?

$$\frac{P_{out}}{P_{in}} = M_1 \times M_2 \times M_3 \times M_4$$
$$= 0.5 \times 2 \times 0.1 \times 5$$
$$= 0.5$$
$$P_{out} = 0.5 \times 1\ watt$$
$$= \frac{1}{2}\ watt$$

Let us consider a simple circuit where a signal generator is connected to a load through an electrical network.

We can represent this by,

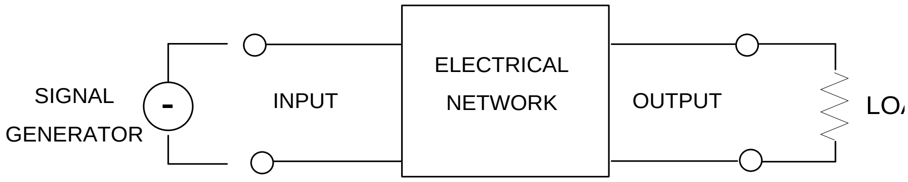


Figure 14.7

We need not know what the electrical network consists of. It has an input and an output.

We will now connect a power meter across the input of the network and another power meter across the output.

What will the meters read?

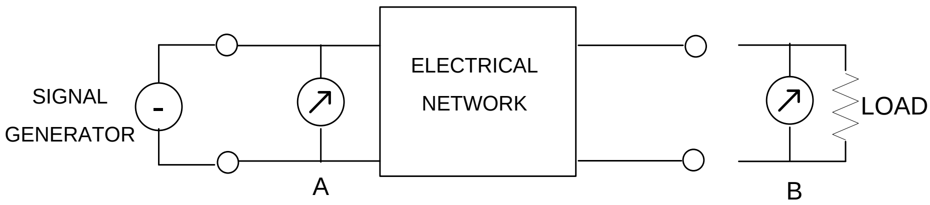


Figure 14.8

Meter A will read the power entering the electrical network.

Meter B will read the power going out of the network into the load.

If A reads P_1 watts and B reads P_2 watts then the ratio of output power to input power is P_2/P_1.

If P_2 is greater than P_1 then we know that the network has caused a power gain. On the other hand if P_2 is less than P_1, then the network has caused a power loss.

Attenuation in Nepers $= \log_e\left(\frac{I_{in}}{I_{out}}\right) = \log_e\left(\frac{E_{in}}{E_{out}}\right) = \frac{1}{2}\log_e\left(\frac{P_{in}}{P_{out}}\right)$

If the resistances at the input and output are equal then attenuation can be readily converted from one notation to the other.

$$dB = 20\log_{10}\left(\frac{I_{in}}{I_{out}}\right)$$

attenuation in dB $\quad = 20\log_{e}\left(\frac{I_{in}}{I_{out}}\right)\times\log_{10} e$

$$= 8.686 \times attenuetion\, in\, nepers$$

Therefore, attenuation in decibels = 8.686 x attenuation in nepers (provided $R_1 = R_2$)

When do we use the unit Neper and when do we use decibel?

The decibel is defined in terms of logarithm to the base 10 and therefore, is very convenient to use in practice.

On the other hand, the Neper is defined in terms of the exponential 'e' and therefore is convenient for theoretical work.

S.A.Q. 14.1

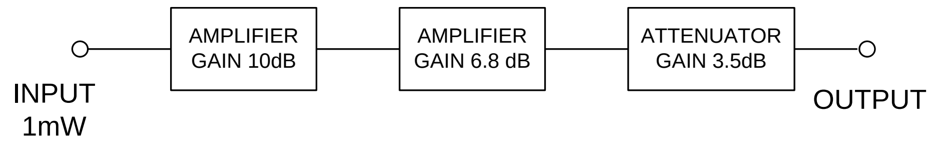


Figure 14.9

Calculate the power at the output in mW.

Answer

Total gain $\quad = 10 + 6.8 – 3.5$

$\quad = 13.3$ dB

To determine the actual output power we must know the gain ratio - that is we must convert dB into a ratio.

$$10\log\left(\frac{P_2}{P_1}\right) = 13.3$$

$$\left(\frac{P_2}{P_1}\right) = anti\log\left(\frac{13.3}{10}\right) = 21.3796$$

Output power = 1mW x 21.3796 = 21.4mW

14.2 DBM AND DBW

In the earlier section we learned how to express a power ratio in decibels. We refer to power ratios when we use amplifiers and attenuators. We may need to know the actual power at a point.

