POST-GRADUATE DIPLOMA IN COMPUTER APPLICATION

PAPER : PGDCA - 5
INTRODUCTION TO COMPUTER NETWORK, INTERNET AND E-COMMERCE

SECTION-A

Department of Distance Education
Punjabi University, Patiala

LESSON NO:

POST-GRADUATE DIPLOMA IN COMPUTER APPLICATION

PAPER : PGDCA - 5
INTRODUCTION TO COMPUTER NETWORK, INTERNET AND E-COMMERCE

LESSON NO. 1.1

AUTHOR: GURINDER GOSAL

COMPUTER NETWORKS

Structure of the Lesson:

Learning Objectives:

The major objectives of this lesson are to:

- Introduce the concept of computer network.
- Describe the general characteristics of a computer network.
- Understand the role of the major components of a computer network.
- Discuss the various benefits and costs of using computer networks.

- Discuss the need of computer networks in different areas.
- Understand the role of network hardware and network software in computer networks.

1.1.1 Introduction

This lesson introduces the fundamental concepts of computer networks. A network is simply a collection of computers or other hardware devices that are connected together, either physically or logically, using special hardware and software, to allow them to exchange information and cooperate. Networking is the term that describes the processes involved in designing, implementing, upgrading, managing and otherwise working with networks and network technologies. The old model of a single computer serving all of the organization's computational needs has been replaced by one in which a large number of separate but interconnected computers do the job called computer networks. A computer network is the infrastructure that allows two or more computers (called hosts) to communicate with each other. The network achieves this by providing a set of rules for communication called protocols, which should be observed by all participating hosts. The need for a protocol should be obvious: it allows different computers from different vendors and with different operating characteristics to 'speak the same language'. Two computers are said to be interconnected if they are able to exchange information. The connection need not be via a copper wire only but fiber optics, microwaves, infrared and communication satellites can also be used. Networks come in many sizes, shapes and forms.

1.1.2 Definition of Computer Network

A computer network is the infrastructure that allows two or more computers (called hosts) to communicate with each other. Figure 1 shows an abstract view of a network and its hosts. The network is made up of two types of components: nodes and communication lines.

The nodes typically handle the network protocols and provide switching capabilities. A node is usually itself a computer (general or special) which runs specific network software.

The communication lines may take many different shapes and forms, even in the same network. Examples include: copper wire cables, optical fiber, radio channels, telephone lines and communication satellites.

A host is connected to the network by a separate communication line which connects it to one of the nodes. In most cases, more than one host may be connected to the same node. From a host's point of view, the entire network may be viewed as a black box, to which many other hosts are connected. Each host has a unique address allocated to it by the network. For a host to communicate with another host, it needs to know the latter's address. All communication between hosts passes through the nodes, which in turn determine how to route messages across the network from one point to another.

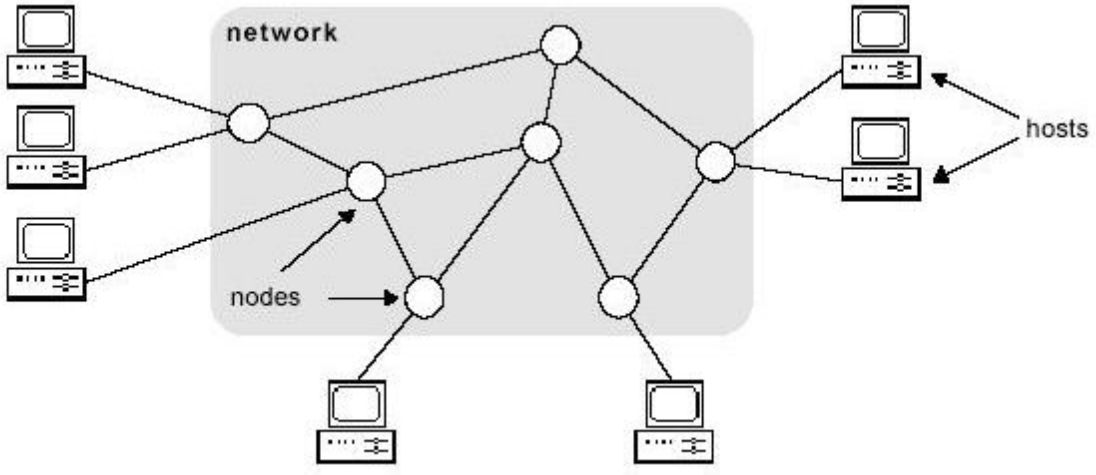


Figure1. An Abstract Network

1.1.3 The Advantages of Networking

A network is not just a bunch of computers with wires running between them. When properly implemented, a network is a system that provides its users with unique capabilities, above and beyond what the individual machines and their software applications can provide.

Most of the benefits of networking can be divided into two generic categories: connectivity and sharing. Networks allow computers, and hence their users, to be connected together. They also allow for the easy sharing of information and resources and cooperation between the devices in other ways. Since modern business depends so much on the intelligent flow and management of information, this tells you a lot about why networking is so valuable.

Here are some of the specific advantages generally associated with networking:

- **Data Sharing:** One of the most important uses of networking is to allow the sharing of data. True networking allows thousands of employees to share data much more easily and quickly than this. More so, it makes possible applications that rely on the ability of many people to access and share the same data, such as databases, group software development and much more. Intranets and extranets can be used to distribute corporate information between sites and to business partners.
- **Connectivity and Communication:** Networks connect computers and the users of those computers. Individuals within a building or work group can be connected. Once connected, it is possible for network users to communicate with each other using technologies such as electronic mail. This makes the transmission of business (or non-business) information easier, more efficient

and less expensive than it would be without the network.

- **Sharing of Hardware:** Networks facilitate the sharing of hardware devices. For example, instead of giving each of 20 employees in a department an expensive color printer, one printer can be placed on the network for everyone to share.
- **Data Security and Management:** In a business environment, a network allows the administrators to manage the company's critical data much better. Instead of having this data spread over dozens or even hundreds of small computers in a haphazard fashion as their users create it; data can be centralized on shared servers. This makes it easy for everyone to find the data, makes it possible for the administrators to ensure that the data is regularly backed up, and also allows for the implementation of security measures to control how one can read or change various pieces of critical information.
- **Access to Internet:** The Internet is itself an enormous network, so whenever you access the Internet, you are using a network. The significance of the Internet on modern society is hard to exaggerate, especially for those of us in technical fields.
- **Sharing of Internet Access:** Small computer networks allow multiple users to share a single Internet connection. Special hardware devices allow the bandwidth of the connection to be easily allocated to various individuals as they need it, and permit an organization to purchase one high-speed connection instead of many slower ones.
- **Enhancing Performance and Balancing:** Under some circumstances, a network can be used to enhance the overall performance of some applications by distributing the computation tasks to various computers on the network.
- **Entertainment:** Networks facilitate many types of games and entertainment. The Internet itself offers many sources of entertainment, of course. In addition, many multi-player games exist that operate over a local area network. Many home networks are set up for this reason and gaming across wide area networks (including the Internet) has also become quite popular.

1.1.4 The Disadvantages (Costs) of Networking

Networking does have some real and significant costs and drawbacks associated with it. Here are a few of the items that balance against the advantages of networking.

- **Setup Costs of Network Hardware, Software:** Computers don't just magically network themselves, of course. Setting up a network requires an investment in hardware and software, as well as funds for planning, designing and implementing the network. For a home with a small network of two or three PCs, this is relatively inexpensive with today's low prices for network hardware, and operating systems already designed for networks. For a large company, cost can easily run into a very big amount.
- **Hardware and Software Management and Administration Costs:** In all but the smallest of implementations, ongoing maintenance and management of

the network requires the care and attention of an IT professional. In a smaller organization that already has a system administrator, a network may fall within this person's job responsibilities, but it will take time away from other tasks. In more substantial organizations, a network administrator may need to be hired, and in large companies an entire department may be necessary.

- **Data Security Concerns:** If a network is implemented properly, it is possible to greatly improve the security of important data. In contrast, a poorly-secured network puts critical data at risk, exposing it to the potential problems associated with hackers, unauthorized access and even sabotage.
- **Undesirable Sharing:** With the good comes the bad; while networking allows the easy sharing of useful information, it also allows the sharing of undesirable data. One significant "sharing problem" in this regard has to do with viruses, which are easily spread over networks and the Internet. Mitigating these effects costs more time, money and administrative effort.
- **Illegitimate or Undesirable Behavior:** Similar to the point above, networking facilitates useful connectivity and communication, but also brings difficulties with it. Typical problems include abuse of company resources, distractions that reduce productivity, downloading of illegal or illicit materials and even software piracy. In larger organizations, these issues must be managed through explicit policies and monitoring, which again, further increases management costs.

Most of these costs and potential problems can be managed. In the end, as with any other decision, whether to network or not is a matter of weighing the advantages against the disadvantages. Of course today, nearly everyone decides that networking is worthwhile.

1.1.5 Need for Computer Networks

Why people are interested in computer networks and what they can be use for. If there is no need or use of them then no body will be interested in computer networks. We will see the use of computer networks in traditional areas as well new advance fields in the following section.

1.1.5.1 In Business Applications

Many companies have a substantial number of computers and usually a company may have separate computers to monitor production, keep track of inventories, and do the payroll. Primarily, each of these computers may have worked in isolation from the others, but at some point, management may have decided to connect them to be able to extract and correlated information about the entire company. The goal of resource sharing is to make all programs, equipment and especially data available to anyone on the network without regard to the physical location. An obvious example is having a group of office workers share a common printer. None of the individuals really needs a private printer and a high-volume networked printer is often cheaper, faster and easier to maintain than a large collection of individual printers

Probably even more important than sharing physical resources is sharing information. Every large and medium-sized company and many small companies are vitally dependent on computerized information. Most companies have a large plethora of information such as customer records, inventories, accounts receivable, financial statements, tax information, and much more online. If all of its computers went down, a bank could not last more. Even a small travel agency or three-person law firm is now highly dependent on computer networks for allowing employees to access relevant information and documents instantly.

The goal of geographical independence is to liberate usage of information and resources from locations. For smaller companies, all the computers are likely to be in a single office or perhaps a single building, but for larger ones, the computers and employees may be scattered over number of offices and plants in different countries. Nevertheless, a sales person in one city might sometimes need access to a product inventory database in other city in different country. In other words, the mere fact that a user happens to be thousands of kilometers away from his data should not prevent him from using the data as though they were local.

Simply one can imagine a company's information system as consisting of one or more databases and some number of employees who need to access them remotely. In this model, the data are stored on powerful computers called servers. Often these are centrally housed and maintained by a system administrator. In contrast, the employees have simpler machines, called clients, on their desks, with which they access remote data. The client and server machines are connected by a network, as illustrated in Fig. 2. Note that the network is shown as a simple oval, without any detail to mean a network in the abstract sense. This whole arrangement is called the client-server model. It is widely used and forms the basis of much network usage. It is applicable when the client and server are both in the same building, but also when they are far apart. For example, when a person at home accesses a page on the World Wide Web, the same model is employed, with the remote Web server being the server and the user's personal computer being the client. Under most conditions, one server can handle a large number of clients.

If we look at the client-server model in detail, we see that two processes are involved, one on the client machine and one on the server machine. Communication takes the form of the client process sending a message over the network to the server process. The client process then waits for a reply message. When the server process gets the request, it performs the requested work or looks up the requested data and sends back a reply as shown in Fig.

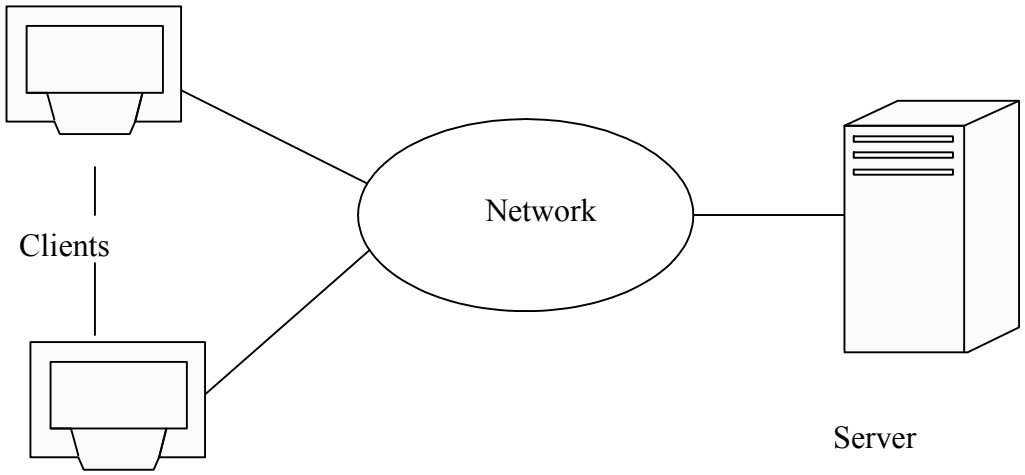


Figure 2: A network with clients and server

Also a goal of setting up a computer network is that it can provide a powerful communication medium among employees. Virtually every company that has two or more computers now has e-mail which employees generally use for a great deal of daily communication. But e-mail is not the only form of improved communication made possible by computer networks. With a network, it is easy for two or more people who work far apart to accomplish a task together. When one worker makes a change to an online document, the others can see the change immediately, instead of waiting several days for a letter. Such a speedup makes cooperation among far-flung groups of people easy where it previously had been impossible. Yet another form of computer-assisted communication is videoconferencing. Using this technology, employees at distant locations can hold a meeting, seeing and hearing each other and even writing on a shared virtual blackboard. Videoconferencing is a powerful tool for eliminating the cost and time previously devoted to travel.

Another goal for increasingly many companies is doing business electronically with other companies, especially suppliers and customers. For example, manufacturers of automobiles, aircraft and computers, among others, buy subsystems from a variety of suppliers and then assemble the parts. Using computer networks, manufacturers can place orders electronically as needed. Being able to place orders in real time reduces the need for large inventories and enhances efficiency. It reduces costs in managing orders and interacting with a wide range of suppliers and trading partners, areas that typically add significant overhead to the cost of products and services.

Also a goal that is starting to become more important is doing business with consumers over the Internet. Airlines, bookstores and music vendors have discovered that many

customers like the convenience of shopping from home. Consequently, many companies provide catalogs of their goods and services online and take orders on-line. This sector is expected to grow quickly in the future. It is called e-commerce (electronic commerce). Electronic commerce is a modem business methodology that addresses the needs of organizations, merchants and consumers to cut costs while improving the quality of goods and services and increasing the speed of service delivery. The term also applies to the use of computer networks to search and retrieve information in support of human and corporate decision making.

EDI is the electronic transfer of structured business documents in an organization - internally among groups of departments or externally with its suppliers, customers and subsidiaries. The documents likely to be used in EDI are invoices, purchase orders, shipping requests acknowledgements and payments. In EDI information is passed electronically from one computer network without having to be read, retyped or printed. The information has a defined structure agreed between your company and the company you send and receive data from.

1.1.5.2 In Home Applications

One of the biggest uses of computers at home is to access the internet. Some of the popular uses of the Internet for home users are as follows:

1.1.5.2.1 Access to Remote Information.

- One of the uses of computer networks at home is to have access to remote information. Access to remote information comes in many forms. It can be surfing the World Wide Web for information or just for fun. Information available includes the arts, business, cooking, government, health, history, hobbies, recreation, science, sports, travel and many others.
- Many newspapers have gone on-line and can be personalized. For example, it is sometimes possible to tell a newspaper that you want everything about corrupt politicians, big fires, scandals involving celebrities and epidemics, but no football, thank you. Sometimes it is even possible to have the selected articles downloaded to your hard disk while you sleep or printed on your printer just before breakfast.
- The next step beyond newspapers (plus magazines and scientific journals) is the on-line digital library. Many professional organizations, such as the ACM (www.acm.org) and the IEEE Computer Society (www.computer.org), already have many journals and conference proceedings on-line. Other groups are following rapidly. Depending on the cost, size and weight of book-sized notebook computers, printed books may become obsolete.

These applications involve interactions between a person and a remote database full of information

1.1.5.2.2 Person-to-Person Communication.

Another broad category of network use is person-to-person communication.

- E-mail is already used on a daily basis by millions of people all over the world and its use is growing rapidly. It already routinely contains audio and video as well as text and pictures.
- Instant messaging is the facility that allows two people to type messages at each other in real time. A multi-person version of this idea is the chat room, in which a group of people can type messages for all to see.
- Worldwide newsgroups, with discussions on every conceivable topic, are already commonplace among a select group of people and this phenomenon will grow to include the population at large. These discussions, in which one person posts a message and all the other subscribers to the newsgroup can read it, run the gamut from humorous to impassioned. Unlike chat rooms, news groups are not real time and messages are saved so that when someone comes back from vacation, all messages that have been posted in the meanwhile are patiently waiting for reading.
- Another type of person-to-person communication often goes by the name of peer-to-peer communication, to distinguish it from the client-server model. In this form, individuals who form a loose group can communicate with others in the group. Every person can, in principle, communicate with one or more other people; there is no fixed division into clients and servers.
- Fans sharing public domain music or sample tracks that new bands have released for publicity purposes, families sharing photos, movies, and genealogical information and teenagers playing multi-person on-line games etc. are legal applications for peer-to-peer communication. This form of communication is expected to grow considerably in the future.
- Other communication-oriented applications include using the Internet to carry telephone calls, video phone and Internet radio, three rapidly growing areas.
- Another application is tele-learning, meaning attending classes without the inconvenience of having to get there. In the long run, the use of networks to enhance human-to-human communication may prove more important than any of the others.

1.1.5.2.3 Interactive Entertainment.

- The interactive entertainment is a huge and growing industry. The major application here is video on demand. A decade or so hence, it may be possible to select any movie or television program ever made, in any country and have it displayed on your screen instantly. New films may become interactive, where the user is occasionally prompted for the story direction with alternative scenarios provided for all cases. Live

television may also become interactive, with the audience participating in quiz shows, choosing among contestants and so on.

- Already we have multi-person real-time simulation games, like hide-and-seek in a virtual dungeon and flight simulators with the players on one team trying to shoot down the players on the opposing team. If games are played with goggles and three-dimensional real-time, photographic quality moving images, we have a kind of worldwide shared virtual reality.

1.1.5.2.4 Electronic Commerce.

Another category of network usage is electronic commerce for home applications.

- Home shopping is already popular and enables users to inspect the on-line catalogs of thousands of companies. Some of these catalogs will soon provide the ability to get an instant video on any product by just clicking on the product's name. After the customer buys a product electronically but cannot figure out how to use it, on-line technical support may be consulted.
- Another area in which e-commerce is already happening is access to financial institutions. Many people already pay their bills, manage their bank accounts and handle their investments electronically. This will surely grow as networks become more secure.
- On-line auctions of second-hand goods have become a massive industry. Unlike traditional e-commerce, which follows the client-server model, on-line auctions are more of a peer-to-peer system, sort of consumer-to-consumer. The most popular forms of e-commerce are listed in figure 3.

Symbol	Type of E-commerce	Example
B2C	Business-to-consumer	Ordering books on-line
B2B	Business-to-business	Car manufacturer ordering tires from supplier
G2C	Government -to-consumer	Government distributing tax forms electronically
C2C	Consumer-to-consumer	Auctioning second-hand products on line
P2P	Peer-to-peer	File sharing

Figure3. Some Forms of E-Commerce

1.1.5.3 Social Issues

The widespread introduction of networking has introduced new social, ethical,

political problems:

- Should subjects be regulated if newsgroups are set up on topics, such as politics, religion, or sex?
- Should network operators be responsible for the contents of what they carry, just as newspapers and magazines are?
- Should employers have the right to read and possibly censor employee messages?

Along with the good comes the bad. Life seems to be like that. The Internet makes it possible to find information quickly, but a lot of it is ill-informed, misleading, or downright wrong. The medical advice you plucked from the Internet may have come from a Nobel Prize winner or from a high school dropout. Computer networks have also introduced new kinds of antisocial and criminal behavior. Electronic junk mail (spam) has become a part of life because people have collected millions of e-mail addresses and sell them on CD-ROMs to would-be marketeers. E-mail messages containing active content (basically programs or macros that execute on the receiver's machine) can contain viruses that wreak havoc.

1.1.6 Network Hardware

There is no generally accepted classification into which all computer networks fit, but two dimensions stand out as important: transmission technology and scale.

1.1.6.1 Transmission Technology

Broadly speaking, there are two types of transmission technology that are widespread use. They are as follows:

- Broadcast links.
- Point-to-point links.

1.1.6.1.1 Broadcast Networks

These have a single communication channel that is shared by all the machines on the network. Short messages called packets in certain contexts, sent by any machine are received by all the others. An address field within the packet specifies the intended recipient. Upon receiving a packet, a machine checks the address field. If the packet is intended for the receiving machine, that machine processes the packet; if the packet is intended for some other machine, it is just ignored.

As an analogy, consider someone standing at the end of a corridor with many rooms off it and shouting "Rajesh, come here. I want you" Although the packet may actually be received (heard) by many people, only Rajesh responds. The others just ignore it. Another analogy is an airport announcement asking all flight 644 passengers to report to gate 12 for immediate boarding.

Broadcast systems generally also allow the possibility of addressing a packet to all destinations by using a special code in the address field. When a packet with this code is transmitted, it is received and processed by every machine on the network. This mode of operation is called broadcasting. Some broadcast systems also support

transmission to a subset of the machines, something known as multicasting. One possible scheme is to reserve one bit to indicate multicasting. The remaining n-1 address bits can hold a group number. Each machine can "subscribe" to any or all of the groups. When a packet is sent to a certain group, it is delivered to all machines subscribing to that group

1.1.6.1.2 Point-to-Point Networks

In contrast, point-to-point networks consist of many connections between individual pairs of machines. To go from the source to the destination, a packet on this type of network may have to first visit one or more intermediate machines. Often multiple routes, of different lengths, are possible, so finding good ones is important in point-to-point networks. As a general rule (although there are many exceptions), smaller, geographically localized networks tend to use broadcasting, whereas larger networks usually are point-to-point. Point-to-point transmission with one sender and one receiver is sometimes called unicasting.

1.1.6.2 Scale

An alternative criterion for classifying networks is their scale. In this we classify multiple processor systems by their physical size. At the top are the personal area networks, networks that are meant for one person. For example, a wireless network connecting a computer with its mouse, keyboard and printer is a personal area network. Also, a PDA that controls the user's hearing aid or pacemaker fits in this category. Beyond the personal area networks come longer-range networks. These can be divided into local, metropolitan and wide area networks. Finally, the connection of two or more networks is called an inter-network. Internet is a well-known example of an inter-network. Distance is important as a classification metric because different techniques are used at different scales

Here is given brief introduction to network hardware.

1.1.6.3 Local Area Networks

The Local Area Network (LAN) is by far the most common type of data network. As the name suggests, a LAN serves a local area (typically the area of a floor of a building, but in some cases spanning a distance of several kilometers). Typical installations are in industrial plants, office buildings, college or university campuses, or similar locations.

Three distinguishable characteristics for LANs:

- **Size:** usually a diameter of not more than a few kilometers, with bounded and known worst-case transmission time, making special design and simple management possible.
- **Transmission technology:** usually a shared cable running at speeds of 10 to 100 Mbps (and even higher), with delay of tens of microseconds and few errors.
- **Topology:** bus (e.g., Ethernet), ring (e.g., IBM token ring), etc.

1.1.6.4 Metropolitan Area Networks

A Metropolitan Area Network (MAN) is one of a number of types of networks. MAN is a bigger version of a LAN and uses similar technology. It uses one or two cables but does not contain switching elements. It covers an entire city and may be related to the local cable TV network.

Three distinguishable characteristics for MANs:

- The network size falls intermediate between LANs and WANs. A MAN typically covers an area of between 5 and 50 km diameter. Many MANs cover an area the size of a city, although in some cases MANs may be as small as a group of buildings or as large as the North of Punjab.
- A MAN (like a WAN) is not generally owned by a single organization. The MAN, its communications links and equipment are generally owned by either a consortium of users or by a single network provider who sells the service to the users. This level of service provided to each user must therefore be negotiated with the MAN operator and some performance guarantees are normally specified.
- A MAN often acts as a high speed network to allow sharing of regional resources (similar to a large LAN). It is also frequently used to provide a shared connection to other networks using a link to a WAN.

1.1.6.5 Wide Area Networks

A WAN spans a large area, often a country or continent. A WAN consists of two parts:

- **Application part:** Machines for running user programs are called hosts.
- **Communication part:** The hosts are connected by the communication subnet or just subnet, whose job is to carry messages from host to host.

The subnet consists of two components:

- Transmission lines (circuits, channels, or trunks) move bits between machines.
- Switching elements (routers) are specialized computers used to connect two or more transmission lines.

Main characteristics:

- A WAN contains numerous cables or telephone lines, each one connecting a pair of routers.
- For those without direct connection, communication takes place indirectly via other routers.
- When a message (a packet) is sent from one router to another, it is received at each intermediate router in its entirety, stored there until the required output line is free and then forwarded.
- A subnet using this principle is called point-to-point, store-and-forward or packet-switched subnet. WANs may also use broadcast channels, such as satellites or ground radio systems.

1.1.6.6 Wireless Networks

Wireless networks can be divided into three main categories:

- System interconnection.
- Wireless LANs.
- Wireless WANs.

Mobile computers, such notebook computers and personal digital assistants (PDAs), are the fastest-growing segment of the computer industry.

Applications using wireless networks:

- Portable offices which allow people to send and receive phone calls, faxes and emails, to read remote files or login remote machines, etc, and to do this from land, sea or air.
- Of great value to fleets of trucks, taxis and repair-persons for keeping in contact with home.
- Important to rescue workers at disaster sites and to the military.

1.1.6.7. Inter-networks

An internetwork is a collection of individual networks, connected by intermediate networking devices, that functions as a single large network. Internetworking refers to the industry, products and procedures that meet the challenge of creating and administering internetworks. The Internet refers to a specific worldwide internet that is widely used to connect universities, government offices, companies and private individuals.

1.1.7 Network Software

The first computer networks were designed with the hardware as the main concern and the software as an afterthought. This strategy no longer works. Network software is now highly structured.

1.1.7.1. Protocol Hierarchies

To reduce their design complexity, most networks are organized as a series of layers or levels. Each layer offers certain services to the higher layers, shielding those layers from the details of how the offered services are actually implemented. Layer on one machine carries on a conversation with layer on another machine. The rules and conventions used in this conversation are collectively known as the layer protocol. The entities comprising the corresponding layers on different machines are called peers, which communicate using the protocol. Between each pair of adjacent layers there is an interface, which defines primitive operations and services the lower layer offers to the upper one. The set of layers and protocols is called the network architecture, which must contain enough information to allow a software/hardware implementation which correctly obeys the appropriate protocol. The lower layers of a protocol hierarchy are frequently implemented (in whole or in part) in hardware or firmware.

1.1.7.2 Design Issues for the Layers

- **A mechanism for identifying senders and receivers:** some form of addressing for both machines and processes.

- **Directions for data transfer:** simplex, half-duplex, full-duplex communication.
- **Logical channels:** at least two per connection.
- **Error control:** both sides must use the same error-detecting and error-correcting codes. Besides, some way is needed to tell which messages have been correctly received and which have not.
- **Message sequencing or ordering:** message pieces are numbered, but what should be done with pieces out of order?
- **Flow control:** keep a fast sender from swamping a slow receiver with data. Some kind of feedback from receiver is needed.
- Mechanisms for disassembling, transmitting and reassembling messages. A related issue is what to do when processes insist upon transmitting data in very small units.
- **Multiplexing:** using the same connection for multiple, unrelated conversations. E.g., a few physical circuits are used for all virtual connections.
- **Routing:** which path should be chosen? The decision may be split over several layers.

1.1.7.3 Interfaces and Services

The real function of each layer is to provide services to the layer above it.

Terminologies

Entities: active elements (e.g., processes, I/O chips) in each layer.

Peer entities: entities in the same layer on different machines.

Service provider: the layer providing certain services.

Service user: the layer using certain services.

One layer can be a service provider and a service user and may provide multiple services.

SAP (service access points): the places where services can be accessed. Each SAP has a unique address.

Analogies: Sockets and phone numbers in the telephone system. Sockets and socket numbers in BSD UNIX.

1.1.7.4 Connection-Oriented and Connectionless Services

Connection-oriented service is modeled after the telephone system. The essence of a connection is that it acts like a tube: the sender pushes objects in at one end and the receiver takes them out in the same order at the other end. Connection-oriented services are suitable for communicating for a long time between two parties.

Connectionless Service is modeled after the postal system. Each message carries the full address and is routed independently. The order is not guaranteed. Connectionless services are suitable for sending short messages.

Quality of service:

Reliable services guarantee they never lose data. This can be implemented by acknowledgements. The overhead introduced are often worth it, but sometimes

undesirable (unreliable services).

A reliable connection-oriented service is appropriate for file transfer.

An unreliable connection-oriented service is appropriate for digitized voice traffic.

A reliable connectionless service (acknowledged datagram service) is appropriate for registered mails.

An unreliable connectionless service (datagram service) is appropriate for electronic junk mail (with a high probability of arrival, but no guarantee).

Another connectionless service is the request-reply, commonly used to implement the client-server model.

1.1.7.5 Service Primitives

A service is formally specified by a set of primitives available to a service user to interact with the service provider. These primitives tell the service provider to perform some action or report on an action taken by a peer entity.

Primitives can have parameters. In request and response primitives, two parties may negotiate some conditions (e.g., message size), which are part of the protocol.

A confirmed service involves an indication, a request, a response and a confirm. An uncomfirmed service involves just a request and an indication.

An example using eight service primitives:

- request a connection to be established.
- Signal the called party.
- Used by the callee to accept/reject calls.
- Tell the caller whether the call was accepted.
- Request that data be sent.
- Signal the arrival of data.
- Request that a connection be released.
- Signal the peer about the request.

In the above example, which services are confirmed and which ones are unconfirmed?

1.1.7.6 The Relationship of Services to Protocols

Services and protocols are distinct concepts.

A service is a set of primitives (operations) that a layer provides to the layer above it. The service defines what operations (not the format and meaning of the exchanged data) the layer is prepared to perform on behalf of its users.

A protocol is a set of rules governing the format and meaning of the frames, packets, or messages that are exchanged by the peer entities within the same layer. Entities use protocols in order to implement their service definitions. The change of protocols may not be visible to the service users (the service remain the same).

Well known networks include the Internet, ATM networks, Ethernet and the IEEE 802.11 wireless LAN.

1.1.8 Summary

A network is a set of hardware devices connected together, either physically or logically to allow them to exchange information. A computer network consists of nodes and communication links which implement its protocols. It interconnects a set of hosts which conform to the network protocols. At a high level, networks are advantageous because they allow computers and people to be connected together, so they can share resources. Some of the specific benefits of networking include communication, data sharing, Internet access, data security and management, application performance enhancement and entertainment. Networking has a few drawbacks that balance against its many positive aspects. Setting up a network has costs in hardware, software, maintenance and administration. It is also necessary to manage a network to keep it running smoothly and to address possible misuse or abuse. Data security also becomes a much bigger concern when computers are connected together. Of course today, nearly everyone decides that networking is worthwhile after weighing costs against benefits.

The use of computer networks in traditional areas as well new advance fields can be very well perceived. In the business context computer networks can be a real asset. Sharing physical resources, sharing information, geographical independence, providing a powerful communication medium among employees, doing business electronically with other companies, especially suppliers and customers are some of the area of network usage in business applications. The computer networks deliver many services to private individuals at home. Home reservations for airplanes, trains, hotels, restaurants, theaters and so on, anywhere in the world with instant confirmation, Home banking and shopping, On-line and personalized electronic newspapers, journals and libraries, Access to WWW (World Wide Web) which contains information about many topics, are few to name.

Networks can be divided up into LANs, MANs, WANs, Wireless Networks and internetworks, with their own characteristics, technologies speeds and niches. Network Software consists of protocols, which are rules by which processes communicate. Protocols are either connectionless or connection-oriented. Well known networks include the Internet, ATM networks, Ethernet and the IEEE 802.11 wireless LAN.

1.1.9 Self Check Exercise

1 Define computer network. Describe the general characteristics of a computer network.

2 Provide three arguments in favor of the use of a computer network in a modern organization and at least one argument against.

3 Discuss the various benefits and costs of using computer networks.

4 What is the rationale behind using computer networks? Give the usage of computer networks in business applications.

5 What are the applications of computer networks in context of home, society?

6 Which are two types of transmission technology that are in widespread use? Discuss.

7 How you can classify networks on basis of scale? Discuss.

8 What includes in network hardware and software? Discuss with examples.

1.1.10 References and Suggested Readings

- Tanenbaum, Andrew.S. 2003, "Computer Networks." Pearson Education Inc.
- Sinha, Pradeep. K. 2003, "Foundations of Computing", BPB Publications.
- Laudon, Kenneth, C. and Traver, carol, Guercio, 2003, "E-commerce", Pearson Education Inc.
- Forouzan, Behrouz. A. 2002, “Data Communications and Networking”, Tata McGraw Hill Ltd.

LESSON NO. 1.2
AUTHOR: GURINDER GOSAL

OSI REFERENCE MODEL

Structure of the Lesson:

Learning Objectives:

The major objectives of this lesson are to:

- Discuss the need of models to explain the roles played by various technologies, and how they interact
- Describe the Open Systems Interconnection (OSI) Reference Model in detail.
- Understand the functions of various layers of OSI model with examples.

1.2.1 Introduction

The International Standards Organization (ISO) has developed a reference model for network design called the **Open Systems Interconnection** (OSI). It proposes seven-

layer architecture for networks. The idea behind the OSI Reference Model is to provide a framework for both designing networking systems and for explaining how they work. The existence of the model makes it easier for networks to be analyzed, designed, built and rearranged by allowing them to be considered as modular pieces that interact in predictable ways rather than enormous complex monoliths.

It should be pointed out that the OSI model is not the only model in use. It is, however, the most-widely respected model and has become a standard benchmark for comparing other network architectures against it. Understanding OSI is very helpful in making sense of networking protocols and technologies. The model is theoretical, but its concepts are employed regularly to describe the operation of real-world networks.

1.2.2 The Open Systems Interconnection (OSI) Reference Model

In the discussion that follows we describe the OSI Reference Model in detail. We begin with a history of the model and a discussion of some general concepts related to the OSI model. We then describe each of the seven layers of the OSI Reference Model and conclude with a summary of the layers and their respective functions.

1.2.2.1 History of the OSI Reference Model

The idea behind the creation of networking standards is to define widely-accepted ways of setting up networks and connecting them together. The OSI Reference Model represented an early attempt to get all of the various hardware and software manufacturers to agree on a framework for developing various networking technologies.

One interesting aspect of the history of the OSI Reference Model is that the original objective was ***not*** to create a model primarily for educational purposes—even though many people today think that this was the case. The OSI Reference Model was intended to serve as the foundation for the establishment of a widely-adopted suite of protocols that would be used by international internetworks—basically, what the Internet became. This was called unsurprisingly the OSI Protocol Suite.

In the late 1970s, two projects began independently, with the same goal: to define a unifying standard for the architecture of networking systems. One was administered by the International Organization for Standardization (ISO), while the other was undertaken by the International Telegraph and Telephone Consultative Committee, or CCITT. These two international standards bodies each developed a document that defined similar networking models. In 1983, these two documents were merged together to form a standard called The Basic Reference Model for Open Systems Interconnection. That's a mouthful, so the standard is usually referred to as the Open Systems Interconnection Reference Model, the OSI Reference Model or even just the OSI Model. It was published in 1984 by both the ISO, as standard ISO 7498.

1.2.2.2 Importance of OSI Model

The OSI Reference Model provides the basis for understanding how technologies like Ethernet and HomePNA have some important similarities; it explains how a PC can

communicate using any of several different sets of protocols, even simultaneously; it is an important part of understanding the differences between interconnection devices such as repeaters, hubs, bridges, switches and routers; and it also explains how many WAN technologies interoperate. Far from being obsolete, the OSI model layers are now showing up more than ever in discussions of technology.

It is just as much a mistake to assign too much importance to the OSI Reference Model as too little. While the model defines a framework for understanding networks, not all networking components, protocols and technologies will necessarily fall into the model's strict layering architecture. There are cases where trying to use the model to describe certain concepts can lead to less clarity rather than more. One should remember that the OSI model is a *tool* and should be used accordingly.

The International Standards Organization (ISO) has developed a reference model for network design called the **Open Systems Interconnection** (OSI). It proposes seven-layer architecture for networks, as summarized by Figure 1. Each layer is characterized by a set of standard protocols which specify its behavior.

1.2.3 OSI Reference Model Layers

Now it is time to take a look at the actual individual layers of the OSI Reference Model. Each layer has certain characteristics that define it, and also various protocols normally associated with it. Understanding the gradation of each layer will help you understand all the technologies that use them.

For each layer we provide its name and layer number, describe its general function in the OSI layer stack and outline the specific types of activities for which each is normally responsible.

1.2.3.1 Physical Layer (Layer 1)

The lowest layer of the OSI Reference Model is layer 1, the *physical layer*; it is commonly abbreviated "PHY". The physical layer is specialy compared to the other layers of the model, because it is the only one where data is physically moved across the network interface. All of the other layers perform useful functions to create messages to be sent, but they must all be transmitted down the protocol stack to the physical layer, where they are actually sent out over the network.

1.2.3.1.1 Understanding the Role of the Physical Layer

The name "physical layer" can be a bit tricky. Because of that name many people who study networking get the impression that the physical layer is only about actual network hardware. Some people may say the physical layer is "the network interface cards and cables". This is not actually the case, however. The physical layer defines a number of network functions, not just hardware cables and cards.

A related notion is that "all network hardware belongs to the physical layer". Again, this isn't strictly accurate. All hardware must have ***some*** relation to the physical layer in order to send data over the network but hardware devices generally implement multiple layers of the OSI model, including the physical layer but also others. For

example, an Ethernet network interface card performs functions at both the physical layer and the data link layer.

1.2.3.1.2 Functions of Physical Layer

The following are the main responsibilities of the physical layer in the OSI Reference Model:

- **Definition of Hardware Specifications:** The details of operation of cables, connectors, wireless radio transceivers, network interface cards and other hardware devices are generally a function of the physical layer (although also partially the data link layer).
- **Encoding and Signaling:** The physical layer is responsible for various encoding and signaling functions that transform the data from bits that reside within a computer or other device into signals that can be sent over the network.
- **Data Transmission and Reception:** After encoding the data appropriately, the physical layer actually transmits the data and of course, receives it. Note that this applies equally to wired and wireless networks, even if there is no tangible cable in a wireless network.
- **Topology and Physical Network Design:** The physical layer is also considered the domain of many hardware-related network design issues, such as LAN and WAN topology.

In general, then, physical layer technologies are ones that are at the very lowest level and deal with the actual ones and zeroes that are sent over the network. For example, when considering network interconnection devices, the simplest ones operate at the physical layer: repeaters, conventional hubs and transceivers. These devices have absolutely no knowledge of the contents of a message. They just take input bits and send them as output. Devices like switches and routers operate at higher layers and look at the data they receive as being more than voltage or light pulses that represent one or zero.

1.2.3.1.3 Relationship between the Physical Layer and Data Link Layer

It's important to point out that while the physical layer of a network technology primarily defines the hardware it uses, the physical layer is closely related to the data link layer. Thus, it is not generally possible to define hardware at the physical layer "independently" of the technology being used at the data link layer. For example, Ethernet is a technology that describes specific types of cables and network hardware, but the physical layer of Ethernet can only be isolated from its data link layer aspects to a point. While Ethernet cables are "physical layer", for example, their maximum length is related closely to message format rules that exist at the data link layer.

Furthermore, some technologies perform functions at the physical layer that are normally more closely associated with the data link layer. For example, it is common to have the physical layer perform low-level (bit level) repackaging of data link layer

frames for transmission. Error detection and correction may also be done at layer 1 in some cases. Most people would consider these "layer two functions".

In many technologies, a number of physical layers can be used with a data link layer. Again here, the classic example is Ethernet, where dozens of different physical layer implementations exist, each of which uses the same data link layer (possibly with slight variations.)

1.2.3.1.4 Physical Layer Sublayers

Finally, many technologies further subdivide the physical layer into *sublayers*. In order to increase performance, physical layer encoding and transmission methods have become more complex over time. The physical layer may be broken into layers to allow different network media to be supported by the same technology, while sharing other functions at the physical layer that are common between the various media. A good example of this is the physical layer architecture used for Fast Ethernet, Gigabit Ethernet and 10-Gigabit Ethernet.

1.2.4 Data Link Layer (Layer 2)

The second-lowest layer (layer 2) in the OSI Reference Model stack is the data link layer, often abbreviated "DLL" (abbreviation has other meanings as well in the computer world). The data link layer, also sometimes just called the link layer, is where many wired and wireless local area networking (LAN) technologies primarily function. For example, Ethernet, Token Ring, FDDI and 802.11 ("wireless Ethernet" or "Wi-Fi') are all sometimes called "data link layer technologies". The set of devices connected at the data link layer is what is commonly considered a simple "network", as opposed to an internetwork.

1.2.4.1 Data Link Layer Sublayers: Logical Link Control (LLC) and Media Access Control (MAC)

The data link layer is often conceptually divided into two sublayers:

- Logical link control (LLC) and
- Media access control (MAC).

This split is based on the architecture used in the IEEE 802 Project, which is the IEEE working group responsible for creating the standards that define many networking technologies (including all of the ones I mentioned above except FDDI). By separating LLC and MAC functions, interoperability of different network technologies is made easier, as explained in our earlier discussion of networking model concepts.

1.2.4.2 Data Link Layer Functions

The following are the key tasks performed at the data link layer:

- **Logical Link Control (LLC):** Logical link control refers to the functions required for the establishment and control of logical links between local devices on a network. As mentioned above, this is usually considered a DLL sublayer; it provides services to the network layer above it and hides the rest of the details of the data link layer to allow different technologies to

work seamlessly with the higher layers. Most local area networking technologies use the IEEE 802.2 LLC protocol.

- **Media Access Control (MAC):** This refers to the procedures used by devices to control access to the network medium. Since many networks use a shared medium (such as a single network cable, or a series of cables that are electrically connected into a single virtual medium) it is necessary to have rules for managing the medium to avoid conflicts. For example. Ethernet uses the CSMA/CD method of media access control, while Token Ring uses token passing.
- **Data Framing:** The data link layer is responsible for the final encapsulation of higher-level messages into frames that are sent over the network at the physical layer.
- **Addressing:** The data link layer is the lowest layer in the OSI model that is concerned with addressing: labeling information with a particular destination location. Each device on a network has a unique number, usually called a hardware address or MAC address that is used by the data link layer protocol to ensure that data intended for a specific machine gets to it properly.
- **Error Detection and Handling:** The data link layer handles errors that occur at the lower levels of the network stack. For example, a cyclic redundancy check (CRC) field is often employed to allow the station receiving data to detect if it was received correctly.

1.2.5 Network Layer (Layer 3)

The third-lowest layer of the OSI Reference Model is the *network layer*. If the data link layer is the one that basically defines the boundaries of what is considered a network, the network layer is the one that defines how *internetworks* (interconnected networks) function. The network layer is the lowest one in the OSI model that is concerned with actually getting data from one computer to another even if it is on a remote network; in contrast, the data link layer only deals with devices that are local to each other.

While all of layers 2 through 6 in the OSI Reference Model serve to act as "fences" between the layers below them and the layers above them, the network layer is particularly important in this regard. It is at this layer that the transition really begins from the more abstract functions of the higher layers—which don't concern themselves as much with data delivery—into the specific tasks required to get data to its destination. The transport layer, which is related to the network layer in a number of ways, continues this "abstraction transition" as you go up the OSI protocol stack.

1.2.5.1 Network Layer Functions

Some of the specific jobs normally performed by the network layer include:

- **Logical Addressing:** Every device that communicates over a network has associated with it a logical address, sometimes called a *layer three* address.

For example, on the Internet, the Internet Protocol (IP) is the network layer protocol and every machine has an IP address. Note that addressing is done at the data link layer as well, but those addresses refer to local physical devices. In contrast, logical addresses are independent of particular hardware and must be unique across an entire internetwork.

- **Routing:** Moving data across a series of interconnected networks is probably the defining function of the network layer. It is the job of the devices and software routines that function at the network layer to handle incoming packets from various sources, determine their final destination and then figure out where they need to be sent to get them where they are supposed to go.
- **Datagram Encapsulation:** The network layer normally encapsulates messages received from higher layers by placing them into *datagrams* (also called *packets*) with a network layer header.
- **Fragmentation and Reassembly:** The network layer must send messages down to the data link layer for transmission. Some data link layer technologies have limits on the length of any message that can be sent. If the packet that the network layer wants to send is too large, the network layer must split the packet up, send each piece to the data link layer, and then have pieces reassembled once they arrive at the network layer on the destination machine. A good example is how this is done by the Internet Protocol.
- **Error Handling and Diagnostics:** Special protocols are used at the network layer to allow devices that are logically connected, or that are trying to route traffic, to exchange information about the status of hosts on the network or the devices themselves

1.2.6 Transport Layer (Layer 4)

The fourth and "middle" layer of the OSI Reference Model protocol stack is the transport layer. It is more often associated with the lower layers, because it concerns itself with the **transport** of data, but its functions are also somewhat high-level, resulting in the layer having a fair bit in common with layers 5 through 7 as well.

Recall that layers 1, 2 and 3 are concerned with the actual packaging, addressing, routing and delivery of data; the physical layer handles the bits; the data link layer deals with local networks and the network layer handles routing between networks. The transport layer, in contrast, is sufficiently conceptual that it no longer concerns itself with these "nuts and bolts" matters. It relies on the lower layers to handle the process of moving data between devices.