Can we use logarithmic units to express actual power?

The decibel is fundamentally a unit of power ratio. However, we can express power available at a point by comparing it with a standard reference value say 1 milliwatt or 1 watt and stating that it is so many times greater than 1 milliwatt or 1 watt. In other words we can express the power ratio relative to 1 milliwatt or 1 watt.

The logarithm of this ratio to the base ten gives a measure of the power level in decibels.

Example 14.3

The power at *X* is 100 milliwatts. Express this power in decibels with reference to 1 mW.

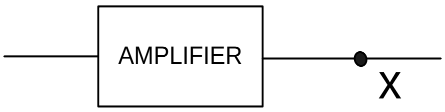


Figure 14.10

$$Ratio = \frac{100\,mW}{1\,mW} = 100$$

In decibels,

10 log 100=20 dB

The power at *X* is 20 decibels more than 1 mW.

We are now expressing the actual power in decibels. Instead of stating that the power at *X* is 20 decibels above 1mw we simply denote writing the decibel value as 20 *dBm*.

The suffix m denotes that the level of reference power is 1 mW.

Example 14.4

If the power at a point in a system is 40 dBm, what is the actual power in milliwatts?

We know by definition that,

$$10\log\left(\frac{P}{1mw}\right)_P = 40$$

$$P = anti\log 4 = 10^4$$

$$Power = \underline{\underline{10,000mW}}$$

Example 14.5

(To illustrate the application of dBW)

What do we mean by - the power at a point is 30 dBW? The suffix W refers to 1 watt. Therefore, power at that point is 30 dB above 1 watt.

$$10\log\left(\frac{P}{1watt}\right)=30$$

$$Power=\underline{\underline{1000watts}}$$

Example 14.6

Calculate the power in dBm at the output of the system shown. in Figure.14.11.

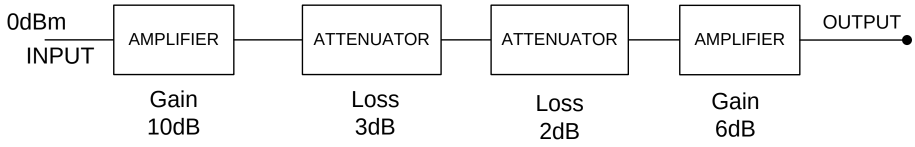


Figure 14.11

Overall gain (or loss?)

= 10 - 3 - 2 + 6 = 11 dB.

As the answer is positive we know that it is a gain. Input level is 0 dBm.

Output power level = 0 dBm + 11 dB = 11 dBm (i.e 11 dB above 1 mW.)

If the input level was 5 dBm instead of 0dBm, The output level would have been

5 dBm + 11 dB = 16 dBm

Example 14.7

How can we add dBm and dB together?

11 dB gain means - whatever the input level, the output level is 11 dB greater than the input. The input in the above example (second) is 5 dBm - which is the absolute value of the power expressed in logarithmic units. Therefore, output is 11 dB higher than 5 dBm.

Example 14.8

What is the actual power of 0 dBm?

$$10\log\frac{P}{1mW}=0$$

P = 1 mW.

Note that 0dBm is not 0 mW.

14.3 REPEATERS

When electrical signal travels along the medium it is subjected to attenuation. This attenuation is a characteristic of medium and also depends on the distance between the transmitter and the receiver. Suppose the strength of the electrical signal received at the receiver is very low, so that the receiver is not being able to reproduce what has been transmitted. The minimum signal strength required by the receiver for satisfactory reproduction is called the receiver THRESHOLD level.

Suppose the distance between the transmitter and the receiver is very high and the signal level required is below the given threshold what improvements would you suggest to maintain communication?

a) to reduce the distance between the two ends until the signal level received is well above receiver threshold. OR

b) to increase the signal level at the sending end. OR

c) to employ some method to raise the level at any intermediate point between the send and receive points.

If you adopt the first method, the distance between send and receive points should be decreased (limiting the distance). This is not practicable. But by adopting (b) and (c) we can satisfactorily overcome the restriction imposed by distance in (a) and can raise the level of signal strength sufficiently to match with the threshold level of the receiver.