The transport layer really acts as a "liaison" of sorts between the abstract world of applications at the higher layers and the concrete functions of layers one to three. Due to this role, the transport layer's overall job is to provide the necessary functions to enable communication between software application processes on different computers. This encompasses a number of different but related duties

Modern computers are multitasking and at any given time may have many different software applications all trying to send and receive data. The transport layer is charged with providing a means by which these applications can all send and receive data using the same lower-layer protocol implementation. Thus, the transport layer is sometimes said to be responsible for end-to-end or host-to-host transport (in fact, the equivalent layer in the TCP/IP model is called the "host-to-host transport layer").

1.2.6.1 Transport Layer Functions

Let's look at the specific functions often performed at the transport layer in more detail:

- **Process-Level Addressing:** Addressing at this layer deals with hardware devices on a local network and layer three addressing identifies devices on a logical internetwork. Addressing is also performed at the transport layer, where it is used to differentiate between software programs. This is part of what enables many different software programs to use a network layer protocol simultaneously, as mentioned above. The best example of transport-layer process-level addressing is the TCP and UDP port mechanism used in TCP/IP, which allows applications to be individually referenced on any TCP/IP device.
- **Multiplexing and Demultiplexing:** Using the addresses just mentioned, transport layer protocols on a sending device multiplex the data received from many application programs for transport, combining them into a single stream of data to be sent. The same protocols receive data and then demultiplex it from the incoming stream of datagrams, and direct each package of data to the appropriate recipient application processes.
- **Segmentation, Packaging and Reassembly:** The transport layer segments the large amounts of data it sends over the network into smaller pieces on the source machine and then reassembles them on the destination machine. This function is similar conceptually to the fragmentation function of the network layer; just as the network layer fragments messages to fit the limits of the data link layer, the transport layer segments messages to suit the requirements of the underlying network layer.
- **Connection Establishment, Management and Termination:** Transport layer connection-oriented protocols are responsible for the series of communications required to establish a connection, maintain it as data is sent over it and then terminate the connection when it is no longer required.
- **Acknowledgments and Retransmissions:** As mentioned above, the transport layer is where many protocols are implemented that guarantee reliable delivery of data. This is done using a variety of techniques, most commonly the combination of acknowledgments and retransmission timers. Each time data is sent a timer is started; if it is received, the recipient sends back an acknowledgment to the transmitter to indicate successful

transmission. If no acknowledgment comes back before the timer expires, the data is retransmitted. Other algorithms and techniques are usually required to support this basic process.

- **Flow Control:** Transport layer protocols that offer reliable delivery also often implement flow control features. These features allow one device in a communication to specify to another that it must "throttle back" the rate at which it is sending data, to avoid bogging down the receiver with data. These allow mismatches in speed between sender and receiver to be detected and dealt with.

1.2.6.2 Relationship between the Transport Layer and Network Layer

In theory, the transport layer and network layer are distinct, but in practice, they are often very closely related to each other. You can see this easily just by looking at the names of common protocol stacks—they are often named after the layer three and four protocols in the suite, implying their close relationship. For example, the name "TCP/IP" comes from the suite's most commonly used transport layer protocol (TCP) and network layer protocol (IP). Similarly, the Novell NetWare suite is often called "IPX/SPX" for its layer three (IPX) and layer four (SPX) protocols. Typically, specific transport layer protocols use the network layers in the same family. You won't often find a network using the transport layer protocol from one suite and the network layer protocol from another.

The most commonly used transport layer protocols are the Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) in the TCP/IP suite, the Sequenced Packet Exchange (SPX) protocol in the NetWare protocol suite.

1.2.7 Session Layer (Layer 5)

The fifth layer in the OSI Reference Model is the session layer. As we proceed up the OSI layer stack from the bottom, the session layer is the first one where pretty much all practical matters related to the addressing, packaging and delivery of data are left behind—they are functions of layers four and below. It is the lowest of the three upper layers, which collectively are concerned mainly with software application issues and not with the details of network and internet implementation.

The name of this layer tells you much about what it is designed to do: to allow devices to establish and manage sessions. In general terms, a session is a persistent logical linking of two software application processes to allow them to exchange data over a prolonged period of time. In some discussions, these sessions are called dialogs; they are roughly analogous to a telephone call made between two people.

2.7.1 Session Layer Functions

As the boundaries between layers start to get very fuzzy once you get to the session layer, which makes it hard to categorize what exactly belongs at layer 5. Some technologies really span layers 5 through 7, and especially in the world of TCP/IP, it is not common to identify protocols that are specific to the OSI session layer.

The term "session" is somewhat vague, and this means that there is sometimes

disagreement on the specific functions that belong at the session layer or even whether certain protocols belong at the session layer or not. To add to this potential confusion, there is the matter of differentiating between a "connection" and a "session". Connections are normally the province of layer four and layer three, yet a Transmission Control Protocol (TCP) connection, for example, can persist for a long time. The longevity of TCP connections makes them hard to distinguish from "sessions".

Application Program Interfaces (APIs)

The primary job of session layer protocols is to provide the means necessary to set up, manage and end sessions. In fact, in some ways, session layer software products are more sets of tools than specific protocols. These session-layer tools are normally provided to higher layer protocols through command sets often called application program interfaces or APIs.

Common APIs include NetBIOS, TCP/IP Sockets and Remote Procedure Calls (RPCs). They allow an application to accomplish certain high-level communications over the network easily by using a standardized set of services. Most of these session-layer tools are of primary interest to the developers of application software. The programmers use the APIs to write software that is able to communicate using TCP/IP without having to know the implementation details of how TCP/IP works.

For example, the Sockets interface lies conceptually at layer five and is used by TCP/IP application programmers to create sessions between software programs over the Internet on the UNIX operating system. Windows Sockets similarly lets programmers create Windows software that is Internet-capable and able to interact easily with other software that uses that interface.

1.2.8 Presentation Layer (Layer 6)

The presentation layer is the sixth layer of the OSI Reference Model protocol stack and second from the top. It is different from the other layers in two key respects.

First, it has a much more limited and specific function than the other layers; it's actually somewhat easy to describe. Second, it is used much less often than the other layers; in many types of connections it is not required.

The name of this layer suggests its main function as well it deals with the presentation of data. More specifically, the presentation layer is charged with taking care of any issues that might arise where data sent from one system needs to be viewed in a different way by the other system. It also takes care of any special processing that must be done to data from the time an application tries to send it until the time it is sent over the network.

1.2.8.1 Presentation Layer Functions

Here are some of the specific types of data handling issues that the presentation layer handles:

- **Translation:** Networks can connect very different types of computers together: PCs, Macintoshes, UNIX systems, AS/400 servers and mainframes can all exist on the same network. These systems have many distinct

characteristics and represent data in different ways; they may use different character sets for example. The presentation layer handles the job of hiding these differences between machines.

- **Compression:** Compression (and decompression) may be done at the presentation layer to improve the throughput of data. (There are some who believe this is not, strictly speaking, a function of the presentation layer.)
- **Encryption:** Some types of encryption (and decryption) are performed at the presentation layer. This ensures the security of the data as it travels down the protocol stack. For example, one of the most popular encryption schemes that are usually associated with the presentation layer is the Secure Sockets Layer (SSL) protocol. Not all encryption is done at layer 6, however; some encryption is often done at lower layers in the protocol stack, in technologies such as IPSec.

1.2.8.2 Presentation Layer Role in the OSI Model

The reason that the presentation layer is not always used in network communications is that the jobs mentioned above are simply not always needed. Compression and encryption are usually considered "optional"and translation features are also only needed in certain circumstances. Another reason why the presentation layer is sometimes not mentioned is that its functions may be performed as part of the application layer.

The fact that the translation job done by the presentation layer isn't always needed means that it is common for it to be "skipped" by actual protocol stack implementations. This means that protocols at layer seven may talk directly with those at layer five. Once again, this is part of the reason why all of the functions of layers five through seven may be included together in the same software package, as described in the overview of layers and layer groupings.

1.2.9 Application Layer (Layer 7)

At the very top of the OSI Reference Model stack of layers, we find layer 7, the application layer. Continuing the trend that we saw in layers 5 and 6, this one too is named very appropriately: the application layer is the one that is used by network applications. These programs are what actually implement the functions performed by users to accomplish various tasks over the network.

It's important to understand that what the OSI model calls an "application" is not exactly the same as what we normally think of as an "application". In the OSI model, the application layer provides services for user applications to employ. For example, when you use your Web browser, that actual software is an application running on your PC. It doesn't really "reside" at the application layer. Rather, it makes use of the services offered by a protocol that operates at the application layer, which is called the Hypertext Transfer Protocol (HTTP). The distinction between the browser and HTTP is subtle, but important.

The reason for pointing this out is because not all user applications use the application layer of the network in the same way. Sure, your Web browser does and so do your e-mail client and your Usenet news reader. But if you use a text editor to open a file on another machine on your network, that editor is not using the application layer. In fact, it has no clue that the file you are using is on the network: it just sees a file addressed with a name that has been mapped to a network somewhere else. The operating system takes care of redirecting what the editor does, over the network.

Similarly, not all uses of the application layer are by applications. The operating system itself can (and does) use services directly at the application layer.

That caveat aside, under normal circumstances, whenever you interact with a program on your computer that is designed specifically for use on a network, you are dealing directly with the application layer. For example, sending an e-mail, firing up a Web browser or using an IRC chat program—all of these involve protocols that reside at the application layer.

There are dozens of different application layer protocols that enable various functions at this layer. Some of the most popular ones include HTTP, FTP, SMTP, DHCP, NFS, Telnet, SNMP, POP3, NNTP and IRC.

As the "top of the stack" layer, the application layer is the only one that does not provide any services to the layer above it in the stack—there isn't one! Instead, it provides services to programs that want to use the network and to you, the user. So the responsibilities at this layer are simply to implement the functions that are needed by users of the network. And of course, to issue the appropriate commands to make use of the services provided by the lower layers.

Table: OSI Reference Model Layer Summary

Group	No	Layer Name	Key Responsibilities	Data Type Handled	Scope
Lower Layers	1	**Physical**	Encoding and Signaling; Physical Data Transmission; Hardware Specifications; Topology and Design	Bits	Electrical or light signals sent between local devices
	2	**Data Link**	Logical Link Control; Media Access Control; Data Framing; Addressing; Error Detection and Handling; Defining Requirements of Physical Layer	Frames	Low-level data messages between local devices
	3	**Network**	Logical Addressing; Routing; Datagram Encapsulation; Fragmentation and Reassembly; Error Handling and Diagnostics	Datagrams / Packets	Messages between local or remote devices
	4	**Transport**	Process-Level Addressing; Multiplexing/ Demultiplexing; Connections; Segmentation and Reassembly; Acknowledgments and Retransmissions; Flow Control	Datagrams / Segments	Communication between software processes
Upper Layers	5	**Session**	Session Establishment, Management and Termination	Sessions	Sessions between local or remote devices
	6	**Presentation**	Data Translation; Compression and Encryption	Encoded User Data	Application data representations
	7	**Application**	User Application Services	User Data	Application data

1.2.10 Summary

Networking models such as the OSI Reference Model provide a framework for breaking down complex internetworks into components that can more easily be understood and utilized. The Open Systems Interconnection Reference Model (OSI Reference Model or OSI Model) was originally created as the basis for designing a universal set of protocols called the OSI Protocol Suite. This suite never achieved widespread success, but the model became a very useful tool for both education and development.

The most fundamental concept in the OSI Reference Model is the division of networking functions into a set of *layers*; from layer one at the bottom to layer seven at the top. As you go up the layer stack, you move away from concrete, hardware-specific functions to ones that are increasingly abstract, until reaching the realm of user applications at layer seven. The seven layers are sometimes divided into groupings: the lower layers (one, two and three) and the upper layers (four through seven). There is some disagreement on whether layer four is a lower or upper layer.

The **OSI model** proposes seven-layer architecture for networks. Each layer is characterized by a set of protocols.

The lowest layer in the OSI Reference Model is the *physical layer*. It is the realm of networking hardware specifications and is the places where technologies reside that perform data encoding, signaling, transmission and reception functions. The physical layer is closely related to the data link layer

The second OSI Reference Model layer is the *data link layer*. This is the place where most LAN and wireless LAN technologies are defined. Layer two is responsible for logical link control; media access control, hardware addressing, error detection and handling, and defining physical layer standards. It is often divided into the logical link control (LLC) and media access control (MAC) sublayers, based on the IEEE 802 Project that uses that architecture.

The OSI Reference Model's third layer is called the *network layer*. This is one of the most important layers in the model; it is responsible for the tasks that link together individual networks into *internetworks*. Network layer functions include internetwork-level addressing, routing, datagram encapsulation, fragmentation and reassembly, and certain types of error handling and diagnostics. The network layer and transport layer are closely related to each other.

The fourth and middle OSI Reference Model layer is the transport layer. This is another very important conceptual layer in the model; it represents the transition point between the lower layers that deal with data delivery issues and the higher layers that work with application software. The transport layer is responsible for enabling end-to-end communication between application processes, which it accomplishes in part through the use of process-level addressing and multiplexing/demultiplexing. Transport layer protocols are responsible for dividing application data into blocks for transmission and may be either connection-oriented or connectionless. Protocols at this layer also

often provide data delivery management services such as reliability and flow control.
The fifth layer in the OSI Reference Model layer is the session layer. As its name suggests, it is the layer intended to provide functions for establishing and managing sessions between software processes. Session layer technologies are often implemented as sets of software tools called application program interfaces (APIs), which provide a consistent set of services that allow programmers to develop networking applications without needing to worry about lower-level details of transport, addressing and delivery.
The sixth OSI model layer is called the presentation layer. Protocols at this layer take care of manipulation tasks that transform data from one representation to another, such as translation, compression and encryption. In many cases, no such functions are required in a particular networking stack; if so, there may not be any protocol active at layer six.
The seventh and highest layer in the OSI Reference Model is the application layer. Application protocols are defined at this layer, which implement specific user applications and other high-level functions. Since they are at the top of the stack, application protocols are the only ones that do not provide services to a higher layer; they make use of services provided by the layers below.

1.2.11 Self-Assessment Questions

1. What is Open Systems Interconnection (OSI) model? What is the history of this model? What is the importance of OSI model?
2. Explain the rationale behind the OSI seven-layer model. Briefly describe the role of each layer and its main functions.
3. Write short notes on the following:
   a) Relationship between the Physical Layer and Data Link Layer
   b) Data Link Layer Sublayers: (LLC) and (MAC)
   c) Relationship between the Transport Layer and Network Layer
   d) Presentation Layer Role in the OSI Model

1.2.12 References and Suggested Readings

- Tanenbaum, Andrew.S. 2003, “Computer Networks.” Pearson Education Inc.
- Sinha, Pradeep. K. 2003, “Foundations of Computing”, BPB Publications
- Laudon, Kenneth, C. and Traver, carol, Guercio, 2003, “E-commerce”, Pearson Education Inc.
- Forouzan, Behrouz. A. 2002, “Data Communications and Networking”, Tata McGraw Hill Ltd.

LESSON NO. 1.3
AUTHOR: GURINDER GOSAL

TCP/IP REFERENCE MODEL

Structure of the Lesson:

Learning Objectives:

The major objectives of this lesson are to:

- Discuss the need of models to explain the roles played by various technologies and how they interact
- Discuss the issues of reference models
- Discuss advantages of networking models
- Describe the TCP/IP Reference Model in detail.
- Understand the functions of various layers of TCP/IP Reference Model with examples.
- Compare TCP/IP model with OSI model on various factors.
- Evaluate TCP/IP Reference Model.

1.3.1 Introduction

Models are useful because they help us understand difficult concepts and complicated systems. When it comes to networking, there are several models that are used to explain the roles played by various technologies and how they interact. The developers of the TCP/IP protocol suite created their own architectural model to help describe its components and functions. This model goes by different names, including the TCP/IP model, the DARPA model (after the agency that was largely responsible for developing TCP/IP) and the DOD model (after the United States Department of Defense, the "D" in "DARPA").

Regardless of the model you use to represent the function of a network—and regardless

of what you call that model -the functions that the model represents are pretty much the same. This means that the TCP/IP and the OSI models are really quite similar in nature even if they don't carve up the network functionality precisely the same way. There is a fairly natural correspondence between the TCP/IP and OSI layers; it just isn't always a "one-to-one" relationship. Since the OSI model is used so widely, it is common to explain the TCP/IP architecture both in terms of the TCP/IP layers and the corresponding OSI layers.

1.3.2 Some Basic Issues of Reference Models

Before getting into the details of how the TCP/IP Reference Model works, it is good to first discuss some of the basic issues related to reference models. We will understand why the model is important and how it benefits networking. Several issues that relate to reference models in general terms and of course, to the TCP/IP Reference Model as well will be discussed. We begin with some overview explanation of why networking models are beneficial and why it is important for us to understand how the TCP/IP model works.

1.3.2.1 The Advantages of Networking Models

One of the important roles of networking models is to make the networking technology simpler to understand. One of the ways in which networking technology is made easier to understand is by splitting it into pieces, each of which plays a particular role or is responsible for a specific job or function. However, if this is to be done, we must have a way of ensuring that these various pieces can interoperate; that is, each must know what is expected of it and also what it can expect from the other pieces. Models split the multitude of tasks required to implement modern networks, into smaller chunks that can be more easily managed.

Networking models represent a framework for dividing up the tasks needed to implement a network, by splitting the work into different levels or *layers*. Hardware and software running at each layer is responsible for interacting with its corresponding hardware and software running on other devices at the same layer. The responsibilities of each hardware or software element are defined in part by specifically delineating lines that exist between the layers.

The result is that you get the benefits of easier training, specialized capabilities at each layer, improved capabilities for modification and modularity. Modularity is particularly important, as it allows you to interchange technologies that run at different layers.

1.3.3 TCP/IP Overview and History

TCP/IP was initially developed in the 1970s as part of an effort to define a set of technologies to operate the fledgling Internet. The name "TCP/IP" came about when the original Transmission Control Program (TCP) was split into the Transmission Control Protocol (TCP) and Internet Protocol (IP). TCP/IP consists of dozens of different protocols, but only a few are the "main" protocols that define the core operation of the suite. Of these key protocols, two are usually considered the most important. The first modern versions of

these two key protocols were documented in 1980 as TCP version 4 and IP version 4.
The *Internet Protocol (IP)* is the primary OSI network layer (layer three) protocol that provides addressing, datagram routing and other functions in an internetwork. The *Transmission Control Protocol (TCP)* is the primary transport layer (layer four) protocol and is responsible for connection establishment and management and reliable data transport between software processes on devices.
Due to the importance of these two protocols, their abbreviations have come to represent the entire suite: "TCP/IP". IP and TCP are important because many of TCP/IP's most critical functions are implemented at layers three and four. However, there is much more to TCP/IP than just TCP and IP. The protocol suite as a whole requires the work of many different protocols and technologies to make a functional network that can properly provide users with the applications they need.

1.3.4 TCP/IP Reference Model

The architecture of the TCP/IP protocol suite is often described in terms of a layered reference model called the *TCP/IP model*, *DARPA model* or *DOD model*. It was first defined in 1974 by Cerf and Kahn. TCP/IP uses its own four-layer architecture that corresponds roughly to the OSI Reference Model and provides a framework for the various protocols that comprise the suite. It also includes numerous high-level applications, some of which are well-known by Internet users who may not realize they are part of TCP/IP, such as HTTP (which runs the World Wide Web) and FTP.

1.3.4.1 Layers of TCP/IP Model

The TCP/IP model uses four layers that logically span the equivalent of the top six layers of the OSI reference model; this is shown in Figure 1.
The following are the TCP/IP model layers, starting from the bottom.

1.3.4.1.1 Network Interface Layer

As its name suggests, this layer represents the place where the actual TCP/IP protocols running at higher layers interface to the local network. This layer is somewhat "controversial" in that some people don't even consider it a "legitimate" part of TCP/IP. This is usually because none of the core IP protocols run at this layer. Despite this, the network interface layer is part of the architecture. It is equivalent to the data link layer (layer two) in the OSI Reference Model and is also sometimes called the *link layer*. You may also see the name *network access layer*.

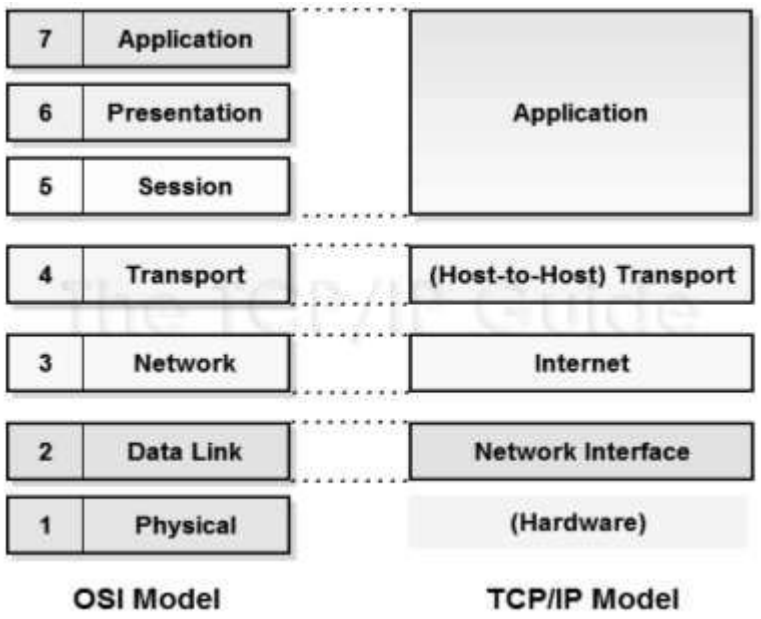


Figure 1: OSI Reference Model and TCP/IP Model Layers

The TCP/IP architectural model has four layers that approximately match six of the seven layers in the OSI Reference Model. The TCP/IP model does not address the physical layer, which is where hardware devices reside. The next three layers network interface, internet and (host-to-host) transport—correspond to layers 2, 3 and 4 of the OSI model. The TCP/IP application layer conceptually "blurs" the top three OSI layers. It's also worth noting that some people consider certain aspects of the OSI session layer to be arguably part of the TCP/IP host-to-host transport layer.

On many TCP/IP networks, there is no TCP/IP protocol running at all on this layer, because it is simply not needed. For example, if you run TCP/IP over an Ethernet, then Ethernet handles layer two (and layer one) functions. However, the TCP/IP standards do define protocols for TCP/IP networks that do not have their own layer two implementation. These protocols, the Serial Line Internet Protocol (SLIP) and the Point-to-Point Protocol (PPP), serve to fill the gap between the network layer and the physical layer. They are commonly used to facilitate TCP/IP over direct serial line connections (such as dial-up telephone networking) and other technologies that operate directly at the physical layer.

1.3.4.1.2 Internet Layer

This layer corresponds to the network layer in the OSI Reference Model (and for that reason is sometimes called the *network layer* even in TCP/IP model discussions). It is responsible for typical layer three jobs, such as logical device addressing, data packaging, manipulation and delivery and last but not least, routing. At this layer we find the Internet Protocol (IP), arguably the heart of TCP/IP, as well as support protocols such as ICMP and

the routing protocols (RIP, OSFP, BGP, etc.) The new version of IP called IP version 6, will be used for the Internet of the future and is of course also at this layer.
This layer is the key player that holds the whole architecture together. Its job is to permit hosts to inject packets into any network and have them travel independently to the destination (potentially on a different network). They may even arrive in a different order than they were sent, in which case it is the job of higher layers to rearrange them, if in-order delivery is desired. Note that "internet" is used here in a generic sense, even though this layer is present in the Internet.
The analogy here is with the (snail) mail system. A person can drop a sequence of international letters into a mail box in one country and with a little luck, most of them will be delivered to the correct address in the destination country. Probably the letters will travel through one or more international mail gateways along the way, but this is transparent to the users. Furthermore, that each country (i.e., each network) has its own stamps, preferred envelope sizes and delivery rules is hidden from the users.
The internet layer defines an official packet format and protocol called IP (Internet Protocol). The job of the internet layer is to deliver IP packets where they are supposed to go. Packet routing is clearly the major issue here as is avoiding congestion.

1.3.4.1.3 (Host-to-Host) Transport Layer

This primary job of this layer is to facilitate end-to-end communication over an internetwork. It is in charge of allowing logical connections to be made between devices to allow data to be sent either unreliably (with no guarantee that it gets there) or reliably (where the protocol keeps track of the data sent and received to make sure it arrives, and re-sends it if necessary). It is also here that identification of the specific source and destination application process is accomplished.
The formal name of this layer is often shortened to just the *transport layer*; the key TCP/IP protocols at this layer are the Transmission Control Protocol (TCP) and User Datagram Protocol (UDP). The TCP/IP transport layer corresponds to the layer of the same name in the OSI model (layer four) but includes certain elements that are arguably part of the OSI session layer. For example, TCP establishes a connection that can persist for a long period of time, which some people say makes a TCP connection more like a session.
The first one, **TCP (Transmission Control Protocol)**, is a reliable connection-oriented protocol that allows a byte stream originating on one machine to be delivered without error on any other machine in the internet. It fragments the incoming byte stream into discrete messages and passes each one on to the internet layer. At the destination, the receiving TCP process reassembles the received messages into the output stream. TCP also handles flow control to make sure a fast sender cannot swamp a slow receiver with more messages than it can handle.
The second protocol in this layer, **UDP (User Datagram Protocol**), is an unreliable, connectionless protocol for applications that do not want TCP's sequencing or flow

control and wish to provide their own. It is also widely used for one-shot, client-server-type request-reply queries and applications in which prompt delivery is more important than accurate delivery, such as transmitting speech or video.

1.3.4.1.4 Application Layer

This is the highest layer in the TCP/IP model. It is a rather broad layer, encompassing layers five through seven in the OSI model. While this seems to represent a loss of detail compared to the OSI model, but in other way the TCP/IP model better reflects the "blurry" nature of the divisions between the functions of the higher layers in the OSI model, which in practical terms often seem rather arbitrary. It really is hard to separate some protocols in terms of which of layers five, six or seven they encompass.

Numerous protocols reside at the application layer. These include application protocols such as HTTP, FTP and SMTP for providing end-user services, as well as administrative protocols like SNMP, DHCP and DNS.

In TCP/IP model, session or presentation layer are not present. Application layer is present on the top of the Transport layer.

The Application Layer defines following protocols:

File Transfer Protocol (FTP)

It was designed to permit reliable transfer of files over different platforms. At the transport layer to ensure reliability, FTP uses TCP. FTP offers simple commands and makes the differences in storage methods across networks transparent to the user. The FTP client is able to interact with any FTP server; therefore the FTP server must also be able to interact with any FTP client. FTP does not offer a user interface, but it does offer an application program interface for file transfer. The client part of the protocol is called as FTP and the server part of the protocol is known as FTPd. The suffix "d" means Daemon this is a legacy from Unix computing where a daemon is a piece of software running on a server that offers a service.

Hyper Text Transfer Protocol

HTTP permits applications such as browsers to upload and download web pages. It makes use of TCP at the transport layer again to check reliability. HTTP is a connectionless protocol that sends a request, receives a response and then disconnects the connection. HTTP delivers HTML documents plus all of the other components supported within HTML such as JavaScript, Visual script and applets.

Simple Mail Transfer Protocol

By using TCP, SMTP sends email to other computers that support the TCP/IP protocol suite. SMTP provides extension to the local mail services that existed in the early years of LANs. It supervises the email sending from the local mail host to a remote mail host. It is not reliable for accepting mail from local users or distributing received mail to recipients this is the responsibility of the local mail system.

SMTP makes use of TCP to establish a connection to the remote mail host, the mail is sent, any waiting mail is requested and then the connection is disconnected. It

can also return a forwarding address if the intended recipient no longer receives email at that destination. To enable mail to be delivered across differing systems, a mail gateway is used.

Simple Network Management Protocol

For the transport of network management information, SNMP is used as standardized protocol. Managed network devices can be cross examined by a computer running to return details about their status and level of activity. Observing software can also trigger alarms if certain performance criteria drop below acceptable restrictions. At the transport layer SNMP protocol uses UDP. The use of UDP results in decreasing network traffic overheads.

In addition to widely known protocols, the application layer includes the following protocols:

- **Domain Name Service (DNS)**. Also called *name service*; this application maps IP addresses to the names assigned to network devices.
- **Routing Information Protocol (RIP).** Routing is central to the way TCP/IP works. RIP is used by network devices to exchange routing information.
- **Network File System (NFS).** A system developed by Sun Microsystems that enables computers to mount drives on remote hosts and operate them as if they were local drives.

1.3.5 A Comparison of the OSI and TCP/IP Models

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

Despite these fundamental similarities, the two models also have many differences. In this section we will focus on the key differences between the two reference models. It is important to note that we are comparing the *reference models* here, not the corresponding *protocol stacks.*

Three concepts are central to the OSI model:

1. Services.
2. Interfaces.
3. Protocols.

Probably the biggest contribution of the OSI model is to make the distinction between these three concepts explicit. Each layer performs some services for the layer above it. The *service* definition tells what the layer does, not how entities above it access it or how the layer works. It defines the layer's semantics. A layer's interface tells the processes above it how to access it. It specifies what the parameters are and what results to expect. It, too, says nothing about how the layer works inside.

Finally, the peer protocols used in a layer are the layer's own business. It can use any protocols it wants to, as long as it gets the job done (i.e., provides the offered services). It can also change them at will without affecting software in higher layers.

These ideas fit very nicely with modern ideas about object-oriented programming. An object, like a layer, has a set of methods (operations) that processes outside the object can invoke. The semantics of these methods define the set of services that the object offers. The methods parameters and results form the object's interface. The code internal to the object is its protocol and is not visible or of any concern outside the object.

The TCP/IP model did not originally clearly distinguish between service, interface and protocol, although people have tried to retrofit it after the fact to make it more OSI-like. For example, the only real services offered by the internet layer are SEND IP packet and RECEIVE IP PACKET. As a consequence, the protocols in the OSI model are better hidden than in the TCP/IP model and can be replaced relatively easily as the technology changes. Being able to make such changes is one of the main purposes of having layered protocols in the first place.

The OSI reference model was devised before the corresponding protocols were invented. This ordering means that the model was not biased toward one particular set of protocols, a fact that made it quite general. The downside of this ordering is that the designers did not have much experience with the subject and did not have a good idea of which functionality to put in which layer. For example, the data link layer originally dealt only with point-to-point networks. When broadcast networks came around, a new sublayer had to be hacked into the model. When people started to build real networks using the OSI model and existing protocols, it was discovered that these networks did not match the required service specifications (wonder of wonders), so convergence sublayers had to be grafted onto the model to provide a place for papering over the differences. Finally, the committee originally expected that each country would have one network, run by the government and using the OSI protocols, so no thought was given to internetworking. To make a long story short, things did not turn out that way.

With TCP/IP the reverse was true: the protocols came first, and the model was really just a description of the existing protocols. There was no problem with the protocols fitting the model. They fit perfectly. The only trouble was that the model did not fit any other protocol stacks. Consequently, it was not especially useful for describing other, non-TCP/IP networks.

Turning from philosophical matters to more specific ones, an obvious difference between the two models is the number of layers: the OSI model has seven layers and the TCP/IP has four layers. Both have (inter)network, transport and application layers, but the other layers are different. Another difference is in the area of connectionless versus connection-oriented communication. The OSI model supports both connectionless and connection-oriented communication in the network layer, but only connection-oriented communication in the transport layer, where it counts (because the transport service is

visible to the users). The TCP/IP model has only one mode in the network layer (connectionless) but supports both modes in the transport layer, giving the users a choice. This choice is especially important for simple request-response protocols.

OSI REFERENCE MODEL		TCP/IP REFERENCE MODEL	
It is the totality of all applications and their relating protocols that use networks and have not yet been represented by the lower layers.	**APPLICATION**	**APPLICATION**	Like OSI Model, it contains all the higher-level protocols
Here are the standards necessary for unambiguously representing data and more generally, syntax of messages to be transmitted (simple text, executable code, pictures...).	**PRESENTATION**		Because no need for them was perceived, Presentation and Session layers are not included in the TCP/IP Model
It establishes a connection with another node and manages the data flow from the higher layers to the lower ones by managing the timing of data transmission and the memory buffer managing, when several applications try to transmit data at the same time.	**SESSION**		
It handles the transmission, reception and error checking of the data.	**TRANSPORT**	**TRANSPORT**	The same as OSI Model
It is concerned with the physical transmission of the data from computer to computer. There is one further level of software to be considered, the network level. It routes the packages across a particular network.	**NETWORK**	**INTERNET**	It is the linchpin that holds the whole architecture together: it permits to send and receive packets, even if they are in random order.
It handles the transmission of a framed set of data (usually a sequence of bits) from one point in a network (node) to another one. This layer also represents the boundary between hardware (e.g. CRC) and software implementation (e.g. physical addressing).	**LINK**	**HOST TO**	The TCP/IP reference model does not really say much about what happens here, except to point out that the host has to
The physical medium used to transmit the information. To specify this layer, it is necessary to define the physical properties of the connection, such as mechanical properties, electrical/optical properties, functional aspects of the data transmission (modulation/demodulation for example) and procedural aspects of data transmission (e.g. bit stuffing to ensure that special signals are unequivocal).	**PHYSICAL**	**NETWORK**	Connect to the network using some protocol so it can send IP packets over it. This protocol is not defined and varies from host to host and network to network.

TABLE: OSI VERSUS TCP/IP REFERENCE MODEL

So, you can see that TCP/IP Reference Model and OSI Reference Model have a lot of things in common. Conceptually, it is useful to envision TCP/IP as a stack, each layer corresponding to a different facet of communication.

1.3.6 An Evaluation of the TCP/IP Reference Model

The TCP/IP model and protocols have their problems too.

- First, the model does not clearly distinguish the concepts of service, interface, and protocol. Good software engineering practice requires differentiating between the specification and the implementation, something that OSI does very carefully, and TCP/IP does not. Consequently, the TCP/IP model is not much of a guide for designing new networks using new technologies.
- Second, the TCP/IP model is not at all general and is poorly suited to describing any protocol stack other than TCP/IP. Trying to use the TCP/IP model to describe Bluetooth, for example, is completely impossible.
- Third, the host-to-network layer is not really a layer at all in the normal sense of the term as used in the context of layered protocols. It is an interface (between the network and data link layers). The distinction between an interface and a layer is crucial and one should not be sloppy about it.
- Fourth, the TCP/IP model does not distinguish (or even mention) the physical and data link layers. These are completely different. The physical layer has to do with the transmission characteristics of copper wire, fiber optics and wireless communication. The data link layer's job is to delimit the start and end of frames and get them from one side to the other with the desired degree of reliability. A proper model should include both as separate layers. The TCP/IP model does not do this.
- Finally, although the IP and TCP protocols were carefully thought out and well implemented, many of the other protocols were ad hoc, generally produced by a couple of graduate students hacking away until they got tired. The protocol implementations were then distributed free, which resulted in their becoming widely used, deeply entrenched and thus hard to replace. Some of them are a bit of an embarrassment now. The virtual terminal protocol, TELNET, for example, was designed for a ten-character per second mechanical Teletype terminal. It knows nothing of graphical user interfaces and mice. Nevertheless, 25 years later, it is still in widespread use.

In summary, despite its problems, the OSI model (minus the session and presentation layers) has proven to be exceptionally useful for discussing computer networks. In contrast, the OSI protocols have not become popular. The reverse is true of TCP/IP: the model is practically nonexistent, but the protocols are widely used.

1.3.7 Summary

The architecture of the TCP/IP protocol suite is often described in terms of a layered reference model called the TCP/IP model, DARPA model or DOD model. The TCP/IP architectural model has four layers that approximately match six of the seven layers in the OSI Reference Model. The TCP/IP model does not address the physical layer, which is where hardware devices reside. The next three layers—network interface, internet and (host-to-host) transport—correspond to layers 2, 3 and 4 of the OSI model. The TCP/IP application layer conceptually "blurs" the top three OSI layers. It's also worth noting that some people consider certain aspects of the OSI session layer to be arguably part of the TCP/IP host-to-host transport layer.

The internet and host-to-host transport layers are usually considered the "core" of TCP/IP architecture, since they contain most of the key protocols that implement TCP/IP internetworks. Network interface layer represents the place where the actual TCP/IP protocols running at higher layers interface to the local network Internet layer is responsible for typical layer three jobs, such as logical device addressing, data packaging, manipulation and delivery and last but not least, routing. Host-to-host transport layer is to facilitate end-to-end communication over an internetwork. Application layer is a rather broad layer, encompassing layers five through seven in the OSI model.

The OSI and TCP/IP reference models have much in common. Both are based on the concept of a stack of independent protocols. Also, the functionality of the layers is roughly similar. Despite these fundamental similarities, the two models also have many differences. The OSI reference model was devised before the corresponding protocols were invented. With TCP/IP the reverse was true: the protocols came first, and the model was really just a description of the existing protocols. The TCP/IP model does not clearly distinguish the concepts of service, interface, and protocol.

1.3.8 Self-Assessment Questions

1. What is a networking model? What are the advantages of networking models?
2. Explain the rationale behind the TCP/IP model. Briefly describe the role of each layer and its main functions.
3. Compare the Open Systems Interconnection (OSI) model with TCP/IP model.
4. Evaluate the TCP/IP model critically.

1.3.9 References and Suggested Readings

- Tanenbaum, Andrew.S. 2003, "Computer Networks." Pearson Education Inc.
- Sinha, Pradeep. K. 2003, "Foundations of Computing", BPB Publications
- Laudon, Kenneth, C. and Traver, carol, Guercio, 2003, "E-commerce", Pearson Education Inc.
- Forouzan, Behrouz. A. 2002, "Data Communications and Networking", Tata McGraw Hill Ltd.

LESSON NO. 1.4 **AUTHOR: GURINDER GOSAL**

TYPES OF NETWORKS

Structure of the Lesson:

Learning Objectives:

The major objectives of this lesson are to:

- Understand the basis of classification of computer networks
- Discuss various types of computer networks.
- Describe the general characteristics of each type of computer network.
- Understand the role of the different types of computer networks in various applications.

1.4.1 Introduction

One of the reasons that understanding networks can be difficult at times is that there are so many different types. When someone talks about a "network", this can mean anything from two computers hooked together in an apartment to a globe-spanning entity with millions of nodes. Every network is unique and each one has an important role to play in filling the communication and data-sharing needs of different individuals and organizations. In fact, the great diversity and flexibility of networking is one of its most important strengths.

In this section we describe the major types of networks that exist by drawing distinctions between them. There is no generally accepted classification into which all computer

networks fit, but one of the many dimensions stand out as important is scale.

1.4.2 Types of Computer Networks

In the criterion for classifying networks based on their **scale** we classify multiple processor systems by their physical size. At the top are the **personal area networks,** networks that are meant for one person. For example, a wireless network connecting a computer with its mouse, keyboard and printer is a personal area network. Also, a PDA that controls the user's hearing aid or pacemaker fits in this category. Beyond the personal area networks come longer-range networks. These can be divided into local, metropolitan and wide area networks. Finally, the connection of two or more networks is called an inter-network. Internet is a well-known example of an inter-network. Distance is important as a classification metric because different techniques are used at different scales

Inter-Processor Distance	Processor located in same	Example
1 m	Square meter	Personal Area Network
10 m	Room	
100 m	Building	} Local Area Network
1 km	Campus	
10 km	City	Metropolitan Area Network
100 km	Country	} Wide Area Network
1 000 km	Continent	
10,000 km	Planet	The internet

Table 1: Classification by Scale

1.4.3 Local Area Network (LAN)

The Local Area Network (LAN) is by far the most common type of data network. As the name suggests, a LAN serves a local area (typically the area of a floor of a building, but in some cases spanning a distance of several kilometers). Typical installations are in industrial plants, office buildings, college or university campuses or similar locations. In these locations, it is feasible for the owning organisation to install high quality, high-speed communication links interconnecting nodes. Typical data transmission speeds are one to 100 megabits per second.

A wide variety of LANs have been built and installed, but a few types have more recently become dominant. The most widely used LAN system is the Ethernet system developed by the Xerox Corporation.

Intermediate nodes (i.e. repeaters, bridges and switches) allow LANs to be connected together to form larger LANs. A LAN may also be connected to another LAN or to WANs and MANs using a "router".

LANs are restricted in size, which means that the worst-case transmission time is bounded and known in advance. Knowing this bound makes it possible to use certain kinds of designs that would not otherwise be possible. It also simplifies network management.

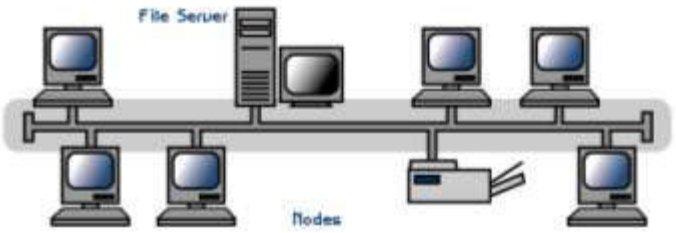
LANs may use a transmission technology consisting of a cable to which all the machines are attached, like the telephone company party lines once used in rural areas. Traditional LANs run at speeds of 10 Mbps to 100 Mbps, have low delay (microseconds or nanoseconds) and make very few errors. Newer LANs operate at up to 10 Gbps.

In summary, a LAN is a communications network which is:

- Local (i.e. One building or group of buildings)
- Controlled by one administrative authority
- Assumes other users of the lan are trusted
- Usually high speed and is always shared

LANs allow users to share resources on computers within an organization and may be used to provide a (shared) access to remote organizations through a router connected to a Metropolitan Area Network (MAN) or a Wide Area Network (WAN).

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



(a) Bus network

(b) Ring

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

Broadcast networks can be further divided into static and dynamic, depending on how the channel is allocated. A typical **static allocation** would be to divide time into discrete intervals and use a round-robin algorithm, allowing each machine to broadcast only when its time slot comes up. Static allocation wastes channel capacity when a machine has nothing to say during its allocated slot, so most systems attempt to allocate the channel dynamically (i.e., on demand).

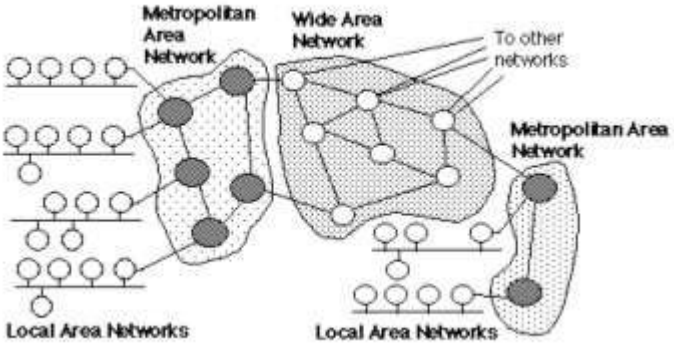
Dynamic allocation methods for a common channel are either centralized or decentralized. In the centralized channel allocation method, there is a single entity, for example, a bus arbitration unit, which determines who goes next. It might do this by accepting requests and making a decision according to some internal algorithm. In the decentralized channel allocation method, there is no central entity; each machine must decide for itself whether to transmit. You might think that this always leads to chaos, but it does not. Later we will study many algorithms designed to bring order out of the potential chaos.

1.4.4 Metropolitan Area Network

A **metropolitan area network,** or **MAN,** covers a city. The best-known example of a MAN is the cable television network available in many cities. This system grew from earlier community antenna systems used in areas with poor over-the-air television reception. In these early systems, a large antenna was placed on top of a nearby hill and signal was then piped to the subscriber's houses.

At first, these were locally-designed, ad hoc systems. Then companies began jumping into the business, getting contracts from city governments to wire up an entire city. The next step was television programming and even entire channels designed for cable only.

Starting when the Internet attracted a mass audience, the cable TV network operators began to realize that with some changes to the system, they could provide two-way Internet service in unused parts of the spectrum. At that point, the cable TV system began to move from a way to distribute television to a metropolitan area network. A typical use of MANs to provide shared access to a wide area network is shown in the figure. Cable television is not the only MAN, recent developments in high-speed wireless Internet access resulted in another MAN.



which share the cost of access to a WAN

1.4.5 Wide Area Networks

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

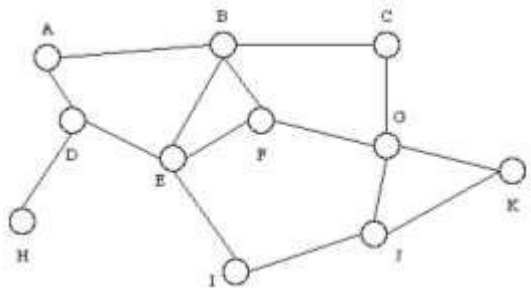


Figure: Typical "mesh" connectivity of a Wide Area Network

A typical network is shown in the figure above. This connects a number of End Systems (ES) (e.g. A, C, H, K) and a number of Intermediate Systems (IS) (e.g. B, D, E, F, G, I, J) to form a network over which data may be communicated between the End Systems (ES).

In most wide area networks, the subnet consists of two distinct components: transmission lines and switching elements. Transmission lines move bits between machines. They can be made of copper wire, optical fiber or even radio links. Switching

elements are specialized computers that connect three or more transmission lines. When data arrive on an incoming line, the switching element must choose an outgoing line on which to forward them. These switching computers have been called by various names in the past; the name router is now most commonly used.