How can you increase the signal levels? This is done by employing an amplifier.

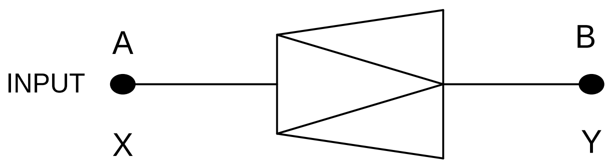


Figure 14.12 Symbol of an amplifier

If the signal level at the amplifier input is *X* and at the output is *Y*, the *Y/X* ratio is defined as the gain of the amplifier.

$$G = \frac{Y}{X} = Gain\ of\ the\ amplifier$$

ie. $Y>X$ always.

If the gain *G* is greater than 10 we refer to the amplifier as a high gain amplifier and *G* is low it is called a low gain amplifier. The design and construction of high gain amplifier is very much more expensive than the design of low gain amplifier.

Is it possible to employ number of low gain amplifiers to make a high gain amplifier? Yes then how?

Let me explain further...

Given three amplifiers each having a gain of "2" how would you connect to get an overall amplification of 8 times.

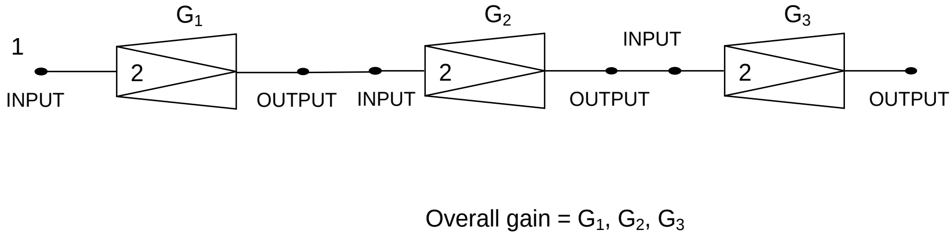
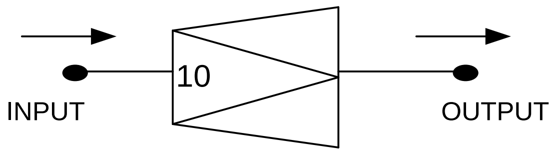


Figure 14.13

By connecting the three amplifiers as shown above an overall gain of 8 can be obtained, between the points" 1" and "2" (in this case output of one amplifier is directly connected to the input of the subsequent amplifier.) This method of connection is known as cascaded connection. In other words we can say three of "2" gain amplifiers are cascaded to make an overall amplification gain of $2^3 \rightarrow 8$.

Why you differentiate the two ends of an amplifier as input and output?

To differentiate the direction of signal flow, the signals to be amplified are fed to the input and the amplified signals are taken out from the output terminal of the amplifier.



A current amplifier

Figure 14.14

Example 14.9

(a) 0.1 ampere current is fed the input of the amplifier gain" 10". What will be current at the output?

By the definition,

Output = gain x input

= 10 x 0.1 = 1 Ampere

Amplifiers, which amplify only current, are called current amplifiers.

Example 14.10

0.1 volt is fed to a voltage amplifier of gain 20 what is the voltage at the output?

Output = 20 x 0.1 = 2 Volts

Can you name the input as output and output as input of this amplifier?

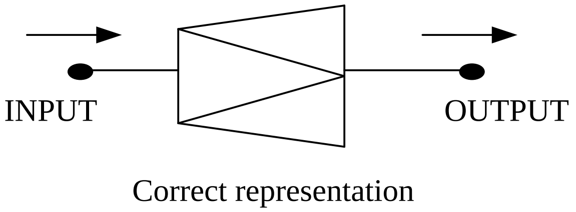


Figure 14.15 (a)

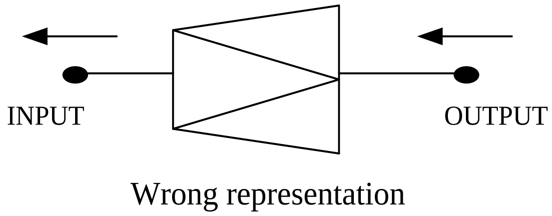


Figure 14.15 (b)

Since the amplifiers that are practicable can amplify signals only in one direction as shown in figure 14.15 (a), and it is not possible to send a signal from output as shown in the figure 14.15 (b). (Amplifiers are unidirectional)

There are special types of amplifiers, which permit signal flow in both directions. The treatment of such amplifiers is not within the scope of this course.