In most WANs, the network contains numerous transmission lines, each one connecting a pair of routers. If two routers that do not share a transmission line wish to communicate, they must do this indirectly, via other routers. When a packet is sent from one router to another via one or more intermediate routers, the packet is received at each intermediate router in its entirety, stored there until the required output line is free, and then forwarded. A subnet organized according to this principle is called a **store-and-forward** or **packet-switched** subnet. Nearly all wide area networks (except those using satellites) have store-and-forward subnets. When the packets are small and all the same size, they are often called **cells.**

The principle of a packet-switched WAN is so important that it is worth devoting a few more words to it. Generally, when a process on some host has a message to be sent to a process on some other host, the sending host first cuts the message into packets, each one bearing its number in the sequence. These packets are then injected into the network one at a time in quick succession. The packets are transported individually over the network and deposited at the receiving host, where they are reassembled into the original message and delivered to the receiving process.

Sr. No.	Characteristic	LAN	WAN
1	**Geographical Distribution**	Limited geographical coverage (few Kms)	Greater geographical coverage (several thousand Kms)
2	**Data Transmission Rate**	Much higher (0,2 Mbps to 1Gbps)	Much slower(1200 bps to 1Mbps)
3	**Communication Link**	Twisted pair, coaxial cable, fiber optics	Telephone lines, Microwave links, satellite channels
4	**Error Rate**	Fewer errors (bit error rates 10^{-8} to 10^{-12}	More errors (bit error rates 10^{-5} to 10^{-7}
5	**Ownership**	Usually owned by a single organization	Usually owned by different organizations
6	**Communication Cost**	Low	High

Table 2: Comparison of LAN with WAN

Not all WANs are packet switched. A second possibility for a WAN is a satellite system. Each router has an antenna through which it can send and receive. All routers can hear the output *from* the satellite and in some cases they can also hear the upward

transmissions of their fellow routers *to* the satellite as well. Sometimes the routers are connected to a substantial point-to-point subnet, with only some of them having a satellite antenna. Satellite networks are inherently broadcast and are most useful when the broadcast property is important.

The comparison of WAN with LAN on some key characteristics is depicted in table 2.

1.4.6 Wireless Networks

Digital wireless communication is not a new idea. Modern digital wireless systems have better performance, but the basic idea is the same.

To a first approximation, wireless networks can be divided into three main categories:

- System interconnection.
- Wireless LANs.
- Wireless WANs.

System interconnection is all about interconnecting the components of a computer using short-range radio. Almost every computer has a monitor, keyboard, mouse and printer connected to the main unit by cables. So many new users have a hard time plugging all the cables into the right little holes that most computer vendors offer the option of sending a technician to the user's home to do it. Consequently, some companies got together to design a short-range wireless network called **Bluetooth** to connect these components without wires. Bluetooth also allows digital cameras, headsets, scanners and other devices to connect to a computer by merely being brought within range. No cables, no driver installation, just put them down, turn them on and they work. For many people, this ease of operation is a big plus.

In the simplest form, system interconnection networks use the master-slave paradigm. The system unit is normally the master, talking to the mouse, keyboard, etc., as slaves. The master tells the slaves what addresses to use, when they can broadcast, how long they can transmit, what frequencies they can use and so on.

The next step up in "wireless networking is the **wireless LANs.** These are systems in which every computer has a radio modem and antenna with which it can communicate with other systems. Often there is an antenna on the ceiling that the machines talk to. However, if the systems are close enough, they can communicate directly with one another in a peer-to-peer configuration. Wireless LANs are becoming increasingly common in small offices and homes, where installing Ethernet is considered too much trouble, as well as in older office buildings, company cafeterias, conference rooms, and other places. There is a standard for wireless LANs, called **IEEE 802.11,** which most systems implement and which is becoming very widespread.

The third kind of wireless network is used in wide area systems (Wireless WANs). The radio network used for cellular telephones is an example of a low-bandwidth wireless system. This system has already gone through three generations. The first generation was analog and for voice only. The second generation was digital and for voice only. The third generation is digital and is for both voice and data. In a certain

sense, cellular wireless networks are like wireless LANs, except that the distances involved are much greater and the bit rates much lower. Wireless LANs can operate at rates up to about 50 Mbps over distances of tens of meters. Cellular systems operate below 1 Mbps, but the distance between the base station and the computer or telephone is measured in kilometers rather than in meters.

In addition to these low-speed networks, high-bandwidth wide area wireless networks are also being developed. The initial focus is high-speed wireless Internet access from homes and businesses, bypassing the telephone system. -This service is often called local multipoint distribution service. A standard for it, called IEEE 802.16, has also been developed.

Almost all wireless networks hook up to the wired network at some point to provide access to files, databases and the Internet. There are many ways these connections can be realized, depending on the circumstances.

1.4.7 Home Networks

Home networking is on the horizon. The fundamental idea is that in the future most homes will be set up for networking. Every device in the home will be capable of communicating with every other device and all of them will be accessible over the Internet. This is one of those visionary concepts that nobody asked for, but once they arrived nobody can imagine how they lived without them.

Many devices are capable of being networked. Some of the more obvious categories (with examples) are as follows:

1. Computers (desktop PC, notebook PC, PDA, shared peripherals).
2. Entertainment (TV, DVD, VCR, camcorder, camera, stereo, MP3).
3. Telecommunications (telephone, mobile telephone, intercom, fax).
4. Appliances (microwave, refrigerator, clock, furnace, lights).
5. Telemetry (utility meter, smoke/burglar alarm, thermostat).

Home computer networking is already here in a limited way. Many homes already have a device to connect multiple computers to a fast Internet connection. Networked entertainment is not quite here, but as more and more music and movies can be downloaded from the Internet, there will be a demand to connect stereos and televisions to it. Also, people will want to share their own videos with friends and family, so the connection will need to go both ways. Telecommunications gear is already connected to the outside world, but soon it will be digital and go over the Internet. The average home probably has a dozen clocks (e.g., in appliances), all of which have to be reset twice a year when daylight saving time (summer time) comes and goes. If all the clocks were on the Internet, that resetting could be done automatically. Finally, remote monitoring of the home and its contents is a likely winner. While one can imagine a separate network for each application area, integrating all of them into a single network is probably a better idea.

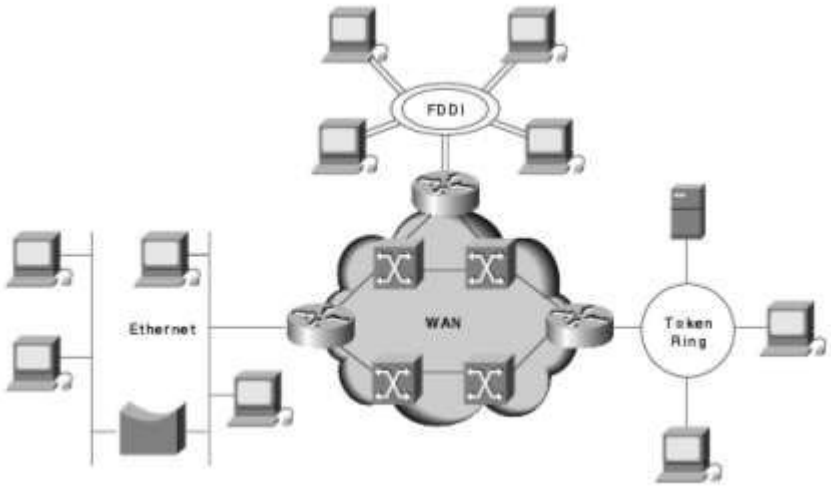
Home networking has some fundamentally different properties than other

network types.

1. First, the network and devices have to be easy to install. A series of phone calls to the vendor's help-desk typically resulted in answers like (1) Read the manual, (2) Reboot the computer, (3) Remove all hardware and software except ours and try again, (4) Download the newest driver from our Web site and if all else fails, (5) Reformat the hard disk and then reinstall Windows from the CD-ROM. Telling the purchaser of an Internet refrigerator to download and install a new version of the refrigerator's operating system is not going to lead to happy customers. Computer users are accustomed to putting up with products that do not work; the car, television and refrigerator-buying public is far less tolerant. They expect products to work for 100% from the word go.
2. Second, the network and devices have to be foolproof in operation. Air conditioners used to have one knob with four settings: OFF, LOW, MEDIUM, and HIGH. Now they have 30-page manuals. Once they are networked, expect the chapter on security alone to be 30 pages. This will be beyond the comprehension of virtually all the users.
3. Third, low price is essential for success. People will not pay a $50 premium for an Internet thermostat because few people regard monitoring their home temperature from work that important. For $5 extra, it might sell, though.
4. Fourth, the main application is likely to involve multimedia, so the network needs sufficient capacity. There is no market for Internet-connected televisions that show shaky movies at 320x240 pixel resolution and 10 frames/sec. Fast Ethernet, the workhorse in most offices, is not good enough for multimedia. Consequently, home networks will need better performance than that of existing office networks and at lower prices before they become mass market items.
5. Fifth, it must be possible to start out with one or two devices and expand the reach of the network gradually. This means no format wars. Telling consumers to buy peripherals with IEEE 1394 (FireWire) interfaces and a few years later retracting that and saying USB 2.0 is the interface-of-the-month is going to make consumers skittish. The network interface will have to remain stable for many years; the wiring (if any) will have to remain stable for decades.
6. Sixth, security and reliability will be very important. Losing a few files to an e-mail virus is one thing; having a burglar disarm your security system from his PDA and then plunder your house is something quite different.

1.4.8 Internetwork

An *internetwork* is a collection of individual networks, connected by intermediate networking devices, that functions as a single large network. Internetworking refers to the industry, products and procedures that meet the challenge of creating and administering internetworks. Figure here illustrates some different kinds of network technologies that can be interconnected by routers and other networking devices to create an internetwork.



Many networks exist in the world, often with different hardware and software. People connected to one network often want to communicate with people attached to a different one. The fulfillment of this desire requires that different and frequently incompatible networks, be connected, sometimes by means of machines called gateways to make the connection and provide the necessary translation, both in terms of hardware and software. Internetworking evolved as a solution to three key problems: isolated LANs, duplication of resources and a lack of network management.

- Isolated LANs made electronic communication between different offices or departments impossible.
- Duplication of resources meant that the same hardware and software had to be supplied to each office or department, as did separate support staff.
- The lack of network management meant that no centralized method of managing and troubleshooting networks existed.

1.4.8.1 Implementation Challenges in Internetworking

Implementing a functional internetwork is no simple task. Many challenges must be faced, especially in the areas of connectivity, reliability, network management, and flexibility. Each area is key area in establishing an efficient and effective internetwork. The challenge when connecting various systems is to support communication among disparate technologies. Different sites, for example, may use different types of media operating at varying speeds or may even include different types of systems that need to communicate.

Because companies rely heavily on data communication, internetworks must provide a certain level of reliability. This is an unpredictable world, so many large internetworks include redundancy to allow for communication even when problems occur.

Furthermore, network management must provide centralized support and

troubleshooting capabilities in an internetwork. Configuration, security, performance, and other issues must be adequately addressed for the internetwork to function smoothly. Security within an internetwork is essential. Many people think of network security from the perspective of protecting the private network from outside attacks. However, it is just as important to protect the network from internal attacks, especially because most security breaches come from inside. Networks must also be secured so that the internal network cannot be used as a tool to attack other external sites.

Early in the year 2000, many major web sites were the victims of distributed denial of service (DDOS) attacks. These attacks were possible because a great number of private networks currently connected with the Internet were not properly secured. These private networks were used as tools for the attackers.

Because nothing in this world is stagnant, internetworks must be flexible enough to change with new demands.

1.4.9 Value Added Networks (VAN)

Some companies specialize in providing value added data transmission services. The value added over and above the standard services of the common carriers may include electronic email, data encryption/decryption and access to commercial databases, and code conversion for communication between incompatible computers. These companies normally lease dedicated lines of a common carrier, do value addition to enhance the communication facilities and then sell the enhanced service. This type of service is popularly known as value-added network (VAN) service.

Value-added networks got their first real foothold in the business world in the area of Electronic Data Interchange (EDI). VANs were deployed to help trading and supply chain partners automate many business-to-business communications and thereby reduce the number of paper transfers needed, cut costs and speed up a wide range of tasks and processes, from inventory and order management to payment.

In today's world, e-commerce is increasingly based on XML, though EDI remains an important part of business and still relies on value-added networks. But other types of VANs have begun to appear, including Web services networks and transaction delivery networks.

A VAN is created when a network provider leases communications lines from a common carrier (such as a telephone company), enhances them by adding services and improvements that facilitate business-to-business application integration, and then resells the network connection and services to others for a fee. Value, then, refers to what the supplier provides beyond the basic network connection and becomes the critical element differentiating one network offering from another.

1.4.9.1 VANs and EDI

In the 1980s, VANs emerged as a way to connect supply chain participants. They offered store-and-forward mailboxes that provided protocol conversion, security and guaranteed delivery.

However, EDI VANs proved to be too costly for most businesses. Only the largest of supply chain participants could afford the expensive setup fees associated with EDI software, not to mention the sometimes exorbitant per-transaction fees. This meant that many small and midsize firms couldn't afford to join electronic, automated supply chains. And since the smaller companies couldn't join in, the larger ones that continued to do business with them still couldn't eliminate a lot of their traditional, paper-based processes. When the Internet and the World Wide Web entered the picture, along with standards like ebXML, some observers felt that VANs might simply disappear.

While traditional EDI is in many ways inferior to newer approaches, it still offers a compelling business model. In fact, the continuing presence of EDI is not a result of its fundamental technology but instead can be attributed to its underlying communications structure, the VAN, which can guarantee and secure B2B interaction over a network.

1.4.9.2 Web Services Networks

In its reincarnation, the EDI VAN model has become the Web services network (WSN, or sometimes WSVAN), which has to meet many of the same requirements and features that EDI users depend on. The final hurdles for business-to-business interaction across the Internet are dependability and security. Most systems and networks within a given company are designed to manage transaction flows behind a secure, centrally managed firewall. Unfortunately, the B2B world exists between firewalls, with partners often sharing no common infrastructure. Moreover, the Internet itself offers little security, reliability or accountability.

1.4.9.3 Transaction Delivery Networks

The newest evolution of VANs, which first appeared in 2000, are the transaction delivery networks (TDN) that provide services for secure end-to-end management of electronic transactions.

Also called transaction processing networks or Internet utility platforms, TDNs can guarantee delivery and nonrepudiation of messages in addition to providing high security and availability, network performance monitoring and centralized directory management.

TDNs typically use a store-and-forward messaging architecture that's designed to adapt readily to a wide range of disparate systems and support any kind of transaction. Most TDNs offer secure encryption using a public-key infrastructure and certificate authorization for trading partners.

TDNs provide standards-based application programming interfaces that developers can use to create custom applications to link internal data sources with the TDN. In addition, most TDNs provide application adapters that plug directly into existing computing environments, such as messaging middleware.

1.4.10 Summary

There is no generally accepted classification into which all computer networks fit, but one of the many dimensions stand out as important is size.

The Local Area Network (**LAN**) is by far the most common type of data network. A LAN serves a local area (typically the area of a floor of a building, but in some cases spanning a distance of several kilometers). Typical installations are in industrial plants, office buildings, college or university campuses, or similar locations. A Metropolitan Area Network (MAN) is one of a number of types of networks. **MAN** is a bigger version of a LAN and uses similar technology. The network size falls intermediate between LANs and WANs. A MAN typically covers an area of between 5 and 50 km diameter. Many MANs cover an area the size of a city, although in some cases MANs may be as small as a group of buildings or as large as the metro city. A **WAN** spans a large area, often a country or continent. Wireless networks can be divided into three main categories: System interconnection, Wireless LANs, Wireless WANs. Almost all wireless networks hook up to the wired network at some point to provide access to files, databases and the Internet. There are many ways these connections can be realized, depending on the circumstances.

Home networking is on the perspective. Every device in the home will be capable of communicating with every other device and all of them will be accessible over the Internet. Home networking offers many opportunities and challenges. Most of them relate to the need to be easy to manage, dependable and secure, especially in the hands of non-technical users, while at the same time delivering high performance at low cost. An *internetwork* is a collection of individual networks, connected by intermediate networking devices, that functions as a single large network. Internetworking refers to the industry, products and procedures that meet the challenge of creating and administering internetworks. Some companies specialize in providing value added data transmission services. These companies normally lease dedicated lines of a common carrier, do value addition to enhance the communication facilities, and then sell the enhanced service. This type of service is popularly known as value-added network (VAN) service.

1.4.11 Self Check Exercise

1. On what basis you can classify computer network? Explain.
2. Discuss these types of computer networks in detail:
   - Local Area Network (LAN)
   - Wide Area Network (WAN)
   - Metropolitan Area Network (MAN)
   - Wireless Networks
   - Value Added network (VAN)
3. What are Home networks? On what characteristics they are fundamentally different from other network types?
4. What is internetwork? What are the challenges faced in implementing these?
5. Compare WAN with LAN on basis of certain key characteristics.

1.4.12 References and Suggested Readings

- Laudon, Kenneth, C. and Traver, carol, Guercio, 2003, “E-commerce”, Pearson Education Inc.
- Sinha, Pradeep. K. 2003, “Foundations of Computing”, BPB Publications
- Kalakota, Ravi, 2006, “Frontiers of Electronic Commerce”,
- Forouzan, Behrouz. A. 2002, “Data Communications and Networking”, Tata McGraw Hill Ltd.

LESSON NO. 1.5
AUTHOR: GURINDER GOSAL

NETWORK TOPOLOGIES

Structure of the Lesson:

Learning Objectives:

The major objectives of this lesson are to:

- Define network topology
- Understand the meaning of physical and logical topologies.
- Understand the different types of computer network topologies and their relative advantages and disadvantages.

1.5.1 Introduction

The network is made up of two types of components: nodes and communication lines. The physical topology of a network refers to the configuration of cables, computers and other peripherals. Physical topology should not be confused with logical topology which is the method used to pass information between workstations. Although the number of possible network topologies is seemingly limitless, the major ones are star network, the ring network, the completely connected network and the multi-access bus network.

1.5.2 Definition of Network Topology

As we know a computer network is the infrastructure that allows two or more computers (called hosts) to communicate with each other. The topology of a network refers to the way in which all of its pieces have been connected.

There are two types of topologies: Physical and Logical.

Physical topology of a network refers to the layout of cables, computers and other

peripherals. Try to imagine yourself in a room with a small network, you can see network cables coming out of every computer that is part of the network, then those cables plug into a hub or switch. What you're looking at is the physical topology of that network

Logical topology is the method used to pass the information between the computers. In other words, looking at that same room, if you were to try to see how the network works with all the computers talking (think of the computers generating traffic and packets of data going everywhere on the network) you would be looking at the logical part of the network. The way the computers will be talking to each other and the direction of the traffic is controlled by the various protocols (like Ethernet) or, if you like, rules.

The physical topology describes the layout of the network, just like a map shows the layout of various roads and the logical topology describes how the data is sent across the network or how the cars are able to travel (the direction and speed) at every road on the map.

1.5.2.1 Factors Affecting Choice of Network Topology

The choice of network topology for installing a computer network depends upon a combination of factors such as:

1. The desired reliability of the entire system.
2. The desired performance of the system.
3. Size (number of nodes and their geographical distribution) of the system.
4. Availability of communication lines e.g. the most common cable in schools is unshielded twisted pair, which is most often used with star topologies
5. Expandability of the system e.g. with a star topology, expanding a network is easily done by adding another concentrator
6. Cost of the components and services required to implement the network. A linear bus network may be the least expensive way to install a network; you do not have to purchase concentrators.
7. Delays involved in routing information from one node to another.

1.5.3 Main Types of Physical Topologies

The most common types of physical topologies which we are going to analyze are:

- Multi-access Bus Network Topology
- Star Network Topology
- Ring Network Topology
- The Completely Connected Network Topology
- The Hybrid Network Topology

1.5.3.1 Multi-access Bus Network Topology

Figure1 shows a multi-access bus network. In this type of network, a single transmission medium is shared by all nodes. That is, all the computers are attached to the same communication line (channel). When a particular computer wants to send a message to another computer, it appends the destination address to the message and checks whether the communication line is free. As soon as the line becomes free, it broadcasts

(places) message on the line. As the message travels on the line, each computer checks whether it is addressed to it. The message is picked up by the addressee computer, which sends an acknowledgement to the source computer and frees the line. This type of network is also known as 'multipoint' or 'multi-drop' or 'broadcasting' network". It is appropriate for use in a local area network where a high-speed communication channel is used and computers are confined to a small area. It is also appropriate when satellite communication is used as one satellite channel may be shared by many computers at a number of geographical locations.

Ethernet bus topologies are relatively easy to install and don't require much cabling compared to the alternatives. 10Base-2 ("ThinNet") and 10Base-5 ("ThickNet") both were popular Ethernet cabling options many years ago for bus topologies. However, bus networks work best with a limited number of devices. If more than a few dozen computers are added to a network bus, performance problems will likely result. In addition, if the backbone cable fails, the entire network effectively becomes unusable.

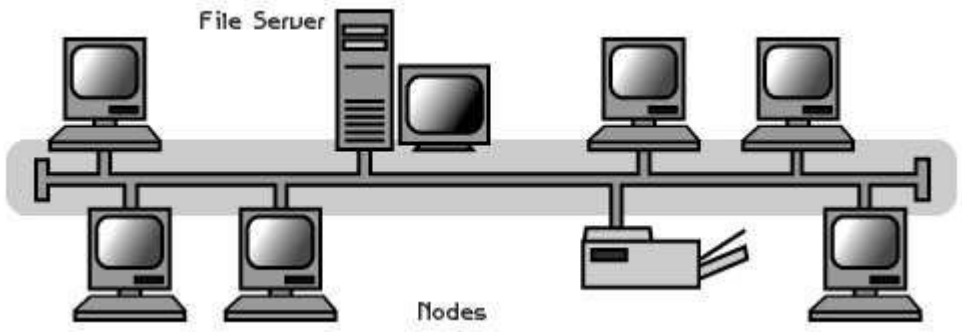


Figure1. Multi-access Bus topology

Single communication line shared by all nodes

Advantages

1. The main advantage of a multi-access bus network is the reduction in physical lines.
2. The failure of a computer in the network does not affect the network functioning for other.
3. Addition of new computers or peripheral to the network is easy.
4. Requires less cable length than a star topology.

Disadvantages

1. All computers in the network must have good communication and decision-making capability.
2. If the communication line fails, the entire system breaks down.
3. Terminators are required at both ends of the backbone cable.
4. Difficult to identify the problem if the entire network shuts down.

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

5.3.2 Star Network Topology

Figure 2 shows the star arrangement of a computer network. In this configuration, multiple computers connected to a host computer. That is, the computers in the network are not linked directly to each other communicate only via the host computer. The routing function is performed by the host computer centrally controls communication between any two other computers by establishing a logical path between them.

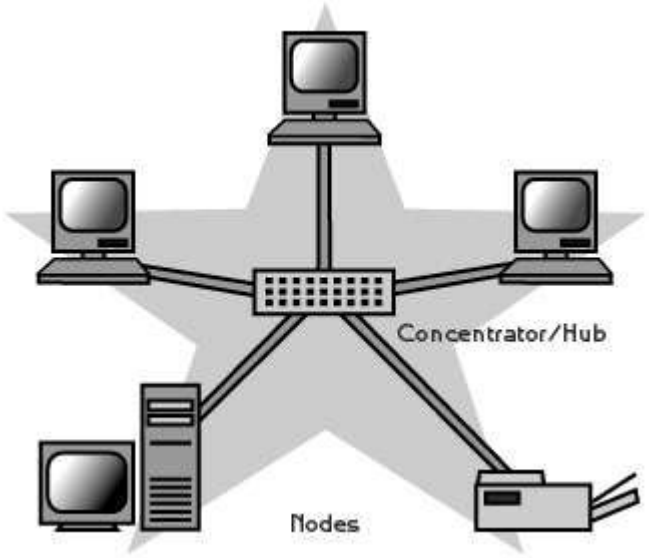


Figure 2: A star configuration of computer network.

Advantages

1. Star topology has minimal line cost because only n-1 lines are required for connecting n nodes.
2. Transmission delays between two nodes do not increase by adding new nodes to the network because any two nodes may be connected via two links only.
3. If any of the local computers fails, the remaining portion of the network is unaffected.
4. Easy to install and wire.
5. Easy to detect faults and to remove parts.

Disadvantages

1. The system crucially depends on the central node. If the host computer fails, the entire network fails.
2. Requires more cable length than a linear topology.
3. More expensive than linear bus topologies because of the cost of the concentrators.

1.5.3.3 Ring Network

Figure3 shows the circular or ring arrangement of a computer network. In this configuration, each computer in the network has communicating subordinates, but within the ring there is no master computer for controlling other computers. A node receives data from one of its two adjacent nodes. The only decision a node has to take is whether the data is for its own use or not. If it is addressed to it, it utilizes it. Otherwise, it merely passes it on to the next node.

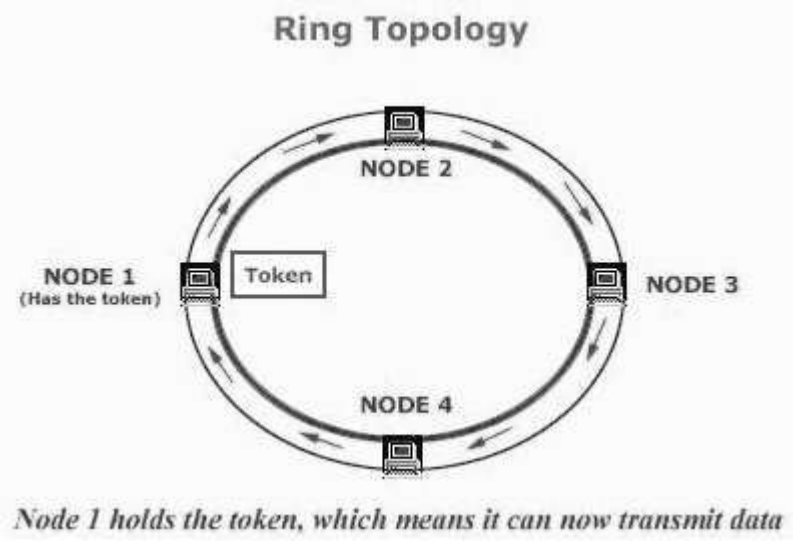


Figure 3: A ring configuration of computer network.

Advantages

I. The ring network works well where there is no central-site computer system.

2. It is more reliable than a star network because communication is not dependent on a single host computer. If a link between any two computers breaks down, or if one of the computers breaks down, alternate routing is possible.

Disadvantages

1. In a ring network, communication delay is directly proportional to the number of nodes in the network. Hence addition of new nodes in the network increases the communication delay.

2. The ring network requires more complicated control software than star network.

1.5.3.4 The Physical Mesh Topology

As shown in Figure 4 a completely connected network or mesh topology has a separate physical link for connecting each node to any other node. Thus, each computer of such a network has a direct dedicated link, called a point-to-point link, with all other computers in the network. The control is distributed with each computer deciding its communication priorities. On a large scale, you can connect multiple LANs using mesh topology with leased telephone lines, thicknet coaxial cable or fiber optic cable. Again,

the big advantage of this topology is its backup capabilities by providing multiple paths through the network.

Advantages

1. This type of network is very reliable, as any link breakdown will affect only communication between the connected computers.
2. Each node of the network need not have individual routing capability.
3. Communication is very fast between any two nodes.

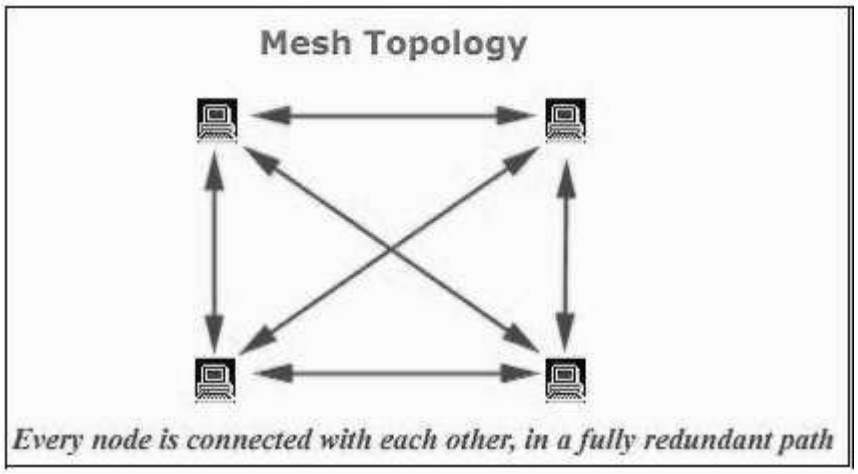


Figure 4: Completely connected topology

Disadvantage

1. It is the most expensive network from the point of view of link cost. If there are n nodes in network, then n (n-1)/2 links are required. Thus, the cost of linking the system grows with the square of the number of nodes.

1.5.3.5 Hybrid Network

Different network configurations have their own advantages and limitations. Hence in reality, a pure star or ring or completely connected network is rarely used. Instead, an organization will use some sort of hybrid network, which is a combination of two or more different network topologies. The exact configuration of the network depends on the needs and the overall organizational structure of the company involved. In some cases, the hybrid network may have components of star, ring and completely connected networks.

1.5.3.5.1 Star-Bus Hybrid Network

As with the hybrid topology, two or more topologies are combined to form a complete network. For example, a hybrid topology could be the combination of a star and bus topology. These are also the most common in use.

In a star-bus topology, several star topology networks are linked to a bus connection. In this topology, if a computer fails, it will not affect the rest of the network. However, if the central component or hub that attaches all computers in a star fails, then you have

big problems since no computer will be able to communicate.

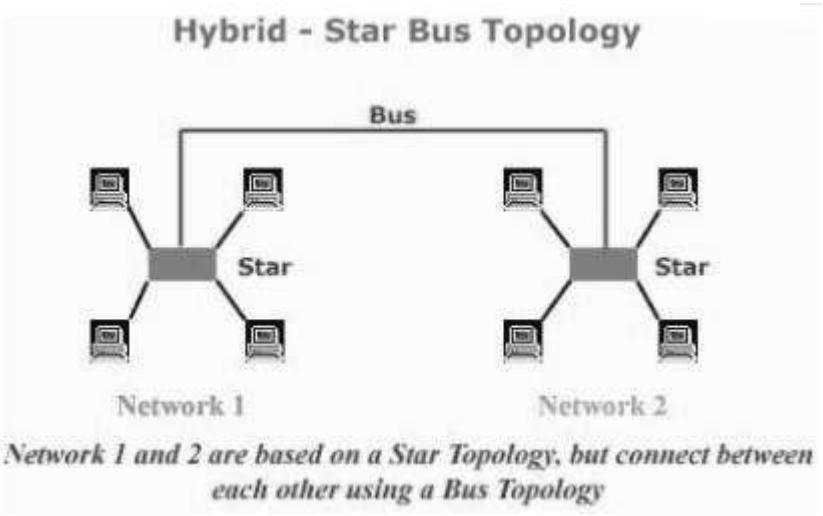


Figure 5: Star-Bus topology

5.3.5.2 Star-Ring Hybrid Network

In the Star-Ring topology, the computers are connected to a central component as in a star network. These components, however, are wired to form a ring network.

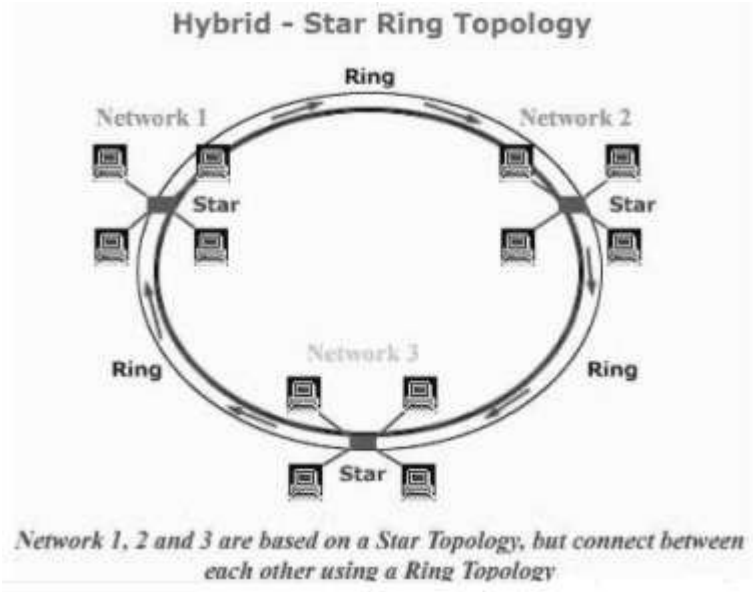


Figure 6: Star-Ring topology

Like the star-bus topology, if a single computer fails, it will not affect the rest of the network. By using token passing, each computer in a star-ring topology has an equal chance of communicating. This allows for greater network traffic between segments than in a star-bus topology.

1.5.4 Summary

A computer network consists of nodes and communication links which implement its protocols. It interconnects a set of hosts which conform to the network protocols. Topologies remain an important part of network design theory. Network topology are the physical layout of the network that the locations of the computers and how the cable is run between them. The term network topology refers to the way in which the nodes of a network are linked together. It determines the data paths that may be used between any pair of nodes in the network. You can probably build a home or small business network without understanding the difference between a bus design and a star design, but understanding the concepts behind these gives you a deeper understanding of important elements like hubs, broadcasts and routes. The choice of network topology for installing a computer network depends upon a combination of factors such as desired reliability, performance, size etc. Physical topology of a network refers to the layout of cables, computers and other peripherals. Logical topology is the method used to pass the information between the computers.

The most common types of physical topologies are: Multi-access Bus Network Topology, Star Network Topology, Ring Network Topology, the Completely Connected Network Topology and The Hybrid Network Topology. In a multi-access bus network, a single transmission medium is shared by all nodes. In star network the computers in the network are not linked directly to each other communicate only via the host computer. In ring network, each computer in the network has communicating subordinates, but within the ring there is no master computer for controlling other computers. In mesh network each computer of such a network has a direct dedicated link, called a point-to-point link, with all other computers in the network. Hybrid network is a combination of two or more different network topologies.

1.5.5 Self Check Exercise

1. Define network topology. Differentiate between the physical and logical network topologies.
2. What are the various types of physical network topologies prevalent? Discuss any two of these topologies in detail.
3. What is a hybrid network? Why they are used?
4. Describe three commonly used network topologies with their relative advantages and disadvantages.

1.5.6 References and Suggested Readings

- Forouzan, Behrouz. A. 2002, “Data Communications and Networking”, Tata McGraw Hill Ltd.
- Laudon, Kenneth, C. and Traver, carol, Guercio, 2003, “E-commerce”, Pearson Education Inc.
- Sinha, Pradeep. K. 2003, “Foundations of Computing”, BPB Publications
- Kalakota, Ravi, 2006, “Frontiers of Electronic Commerce”,

LESSON NO. 1.6 AUTHOR : GURINDER GOSAL

TRANSMISSION MEDIA

Structure of the Lesson:

Learning Objectives:

The major objectives of this lesson are to:

- Discuss transmission fundamentals.
- Understand how data is transmitted and the basic techniques that this process involves.
- Discuss the factors influencing the choice of transmission media.
- Have a broad understanding of the different physical transmission media and their characteristics.

6.1 Introduction

Transmission is the act of transporting information from one location to another via a signal. The signal may be analog or digital and may travel in different media. Digital data can be transmitted over many different types of media and choice of a transmission medium is guided by comparing transmission requirements against the medium's characteristics. Different important criteria influence the choice of medium.

Bandwidth, reliability, cost and coverage serve as the basic criteria for the choice. Transmission media can be divided into categories of guided and unguided media. Guided transmission media are those that provide a conduit from one device to another. Unguided transmission media or wireless communications are those that transport electromagnetic waves without using a physical conductor. We will discuss here different types of guided transmission media that are Magnetic media, Copper wire and Fiber optics.

6.2 Transmission Fundamentals

Since digital computers play a central role in data communication, in nearly all cases, digital signals are used. Analog signals are used in cases of equipment which date back to before the advent of digital technology. Existing analog telephone networks are a good example of the latter. Modulation is used for transmission of digital data over an analog line. Three basic types of modulation are AM, FM and PM. In digitization an analog signal is converted into digital format through a process of sampling. It is much easier to reliably transmit a digital signal over a long distance than an analog signal. The process of synchronization is to maintain an agreement about the exact distribution of data over time between receiver and transmitter. Here we discuss certain fundamentals related to transmission.

1.6.2.1 Connection Types

Connections between devices may be classified into three categories:

1. **Simplex.** This is a unidirectional connection, i.e., data can only travel in one direction. Simplex connections are useful in situations where a device only receives or only sends data (e.g., a printer).
2. **Half-duplex.** This is a bidirectional connection, with the restriction that data can travel in one direction at a time.
3. **Full-duplex.** This is a bidirectional connection in which data can travel in both directions at once. A full-duplex connection is equivalent to two simplex connections in opposite directions.

6.2.2 Signal Types

All signals are either analog or digital. An **analog signal** is one in which information appears as a continuous variation of some property. Human speech is an example: it produces a continuous variation of air pressure. A **digital signal**, on the other hand, is one in which information appears as a sequence of binary values 0 and 1. To represent these two values, a signal is used in which only two wave shapes are allowed, one representing the binary value 0 and the other representing the binary value 1. By definition, therefore, a digital signal is a restricted form of an analog signal. A human speaker who only utters the two words *zero* and *one* is a crude example of a digital signal. In electrical terms, signals appear as variation of some electrical property (e.g., voltage).

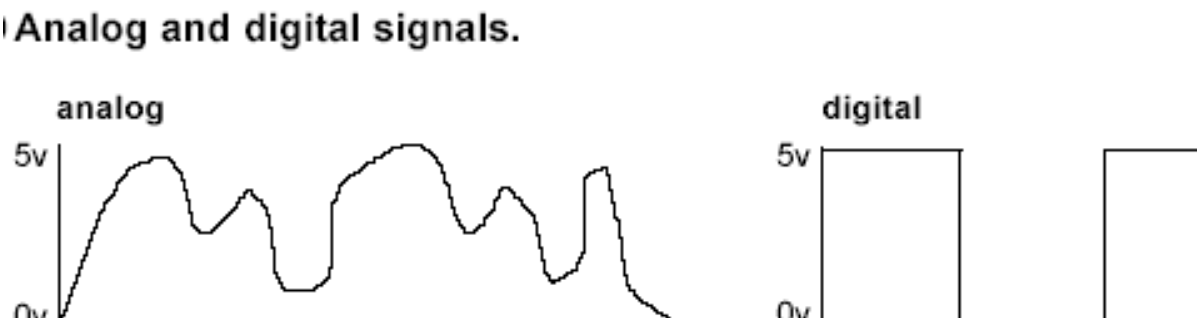


Figure: Analog and Digital Signals

6.2.3 Modulation

Transmission of digital data over an analog line is achieved using a technique called **modulation**, where the digital bit stream is modulated over an analog carrier signal. A **modem** (modulator and demodulator) is a commonly used device which employs this technique. A modem converts the outgoing digital bit stream from a device into an analog signal and converts the incoming analog signal into a digital bit stream.

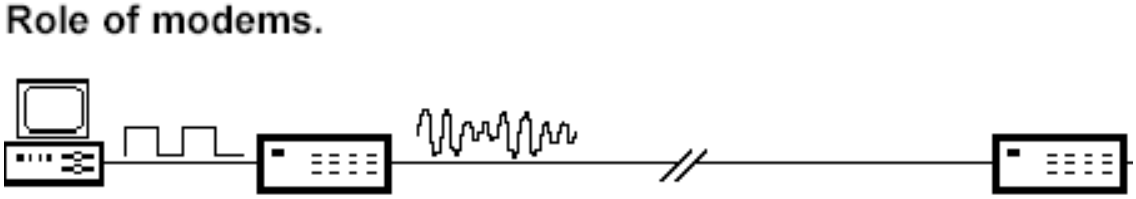


Figure: Role of modems

Three basic types of modulation are possible:

1. **Amplitude Modulation (AM).** In AM, the carrier signal's *amplitude* is changed according to the modulating digital signal's bit value. For example, two amplitude sizes (a small and a large one) may be used to, respectively, represent bit values 0 and 1. AM's main weakness is its susceptibility to distortion.
2. **Frequency Modulation (FM).** In FM, the carrier signal's *frequency* is changed according to the modulating digital signal's bit value. For example, two frequency values (a low and a high one) may be used to, respectively, represent bit values 0 and 1. FM is more resistant to distortion than AM.
3. **Phase Modulation (PM)**. In PM, the carrier signal's *phase* is changed according to the modulating digital signal's bit value. A change in the carrier signal's phase indicates a change in the modulating digital signal's bit value from 0 to 1 or from 1 to 0.

1.6.2.4 Digitization

Digitization is essentially the opposite of modulation. Whereas in modulation a digital signal is modulated over an analog signal for transmission, in digitization an analog signal is converted into digital format through a process of sampling. For example, the

analog signal resulting from human speech can be sampled and converted into digital data, transmitted over digital lines and converted back to analog signal at the other end. These two functions are performed by a device called **codec** (coder/decoder).

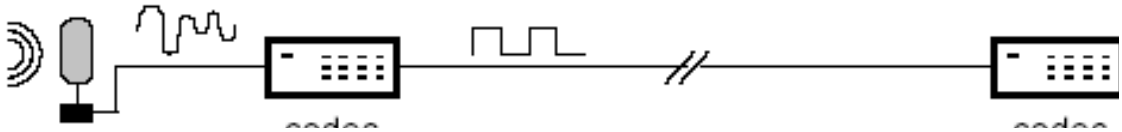


Figure: Role of Codec

It is worth noting that, unlike modulation (which is an exact process since the digital signal at the source and the digital signal received at the destination are identical), digitization is only an approximate process because of sampling. The process (of representing a continuous value with a discrete value) is called **quantization**. The relatively small loss of information inherent in the process is called **quantization error**.

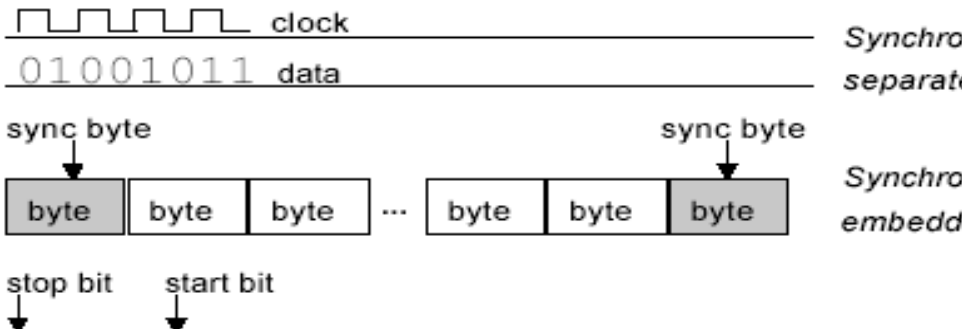
The coding process generates the sample data from the analog signal. The decoding process regenerates an approximation of the original signal by fitting a smooth curve to the sampled points. The quality of the regenerated signal can be improved by increasing the sampling rate (i.e., reducing the sampling interval), but up to a limit dictated by the Nyquist's theorem. This limit is exercised by a popular digitization technique called **Pulse Code Modulation** (PCM) which uses a sampling rate twice that of the original signal frequency. For example, a 4 kHz speech signal is sampled at a rate of 8000 samples per second.

The main advantage of digitization is that, due to its resistance to distortion, it is much easier to reliably transmit a digital signal over a long distance than an analog signal.

1.6.2.5 Synchronization

When two devices are about to communicate, the transmitter should somehow notify the receiver as to when to expect to receive data. This allows the receiver to prepare it for receiving the data. Furthermore, such notifications should occur frequently enough so that both devices maintain an agreement about the exact distribution of data over time. This process is called **synchronization**. There are two basic methods of synchronization:

- Synchronous transmission and
- Asynchronous transmission.



In **synchronous transmission**, a clock signal is used as a common source of reference by both the transmitter and the receiver. By tying the data signal to the clock signal, either device can look at the clock signal to know where data bits may begin or end. The clock signal may be provided on a separate line or be embedded in the data signal itself. Because having a separate clock line increases the costs, it is only used for covering very short distances (e.g., for connecting personal computers).

In **asynchronous transmission,** the beginning and end of each byte of data is marked by start and stop bits. This enables the receiver to work out the byte boundaries (see Figure) Because of its simplicity, asynchronous transmission is cheaper to implement and is therefore more widely used.

1.6.3 Transmission Media

Digital data can be transmitted over many different types of media. Selecting a transmission medium is guided by comparing transmission requirements against the medium's characteristics.

1.6.3.1 Factors affecting Choice of Media

Four important criteria influence the choice:

1. **Bandwidth.** Bandwidth is the maximum frequency range that can be practically supported by a medium. This is usually expressed in kilo Hz (kHz) or mega Hz (MHz). For example, analog transmission of human speech typically requires a bandwidth of 4 kHz. Also related, is the notion of **data rate**, which denotes the maximum number of bits per second (bps) that can be transmitted. For example, a data rate of 10 mbps means that 10 million bits of data can be transmitted in each second. Because of their obvious relationship, the terms bandwidth and data rate are sometimes used interchangeably. Because of distortion factors, bandwidth and data rate are usually inversely proportional to the communication distance.
2. **Cost.** Two types of cost are relevant: (i) the cost of installing the medium, including the medium-specific equipment that may be needed, and (ii) the cost of running and maintaining the medium and its equipment. There is usually a need for tradeoff between cost, bandwidth and distance.

3. Reliability.

Some media, by their physical nature, transmit data more reliably than others. Low reliability translates into a higher number of errors, which needs to be balanced against the potential cost of recovering from the errors (e.g., retransmission, more complex hardware and software).