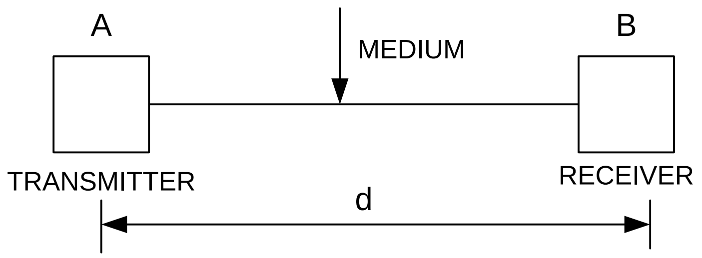


Figure 14.16 (a)

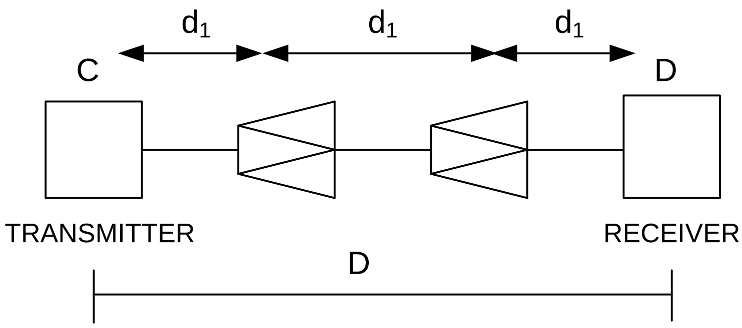


Figure 14.16 (b)

Figure 14.16(a) shows a simple communication system between A and B points. Assume the threshold level of the receiver is "*X*". Suppose the same transmitter and the receiver is used between *C* and *D* points (length is $D > d_1$) without amplifiers. In that case, the signal level received at *D* will be less than "*X*". Hence this system cannot be used between *C* and *D* points. By employing two amplifiers one at d away from '*C*' and another $2d_1$ away from *C* the signal strength received at *D* can be increased up to the required level of the receiver. (figure 4.16 (b) shows the configuration with two amplifiers.)

Is it possible to establish communication using only one amplifier? Yes depending on the threshold level of the amplifier.

What is a repeater?

We can define the repeater as a device where the energy dissipated by the signal when traversing through the transmission medium is restored without destroying the shape of the signal.

A Repeater is primarily an amplifying station for signals but there are other functions too. Such as,

(a) amplitude/frequency correction of the system.

(b) delay/frequency correction of the system.

(c) temperature correction in case of underground repeater etc.

The details of (a), (b) and (c) will not be explained in this course.

Example 14.11

A transmitter and a receiver are placed between two points *A* and *B* as shown in Figure 14.17 and the following data are given;

distance between A and B = 10km

signal attenuation = 2 dB per km

transmitting voltage 1 volt

receiver threshold level 0.2 volts.

Calculate the gain of each amplifier. (Both amplifiers have equal gain).

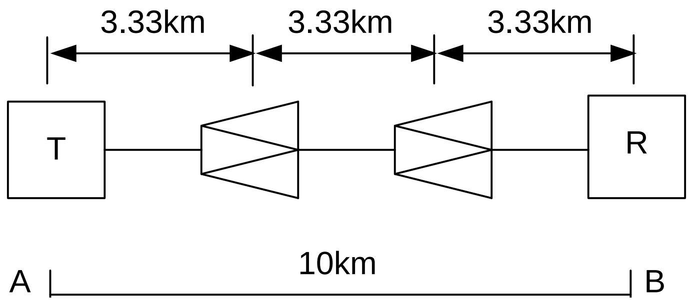


Figure 14.17

Solution

Let us take power $= V^2$

$P = V^2$

10 log P	=	20 log V
$\therefore$ power	=	20 log V dB

[according to the definition of dB]

Let the gain of both the amplifiers be G dB.