4. Coverage.

The physical characteristics of a medium dictate how long a signal can travel in it before it is distorted beyond recognition. To cover larger areas, repeaters are needed to restore the signal, and this increases the costs.

1.6.3.2 Types of Transmission Media

Transmission media can be roughly grouped into following categories:

- Guided Media
- Unguided

1.6.4 Guided Transmission Media

Guided transmission media are those that provide a conduit from one device to another. Many different types of guided transmission media are there over which digital data can be transmitted such as Magnetic media, Copper wire and Fiber optics. **Unguided transmission media** or **wireless communications** are those that transport electromagnetic waves without using a physical conductor. We will discuss wireless communication in our next lesson. Here we discuss various guided media.

1.6.4.1 Magnetic Media

One of the most common ways to transport data from one computer to another is to write them onto magnetic tape or removable media, physically transport the tape or disks to the destination machine and read them back in again. Although this method is not as sophisticated as using a geosynchronous communication satellite, it is often more cost effective, especially for applications in which high bandwidth or cost per bit transported is the key factor.

For a bank with many gigabytes of data to be backed up daily on a second machine (so the bank can continue to function even in the face of a major flood or earthquake), it is likely that no other transmission technology can even begin to approach magnetic tape for performance. Of course, networks are getting faster, but tape densities are increasing, too.

1.6.4.2 Copper Wire.

This is the oldest form of electronic transmission medium. Its use dates back to the development of telegraph in the 1800s and earliest telephone systems. Early installations used open wires, but these were superseded by twisted pairs, which consist of a pair of insulated and twisted wires Twisted pairs are superior because of reduced crosstalk. They are very effective for relatively short distances (a few hundred

feet), but can be used for up to a few kilometers. A twisted pair has a bandwidth to distance ratio of about 1 MHz per kilometer. The performance of the twisted pair can be substantially improved by adding a metallic shield around the wires. Shielded wires are much more resistant to thermal noise and crosstalk effects. Twisted pairs used for long distance connections (e.g., telephone lines) are usually organized as a much larger cable containing numerous twisted pairs.

Although the bandwidth characteristics of magnetic tape are excellent, the delay characteristics are poor. Transmission time is measured in minutes or hours, not milliseconds. For many applications an on-line connection is needed. One of the oldest and still most common transmission media is **twisted pair.** A twisted pair consists of two insulated copper wires, typically about 1 mm thick. The wires are twisted together in a helical form, just like a DNA molecule. Twisting is done because two parallel wires constitute a fine antenna. When the wires are twisted, the waves from different twists cancel out, so the wire radiates less effectively.\

The most common application of the twisted pair is the telephone system. Nearly all telephones are connected to the telephone company (telco) office by a twisted pair. Twisted pairs can run several kilometers without amplification, but for longer distances, repeaters are needed. When many twisted pairs run in parallel for a substantial distance, such as all the wires coming from an apartment building to the telephone company office, they are bundled together and encased in a protective sheath. The pairs in these bundles would interfere with one another if it were not for the twisting. In parts of the world where telephone lines run on poles above ground, it is common to see bundles several centimeters in diameter.

Twisted pairs can be used for transmitting either analog or digital signals. The bandwidth depends on the thickness of the wire and the distance traveled, but several megabits/sec can be achieved for a few kilometers in many cases. Due to their adequate performance and low cost, twisted pairs are widely used and are likely to remain so for years to come.

Twisted pair cabling comes in several varieties, two of which are important for computer networks.

Unshielded Twisted-Pair {UTP) Cable

Unshielded twisted-pair (UTP) cable is the most common type of telecommunication medium in use today. Although most familiar from its use in telephone systems its frequency range is suitable for transmitting both data and voice. A twisted pair consists of two conductors (usually copper) each with its own colored plastic insulation. The plastic insulation is color-banded for identification. Colors are used both to identify the specific conductors in a cable and to indicate which wires belong in pairs

and how they relate to other pairs in a larger bundle. In the past, two parallel flat wires were used for communication. However, electromagnetic interference from devices such as a motor can create noise over those wires. If the two wires are parallel, the wire closest to the source of the noise gets more interference and ends up with a higher voltage level than the wire farther away, which results in an uneven load and a damaged signal.

If, however, the two wires are twisted around each other at regular intervals (between 2 and 12 twists per foot), each wire is closer to the noise source for half the time and farther away for the other half. With twisting, therefore, the cumulative effect of the interference is equal on both wires. Each section of wire has a "load" of 4 when it is on the top of the twist and 3 when it is on the bottom. The total effect of the noise at the receiver is therefore 0 (14 - 14). Twisting does not always eliminate the impact of noise, but it does significantly reduce it.

Advantages of UTP are its cost and ease of use. UTP is cheap, flexible, and easy to install. Higher grades of UTP are used in many LAN technologies; including Ethernet and Token Ring.

The Electronic Industries Association (EIA) has developed standards to grade UTP cables by quality. Categories are determined by cable quality, with 1 as the lowest and 5 as the highest. Each EIA category is suitable for certain uses and not for others:

Category 1: The basic twisted-pair cabling used in telephone systems. This level of quality is fine for voice but inadequate for all but low-speed data communication.

Category 2: The next higher grade, suitable for voice and for data transmission of up to 4 Mbps.

Category 3: Required to have at least three twists per fool and can be used for data transmission of up to 10 Mbps. It is now the standard cable for most telephone systems.

Category 4: Must also have at least three twists per foot as well as other conditions to bring the possible transmission rate to 16 Mbps.

Category 5: Used for data transmission up to 100 Mbps.

UTP Connectors UTP is most commonly connected to network devices via a type of snap-in plug like that used with telephone jacks. Connectors are either male (the plug) or female (the receptacle). Male connectors snap into female connectors and have a repressible tab (called a key) that locks them in place. Each wire in a cable is attached to one conductor (or pin) in the connector. The most frequently used of these plugs is an RJ45 connector with eight conductors, one for each wire of four twisted pairs.

Shielded Twisted-Pair (STP) Cable

Shielded twisted-pair (STP) cable has a metal foil or braided-mesh covering that encases each pair of insulated conductors. The metal casing prevents the penetration of electromagnetic noise. It also can eliminate a phenomenon called crosstalk, which is the undesired effect of one circuit (or channel) on another circuit (or channel). It

occurs when one line (acting as a kind of receiving antenna) picks up some of the signals traveling down another line (acting as a kind of sending antenna). This effect can be experienced during telephone conversations when one can hear other conversations in the background. Shielding each pair of a twisted-pair cable can eliminate most crosstalk.

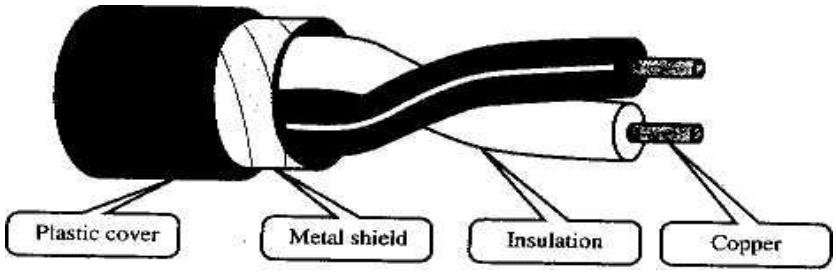


Figure: Shielded twisted-pair (STP) cable

STP has the same quality considerations and uses the same connectors as UTP, but the shield must be connected to a ground. Materials and manufacturing requirements make STP more expensive than UTP but less susceptible to noise.

1.6.4.3 Co-axial Cable

Another common transmission medium is the **coaxial cable.** It has better shielding than twisted pairs, so it can span longer distances at higher speeds. Two kinds of coaxial cable are widely used.

- 50-ohm cable, is commonly used when it is intended for digital transmission from the start.
- The other kind, 75-ohm cable, is commonly used for analog transmission and cable television but is becoming more important with the advent of Internet over cable. This distinction is based on historical, rather than technical, factors (e.g., early dipole antennas had an impedance of 300 ohms, and it was easy to use existing 4:1 impedance matching transformers).

A coaxial cable consists of a stiff copper wire as the core, surrounded by an insulating material. The insulator is encased by a cylindrical conductor, often as a closely-woven braided mesh. The outer conductor is covered in a protective plastic sheath. A view of a coaxial cable is shown in Fig.

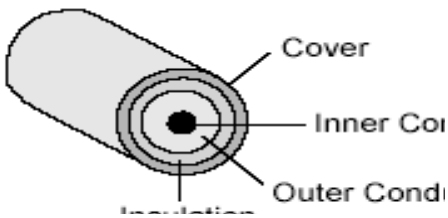


Figure: Co-axial cable

The construction and shielding of the coaxial cable give it a good combination of high bandwidth and excellent noise immunity. The bandwidth possible depends on the cable quality, length, and signal-to-noise ratio of the data signal. Modern cables have a bandwidth of close to 1 GHz. Coaxial cables used to be widely used within the telephone system for long-distance lines but have now largely been replaced by fiber optics on long-haul routes. Coax is still widely used for cable television and metropolitan area networks, however.

1.6.4.4 Fiber Optics

Many people in the computer industry take enormous pride in how fast compute technology is improving. In the race between computing and communication, communication won. The new conventional wisdom should be that all computers are hopelessly slow and that networks should try to avoid computation at all costs, no matter how much bandwidth that wastes. Here we will study fiber optics to see how that transmission technology works.

An optical transmission system has three key components: the light source, the transmission medium and the detector. Conventionally, a pulse of light indicates a 1 bit and the absence of light indicates a 0 bit. The transmission medium is an ultra-thin fiber of glass. The detector generates an electrical pulse when light falls on it. By attaching a light source to one end of an optical fiber and a detector to the other, we have a unidirectional data transmission system that accepts an electrical signal, converts and transmits it by light pulses and then reconverts the output to an electrical signal at the receiving end.

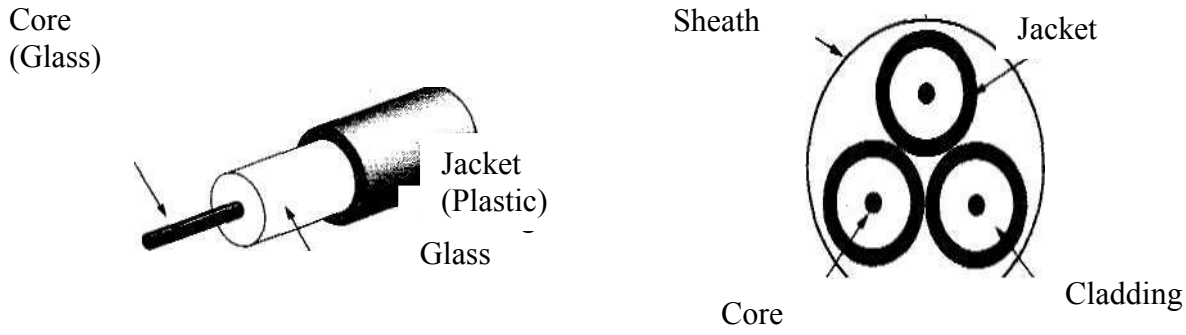


Figure (a): Side View of a Single Fiber (b): End View of a Sheath with 3 fibers

Fiber optic cables are similar to coax, except without the braid. Figure shows a single fiber viewed from the side. At the center is the glass core through which the light propagates. In multimode fibers, the core is typically 50 microns in diameter, about the thickness of a human hair. In single-mode fibers, the core is 8 to 10 microns. The core is surrounded by a glass cladding with a lower index of refraction than the core, to keep all the light in the core. Next comes a thin plastic jacket to protect the cladding. Fibers are typically grouped in bundles, protected by an outer sheath. Figure (b) shows a sheath with three fibers.

Terrestrial fiber sheaths are normally laid in the ground within a meter of the surface, where they are occasionally subject to attacks by backhoes or gophers. Near the shore, transoceanic fiber sheaths are buried in trenches by a kind of seaplow. In deep water, they just lie on the bottom, where they can be snagged by fishing trawlers or attacked by giant squid.

Fibers can be connected in three different ways:

First, they can terminate in connectors and be plugged into fiber sockets. Connectors lose about 10 to 20 percent of the light, but they make it easy to reconfigure systems.

- Second, they can be spliced mechanically. Mechanical splices just lay the two carefully-cut ends next to each other in a special sleeve and clamp them in place. Alignment can be improved by passing light through the junction and then making small adjustments to maximize the signal. Mechanical splices take trained personnel about 5 minutes and result in a 10 percent light loss.
- Third, two pieces of fiber can be fused (melted) to form a solid connection. A fusion splice is almost as good as a single drawn fiber, but even here, a small amount of attenuation occurs.

For all three kinds of splices, reflections can occur at the point of the splice, and the reflected energy can interfere with the signal.

Two kinds of light sources are typically used to do the signaling, LEDs (Light Emitting Diodes) and semiconductor lasers. They have different properties, as shown in Figure.

Characteristics	Semiconductor	LED
Distance	Long	Short
Lifetime	Short life	Long life
Data rate	High	Low
Fiber type	Multimode or single	Multimode
Temperature	Substantial	Minor
Cost	Expensive	Low cost

Figure: A comparison of semiconductor diodes and LEDs as light sources

The receiving end of an optical fiber consists of a photodiode, which gives off an electrical pulse when struck by light. The typical response time of a photo-diode is 1 nsec, which limits data rates to about 1 Gbps. Thermal noise is also an issue, so a pulse of light must carry enough energy to be detected. By making the pulses powerful enough, the error rate can be made arbitrarily small.

Fiber optics can be used for LANs as well as for long-haul transmission, although tapping into it is more complex than connecting to an Ethernet. One way around the problem is to realize that a ring network is really just a collection of point-to-point links. The interface at each computer passes the light pulse stream through to the next link and also serves as a T junction to allow the computer to send and accept messages.

1.6.5 Summary

A signal may be analog (continuous variation of some property) or digital (sequence of binary values 0 and 1). Digital data is transmitted over analog lines using modulation and converted back to digital format using demodulation. These two functions are performed by a modem. Modulation method is classified into AM, FM, and PM. Converting an analog signal into digital is called digitization and is performed by a codec. PCM is a popular digitization method for voice signals. Transmission methods are classified into synchronous (clock-based) and synchronous. Popular transmission media include: copper wire, coaxial cable, optical fiber, radio and infra-red.

A coaxial cable consists of four concentric cylinders: an inner conductor, surrounded by an insulating cylinder, surrounded by an outer conductor, surrounded by a final protective cover. This combination is called a coax. An optical fiber consists of two concentric cylinders: an inner core surrounded by a cladding. Both the core and the cladding are made of transparent plastic or glass material. The core is used for

guiding a light beam, whereas the cladding (which has a different refractive index) acts as a reflector to prevent the light from escaping from the core. Because optical fiber uses a light signal instead of electrons, it does not suffer from the various noise problems associated with electromagnetic signals. The signal is usually generated by a laser or Light Emitting Diode (LED). Optical fibers can provide bandwidth to distance ratios in order of 100s of MHz per kilometer. Like other cables, hundreds of optical fibers are usually housed within one cable. They are being increasingly used by telecommunication carriers for long distance digital trunk lines. Current trends promise that they will replace twisted pair residential loops in the near future.

1.6.6 Self-Assessment Questions

1. State the differences between an analog signal and a digital signal. Provide an example of either signal type.
2. What are the factors that influence the choice of transmission media? Discuss.
3. What are the differences between modulation and digitization? Name the devices that perform these functions.
4. Write short note on the following:
   a) Connection types b) Synchronization
   c) Magnetic Media d) Twisted Pair Cable
5. Discuss Fiber optic as a transmission media. .

1.6. **Multiple Choice Questions**

i) Transmission media are usually categorized as ----------.
   a) Fixed or unfixed
   b) Determinate or indeterminate
   c) Guided or unguided
   d) Metallic or nonmetallic

ii) In Fiber optics the signal source is----------.
   a) Radio
   b) Light
   c) Infrared
   d) Very low frequency

iii) Which of the following is not a guided medium?
   a) Twisted pair cable
   b) Co-axial cable
   c) Fiber optic cable
   d) Atmosphere

iv) The inner core of an optical fiber is ---------- in composition.
   a) Glass or plastic
   b) Copper
   c) Bimetallic

d) Liquid

v) if a satellite in geosynchronous orbit, it completes one orbit in----------.

a) One hour

b) 24 Hour

c) One month

d) One year

1.6.7 References and Suggested Readings

- Forouzan, B.A. 2003, " Data Communication and Networking", Tata McGraw Hill.
- Sinha, Pradeep. K. 2003, "Foundations of Computing", BPB Publications
- Tanenbaum, Andrew.S. 2003, "Computer Networks." Pearson Education Inc.
- Laudon, Kenneth, C. and Traver, carol, Guercio, 2003, "E-commerce", Pearson Education Inc.

LESSON NO. 1.7

AUTHOR : GURINDER GOSAL

WIRELESS TRANSMISSION

Structure of the Lesson:

Learning Objectives:

The major objectives of this lesson are to:

- Define various types of unguided media
- Discuss wireless transmission fundamentals.
- Have a broad understanding of the different wireless transmission media
- Discuss the characteristics of the different wireless transmission media.

1.7.1 Introduction

Guided transmission media are those that provide a conduit from one device to another. Different types of guided transmission media were discussed in our last lesson. **Unguided transmission media** or **wireless communications** are those that transport electromagnetic waves without using a physical conductor. We will discuss wireless communication in this lesson. Some people believe that the future holds only

two types of communication: fiber and wireless. All fixed or non-mobile computers, telephones, faxes etc. will use fiber and all mobile one will use wireless. Wireless has advantages even for fixed devices in some circumstances. Various wireless transmissions are prevalent. **Radio signals** have been used for a long time to transmit analog information. They are particularly attractive for long distance communication over difficult terrain or across the oceans, where the cost of installing cables can be too prohibitive.

An important form of microwave system is a **satellite** system, which is essentially a microwave system plus a large repeater in the sky. The signals transmitted by earth stations are received, amplified and retransmitted to other earth stations by the satellite. Another increasingly-popular form of radio is **cellular radio**, which is currently being used by carriers for providing mobile telephone networks. These operate in the VHF band and subdivide their coverage area into conceptual cells, where each cell represents a limited area which is served by a low-power transmitter and receiver station. As the mobile user moves from one cell area to another, its communication is handed over from one station to other. **Infra-red** signals are suitable for transmission over relatively short distances (the signal is easily reflected by hard objects). The signal is generated and received using optical transceivers.

1.7.2 Wireless Transmission

Radio waves are easy to generate, can travel long distances and can penetrate buildings easily, so they are widely used for communication, both indoors and outdoors. A minimum radio system consists of a transmitter and a receiver. It may operate at a variety of frequency bands, ranging from hundreds of Hz to hundreds of giga Hz (GHz). A huge range of transmission bandwidths are therefore possible. Microwave is by far the most widely used form of radio transmission. It operates in the GHz range with data rates in order of 100s of mbps per channel. Telecommunication carriers and TV stations are the primary users of microwave transmission. A satellite is one of the important forms of microwave systems and like other microwave systems, the bandwidth is subdivided into channels of 10s of MHz each, providing data rates in order of 100s of mbps. Because of their high bandwidths, satellites are capable of supporting an enormous number and variety of channels, including TV, telephone and data. The satellite itself, however, represents a major investment and typically has a limited lifetime (at most a few decades). Infra-red systems represent a cheap alternative to most other methods, because there is no cabling involved and the necessary equipment is relatively cheap. Data rates similar to those of twisted pairs are easily possible. However, applications are limited because of distance limitations (of about one kilometer). One recent use of infra-red has been for interfacing hand-held and portable computing devices to LANs.

1.7.3 Radio Transmission

Radio waves also are omni directional, meaning that they travel in all directions from the source, so the transmitter and receiver do not have to be carefully aligned physically. Sometimes omni directional radio is good, but sometimes it is bad. The properties of radio waves are frequency dependent. At low frequencies, radio waves pass through obstacles well, but the power falls off sharply with distance from the source, roughly as $1/r^2$ in air. At high frequencies, radio waves tend to travel in straight lines and bounce off obstacles. They are also absorbed by rain. At all frequencies, radio waves are subject to interference from motors and other electrical equipment.

Due to radio's ability to travel long distances, interference between users is a problem. For this reason, all governments tightly license the use of radio transmitters, with one exception, discussed below.

In the VLF, LF, and MF bands, radio waves follow the ground, as illustrated in Fig. (a). These waves can be detected for perhaps 1000 km at the lower frequencies, less at the higher ones.

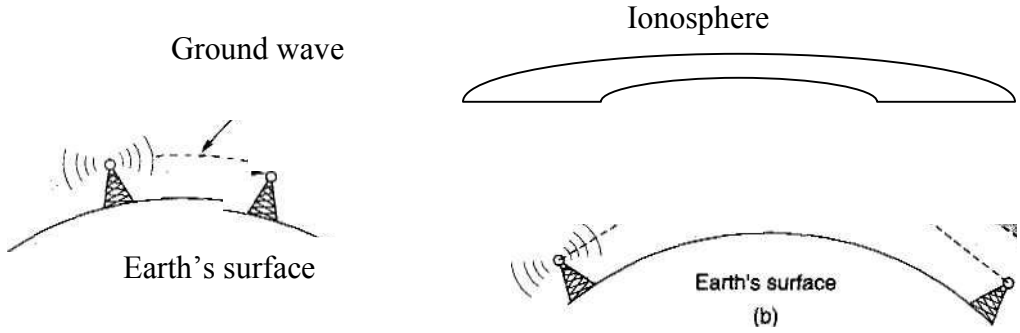


Figure a): In the VLF, LF, and MF bands, b) In the HF band, they -radio waves follow the curvature of the earth, bounce off the ionosphere.

In the HF and VHF bands, the ground waves tend to be absorbed by the earth. However, the waves that reach the ionosphere, a layer of charged particles circling the earth at a height of 100 to 500 km, are refracted by it and sent back to earth, as shown in Fig (b). Under certain atmospheric conditions, the signals can bounce several times. Amateur radio operators (hams) use these bands to talk long distance. The military also communicate in the HF and VHF bands.

1.7.3.1 Propagation of Radio Waves

1.7.3.1.1 Types of Propagation

Radio wave transmission utilizes five different types of propagation: surface, tropospheric, ionospheric, line-of-sight and space.

Radio technology considers the earth as surrounded by two layers of atmosphere: the troposphere and the ionosphere. The troposphere is the portion of the atmosphere extending outward approximately 30 miles from the earth's surface (in radio terminology, the troposphere includes the high-altitude layer called the stratosphere) and contains what we generally think of as air. Wind, temperature variations and weather in general occur in the troposphere, as does jet plane travel. The ionosphere is the layer of atmosphere above the troposphere but below space. It is beyond what we think of as atmosphere and contains free electrically charged particles (hence the name).