Total attenuation 2x10dB = 20dB

The signal level at transmitter point	= 1V = 0 dB
$\therefore$ The signal level at the receiver point	= 0 - 20 + G dB
	= -20 + 2G dB
But receiver threshold level	= 0.2V
	= 20 log 0.2
	= -13.98 dBW

For satisfactory reproduction at the receiver,

Threshold level ≤ Actual receiver level

$-13.98 \le -20 + 2G$

$G \ge 3.01 dB$

$10 \log G \ge 3.01$

$G \ge 2$

$Minimum \;\; gain = 2$

Note:

The positions of the repeaters are independent from the answer.

Thus, the position of the repeater can be anywhere in between A & B. But for uniformity they have placed in equal distances.

In this particular case the signals will flow in the T to R direction. But a telecommunication system should be provided with both way communication facility. This can be achieved by employing a second system for the opposite direction. (Figure 14.18)

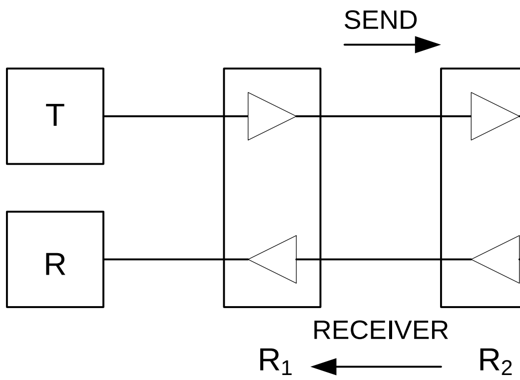


Figure 14.18

Here we have two independent systems, one for 'Send' direction and the second for 'Receive' direction. The repeaters R_1, and R_2 have two amplifiers one amplifying signals in the Send direction and the other is amplifying the Receive direction signals.

Now let me explain a situation where only one path is provided for both directions. At a repeating point of the medium of transmission the two signals. Send and receive signals have to be separated for the purpose of amplification and after amplification two signals have to be recombined and send in the common path.

How you can separate the signals 'Send' and 'Receive'?

The device which is being used to separate the paths is called a hybrid. How could you recombine the two paths after amplification?

Same hybrid can also be used for this purpose too.

Now we can explain the functions of a hybrid circuit as

(a). To separate the "Send and Receive" signals into two individual paths and

(b). To recombine the separated paths into one path.

Figure 14.19 shows two circuit symbols for a hybrid circuit.

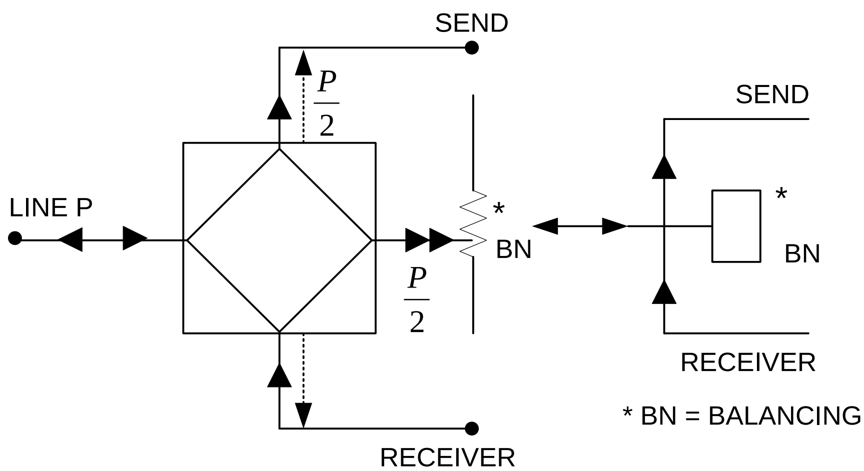


Figure 14.19

SUMMARY

In this session we studied how to perform the power calculations of a communications system. We learned different types of logarithmic units used to represent power values. Also, we studied about the function of repeaters and their operation in communication systems.

LEARNING OUTCOMES

Now you will be able to,

- Determine overall gain or loss of a system using decibels.
- Express absolute power levels in dBm and dBW.
- Identify functions of a repeater.
- Carry out simple calculations of communication systems in decibles

Pleace print

....
Subject Officer

....
Approved by

....
Date