Surface Propagation

In surface propagation, radio waves travel through the lowest portion of the atmosphere, hugging the earth. At the lowest frequencies, signals emanate in all directions from the transmitting antenna and follow the curvature of the planet. Distance depends on the amount of power in the signal: the greater the power, the greater the distance. Surface propagation can also take place in seawater.

Tropospheric Propagation

Tropospheric propagation can work two ways. Either a signal can be directed in a straight line from antenna to antenna (line-of-sight), or it can be broadcast at an angle into the upper layers of the troposphere where it is reflected back down to the earth's surface. The first method requires that the placement of the receiver and the transmitter be within line-of-sight distances, limited by the curvature of the earth in relation to the height of the antennas. The second method allows greater distances to be covered.

Ionospheric Propagation

In ionospheric propagation, higher-frequency radio waves radiate upward into the ionosphere where they are reflected back to earth. The density difference between the troposphere and the ionosphere causes each radio wave to speed up and change direction, bending back to earth. This type of transmission allows for greater distances to be covered with lower power output.

Line-of-Sight Propagation

In line-of-sight propagation, very high frequency signals are transmitted in straight lines directly from antenna to antenna. Antennas must be directional, facing each other and either tall enough or close enough together not to be affected by the curvature of the earth. Line-of-sight propagation is tricky because radio transmissions cannot be completely focused. Waves emanate upward and downward as well as forward and can reflect off the surface of the earth or parts of the atmosphere. Reflected waves that arrive at the receiving antenna later than the direct portion of the transmission can corrupt the received signal.

Space Propagation

Space propagation utilizes satellite relays in place of atmospheric refraction. A broadcast signal is received by an orbiting satellite, which rebroadcasts the signal to the intended receiver back on the earth. Satellite transmission is basically line-of-sight with an intermediary (the satellite). The distance of the satellite from the earth makes it the equivalent of a super-high-gain antenna and dramatically increases the distance coverable by a signal.

1.7.3.1.2 Propagation of Specific Signals

The type of propagation used in radio transmission depends on the frequency (speed) of the signal. Each frequency is suited for a specific layer of the atmosphere and is most efficiently transmitted and received by technologies adapted to that layer.

VLF Very low frequency (VLF) waves are propagated as surface waves, usually through air but sometimes through seawater. VLF waves do not suffer much attenuation in transmission but are susceptible to the high levels of atmospheric noise (heat and electricity) active at low altitudes. VLF waves are used mostly for long-range radio navigation and for submarine communication.

LF Similar to VLF, low frequency (LF) waves are also propagated as surface waves. LF waves are used for long-range radio navigation and for radio beacons of navigational locators. Attenuation is greater during the daytime, when absorption of waves by natural obstacles increases.

MF Middle frequency (MF) signals are propagated in the troposphere. These frequencies are absorbed by the ionosphere. The distance they can cover is therefore limited by the angle needed to reflect the signal within the troposphere without entering the ionosphere. Absorption increases during the daytime, but most MF transmissions rely on line-of-sight antennas to increase control and avoid the absorption problem altogether. Uses for MF transmissions include AM radio, maritime radio, -radio direction finding (RDF), and emergency frequencies.

HF High frequency (HF) signals use ionospheric propagation. These frequencies move into the ionosphere, where the density difference reflects them back to earth. Uses for HF signals include amateur_radio (ham radio), citizen's band (CB) radio, international broadcasting, military communication, long-distance aircraft and ship communication, telephone, telegraph, and facsimile.

VHF Most very high frequency (VHF) waves use line-of-sight propagation. Uses for VHF include VHF television, FM radio, aircraft AM radio and aircraft navigational aid.

UHF Ultrahigh frequency (UHF) waves always use line-of-sight propagation. Uses for UHF include UHF television, mobile telephone, cellular radio, paging, and microwave links. Note that microwave communication begins at 1 GHz in the UHF band and continues into the SHF and EHF bands.

SHF Super high frequency (SHF) waves are transmitted using mostly line-of-sight and some space propagation. Uses for SHF include terrestrial and satellite microwave and radar communication.

EHF Extremely high frequency (EHF) waves use space propagation. Uses for EHF are predominantly scientific and include radar, satellite and experimental communications.

1.7.4 Microwave Transmission

Above 100 MHz, the waves travel in nearly straight lines and can therefore be narrowly focused. Concentrating all the energy into a small beam by means of a parabolic antenna (like the familiar satellite TV dish) gives a much higher signal-to-noise ratio, but the transmitting and receiving antennas must be accurately aligned with each other. In addition, this directionality allows multiple transmitters lined up in a row to communicate with multiple receivers in a row without interference, provided some minimum spacing rules are observed. Before fiber optics, for decades these microwaves formed the heart of the long-distance telephone transmission system. In fact, MCI, one of AT&T's first competitors after it was deregulated, built its entire system with microwave communications going from tower to tower tens of kilometers apart. Even the company's name reflected this (MCI stood for Microwave Communications, Inc.). MCI has since gone over to fiber and merged with WorldCom.

Since the microwaves travel in a straight line, if the towers are too far apart, the earth will get in the way. Consequently, repeaters are needed periodically. The higher the towers are, the farther apart they can be. The distance between repeaters goes up very roughly with the square root of the tower height. For 100-meter-high towers, repeaters can be spaced 80 km apart.

Unlike radio waves at lower frequencies, microwaves do not pass through buildings well. In addition, even though the beam may be well focused at the transmitter, there is still some divergence in space. Some waves may be refracted off low-lying atmospheric layers and may take slightly longer to arrive than the direct waves. The delayed waves may arrive out of phase with the direct wave and thus cancel the signal. This effect is called multipath fading and is often a serious problem. It is weather and frequency dependent. Some operators keep 10 percent of their channels idle as spares to switch on when multipath fading wipes out some frequency band temporarily.

The demand for more and more spectrum drives operators to yet higher frequencies. Bands up to 10 GHz are now in routine use, but at about 4 GHz a new problem sets in: absorption by water. These waves are only a few centimeters long and are absorbed by rain. This effect would be fine if one were planning to build a huge outdoor microwave oven for roasting passing birds, but for communication, it is a

severe problem. As with multipath fading, the only solution is to shut off links that are being rained on and route around them.

Microwave communication is so widely used for long-distance telephone communication, mobile phones, television distribution, and other uses that a severe shortage of spectrum has developed. It has several significant advantages over fiber. The main one is that no right of way is needed and by buying a small plot of ground every 50 km and putting a microwave tower on it, one can bypass the telephone system and communicate directly. This is how MCI managed to get started as a new long-distance telephone company so quickly.

Microwave is also relatively inexpensive. Putting up two simple towers (may be just big poles with four guy wires) and putting antennas on each one may be cheaper than burying 50 km of fiber through a congested urban area or up over a mountain, and it may also be cheaper than leasing the telephone company's fiber, especially if the telephone company has not yet even fully paid for the copper it ripped out when it put in the fiber.

1.7.5 Infrared Waves Transmission

Unguided infrared and millimeter waves are widely used for short-range communication. The remote controls used on televisions, VCRs, and stereos all use infrared communication. They are relatively directional, cheap, and easy to build but have a major drawback: they do not pass through solid objects (try standing between your remote control and your television and see if it still works). In general, as we go from long-wave radio toward visible light, the waves behave more and more like light and less and less like radio.

On the other hand, the fact that infrared waves do not pass through solid walls well is also a plus. It means that an infrared system in one room of a building will not interfere with a similar system in adjacent rooms or buildings: you cannot control your neighbor's television with your remote control. Furthermore, security of infrared systems against eavesdropping is better than that of radio systems precisely for this reason. Therefore, no government license is needed to operate an infrared system in contrast to radio systems, which must be licensed outside the ISM bands. Infrared communication has a limited use on the desktop, for example, connecting notebook computers and printers but it is not a major player in the communication game.

1.7.6 Lightwave Transmission

Unguided optical signaling has been in use for centuries. A modern application is to connect the LANs in two buildings via lasers mounted on their rooftops. Coherent optical signaling using lasers is inherently unidirectional, so each building needs its own laser and its own photodetector. This scheme offers very high bandwidth and very low cost. It is also relatively easy to install and unlike microwave, does not require an FCC license.

The laser's strength, a very narrow beam, is also its weakness here. Aiming a laser beam 1-mm wide at a target the size of a pin head 500 meters away requires the marksmanship. Usually, lenses are put into the system to defocus the beam slightly. A disadvantage is that laser beams cannot penetrate rain or thick fog, but they normally work well on sunny days.

1.7.7 Satellite Communication Transmission

Satellite transmission is much like line-of-sight microwave transmission in which one of the stations is a satellite orbiting the earth. The principle is the same as terrestrial microwave, with a satellite acting as a super tall antenna and repeater. Although in satellite transmission signals must still travel in straight lines, the limitations imposed on distance by the curvature of the earth are reduced. In this way, satellite relays allow microwave signals to span continents and oceans with a single bounce.

Satellite microwave can provide transmission capability to and from any location on earth, no matter how remote. This advantage makes high-quality communication available to undeveloped parts of the world without requiring a huge investment in ground-based infrastructure. Satellites themselves are extremely expensive, of course, but leasing time or frequencies on one can be relatively cheap.

1.7.7.1 Geosynchronous Satellites

Line-of-sight propagation requires that the sending and receiving antennas be locked onto each other's location at all times (one antenna must have the other in sight). For this reason, a satellite that moves faster or slower than the earth's rotation is useful only for short periods of time (just as a stopped clock is accurate twice a day). To ensure constant communication, the satellite must move at the same speed as the earth so that it seems to remain fixed above a certain spot. Such satellites are called geosynchronous.

Because orbital speed is based on distance from the planet, only one orbit can be geosynchronous. This orbit occurs at the equatorial plane and is approximately 22,000 miles from the surface of the earth.

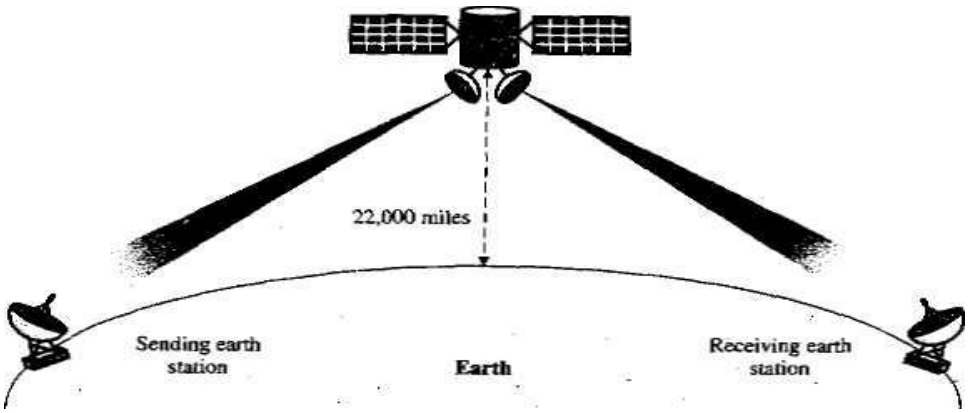


Figure: Satellite Communication

But one geosynchronous satellite cannot cover the whole earth. One satellite in orbit has line-of-sight contact with a vast number of stations, but the curvature of the earth still keeps much of the planet out of sight. It takes a minimum of three satellites equidistant from each other in geosynchronous orbit to provide full global transmission. Figure shows three satellites, each 120 degrees from another in geosynchronous orbit around the equator. The view is from the North Pole

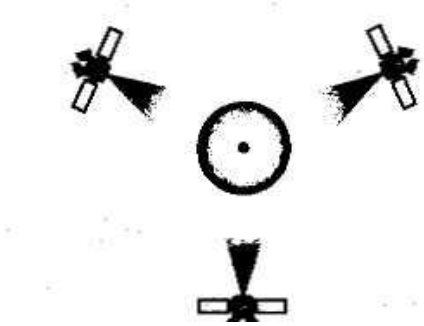


Figure: Satellites in geosynchronous orbits

1.7.8 Cellular Telephony

Cellular telephony is designed to provide stable communications connections between two moving devices or between one mobile unit and one stationary (land) unit. A service provider must be able to locate and track a caller, assign a channel to the call, and transfer the signal from channel to channel as the caller moves out of the range of one channel and into the range of another.

To make this tracking possible, each cellular service area is divided into small regions called cells. Each cell contains an antenna and is controlled by a small office called the cell office. Each cell office, in turn, is controlled by a switching office called a mobile telephone switching office (MTSO). The MTSO coordinates communication between all of the cell offices and the telephone central office. It is a computerized center that is responsible for connecting calls as well as recording call information and billing. Cell size is not fixed and can be increased or decreased depending on the population of the area. The typical radius of a cell is 1 to 12 miles. High-density areas require more, geographically smaller cells to meet traffic demands than do lower density areas. Once determined, cell size is optimized to prevent the interference of adjacent cell signals. The transmission power of each cell is kept low to prevent its signal from interfering with those of other cells.

1.7.8.1 Cellular Bands

Traditional cellular transmission is analog. To minimize noise, frequency modulation (FM) is used for communication between the mobile telephone itself and the cell office. The FCC has assigned two bands for cellular use. The band between 824 and 849 MHz carries those communications that initiate from mobile phones. The band

between 869 and 894 MHz carries those communications that initiate from land phones. Carrier frequencies are spaced every 30 KHz, allowing each band to support to 833 carriers. However, two carriers are required for full-duplex communication, which doubles the required width of each channel to 60 KHz and leaves only 416 channels available for each band.

Each band, therefore, is divided into 416 FM channels (for a total of 832 channels). Of these, some are reserved for control and setup data rather than voice communication. In addition, to prevent interference, channels are distributed among the cells in such a way that adjacent cells do not use the same channels. This restriction means that each cell normally has access to only 40 channels.

1.7.8.2 Transmitting

To place a call from a mobile phone, the caller enters a code of 7 or 10 digits (a phone number) and presses the send button. The mobile phone then scans the band, seeking a setup channel with a strong signal, and sends the data (phone number) to the closest cell office using that channel. The cell office relays the data to the MTSO. The MTSO sends the data on to the telephone central office. If the called party is available, a connection is made and the result is relayed back to the MTSO. At this point, the MTSO assigns an unused voice channel to the call and a connection is established. The mobile phone automatically adjusts its tuning to the new channel and voice communication can begin.

1.7.8.3 Receiving

When a land phone places a call to a mobile phone, the telephone central office sends the number to the MTSO. The MTSO searches for the location of the mobile phone by sending query signals to each cell in a process called paging. Once the mobile phone is found, the MTSO transmits a ringing signal and when the mobile phone is answered, assigns a voice channel to the call, allowing voice communication to begin.

1.7.8.4 Handoff

It may happen that, during a conversation, the mobile phone moves from one cell to another. When it does, the signal may become weak. To solve this problem, the MTSO monitors the level of the signal every few seconds. If the strength of the signal diminishes, the MTSO seeks a new cell that can accommodate the communication better. The MTSO then changes the channel carrying the call (hands the signal off from the old channel to a new one). Handoffs are performed so smoothly that most of the time they are transparent to the users.

1.7.8.5 Digital

Analog (FM) cellular services are based on a standard called analog circuit switched cellular (ACSC). To transmit digital data using an ACSC service requires a modem with a maximum speed of 9600 to 19,200 bps.

Since 1993, however, several service providers have been moving to a cellular data standard called cellular digital packet data (CDPD). CDPD provides low-speed digital service over the existing cellular network. It is based on the OSI model.

To use the existing digital services, such as 56K switched service, CDPD uses what is called a trisector. A trisector is a combination of three cells each using 19.2 Kbps, for a total of 57.6 Kbps.

1.7.8.6 Integration with Satellites and PCs

Cellular telephony is moving fast toward integrating the existing system with satellite communication. This integration will make it possible to have mobile communication between any two points on the globe. Another goal is to combine cellular telephony and personal computer communication under a scheme called mobile personal communication to enable people to use small, mobile personal computers to send and receive data, voice, image and video.

1.7.9 Summary

Unguided media or wireless communication transport electromagnetic waves without using a physical conductor. Instead, signals are broadcast through air (or, in a few cases, water), and thus are available to anyone who has a device capable of receiving them.

Radio signals have been used for a long time to transmit analog information. They are particularly attractive for long distance communication over difficult terrain or across the oceans, where the cost of installing cables can be too prohibitive.

Microwaves do not follow the curvature of the earth and therefore require line-of-sight transmission and reception equipment. The distance coverable by a line-of-sight signal depends to a large extent on the height of the antenna: the taller the antennas, the longer the sight distance.

Satellite transmission is much like line-of-sight microwave transmission in which one of the stations is a satellite orbiting the earth. The principle is the same as terrestrial microwave, with a satellite acting as a super tall antenna and repeater. To ensure constant communication, the satellite must move at the same speed as the earth so that it seems to remain fixed above a certain spot. Such satellites are called **geosynchronous**.

Cellular telephony is designed to provide stable communications connections between two moving devices or between one mobile unit and one stationary (land) unit

1.7.10 Self-Assessment Questions

6. What is wireless transmission? Discuss radio transmission in detail.
7. What are the methods used to propagate radio waves?
8. Describe the layers of the atmosphere. What types of radio communication utilizes each?

9. What is satellite communication? Describe geosynchronous satellites functioning?
10. Write a note on cellular telephony.

6. Multiple Choice Questions

i) If a satellite is in geosynchronous orbit, its distance from the sending station ---
   a) Is constant
   b) Varies according to the radius of the orbit
   c) Varies according to the time of a day
   d) None of the above

ii) The radio communication spectrum is divided into bands based on ----------
   a) Frequency
   b) Amplitude
   c) Cost and Hardware
   d) Transmission medium

iii) If a satellite is in geosynchronous orbit, it completes one orbit in ----------.
   a) 24 hours
   b) One hour
   c) One month
   d) One year

iv) When we talk about unguided media, usually we are referring to ----------.
   a) Nonmetallic wires
   b) Metallic wires
   c) The atmosphere
   d) None of the above

1.7.11 References and Suggested Readings

- Forouzan, B.A. 2003, " Data Communication and Networking", Tata McGraw Hill.
- Sinha, Pradeep. K. 2003, "Foundations of Computing", BPB Publications
- Tanenbaum, Andrew.S. 2003, "Computer Networks." Pearson Education Inc.
- Laudon, Kenneth, C. and Traver, carol, Guercio, 2003, "E-commerce", Pearson Education Inc.

LESSON NO. 1.8 | **AUTHOR : GURINDER GOSAL**

SWITCHING TECHNIQUES

Structure of the Lesson:

Learning Objectives:

The major objectives of this lesson are to:

- Discuss switching.
- Describe the general switching technologies available.
- Understand the underlying principles of circuit switching and packet switching.
- Discuss the generic methods of packet switching.
- Compare the circuit switching and packet switching on various parameters.

1.8.1 Introduction

Switching is the generic method for establishing a path for point-to-point communication in a network. Many characteristics of switched communication networks are directly dependent on how data is relayed over the wires. Early networks carried continuous bitstreams over physical links in a technique called circuit switching, well suited to transmit voice or real time data from a single sender to a single receiver (unicast communication). However, a physical link failure in circuit switching networks has dramatic consequences leading to the interruption of all communications using the failed link. Datagram packet switching networks like the Internet fix these drawbacks by cutting data into small chunks called *packets*. In datagram packet switching networks, two consecutive packets from the same communication are independently handled by the network. Therefore, when a link fails, packets previously sent on the failed link can be *rerouted* to avoid the failed link

and communications are not interrupted. Datagram packet switching networks are said to be *resilient* to link failures because link failures are hidden to end-users. On the other hand, it is more difficult to manage end-to-end flows of data in datagram packet switching networks than in circuit switching networks due to the lack of a separate circuit for each flow.

1.8.2 Switching Technologies

The different switching techniques used in networks are: circuit switching, message switching, datagram packet switching and virtual circuit packet switching. These are separately discussed below:

1.8.3 Circuit Switching

In circuit switching, a caller must first establish a connection to a callee before any communication is possible. During the connection establishment, resources are allocated between the caller and the callee.

In circuit switching, two communicating stations are connected by a *dedicated* communication path which consists of intermediate nodes in the network and the links that connect these nodes. What is significant about circuit switching is that the communication path remains intact for the duration of the connection, engaging the nodes and the links involved in the path for that period. (However, these nodes and links are typically capable of supporting many channels, so only a portion of their capacity is taken away by the circuit.).

Figure1 shows a simple circuit switch which consists of a 3×3 matrix, capable of connecting any of its inlets (*a*, *b*, and *c*) to any of its outlets (*d*, *e*, and *f*).

Each cross-point appears as a circle. A hollow circle means that the cross-point is *off* (i.e., the two crossing wires are not connected). A solid circles means that the cross-point is *on* (i.e., the crossing wires are connected). The switch can support up to three simultaneous but independent connections. (Although we have used an equal number of inlets and outlets here, in general, this need not be the case. Switches may also have more inlets than outlets, or more outlets than inlets.)

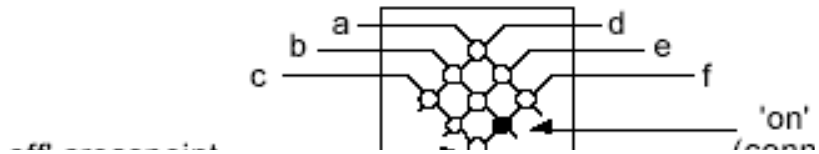


Figure1: A Simple Circuit Switch

Figure 2 show a simple circuit-switched network built using the switch in Figure1. When the two hosts shown in the figure initiate a connection, the network determines a path through the intermediate switches and establishes a circuit which is maintained for the duration of the connection. When the hosts disconnect, the network releases the circuit.

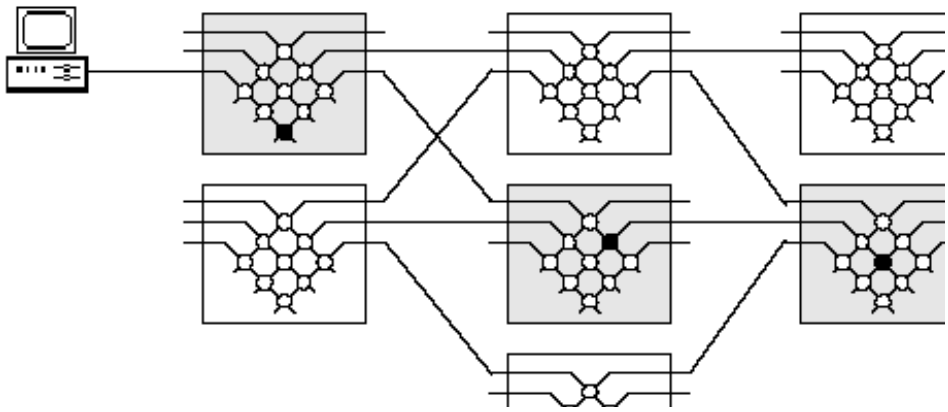


Figure2: Circuit Switching using Switch of Figure1

In circuit switching, resources remain allocated during the full length of a communication, after a circuit is established and until the circuit is terminated and the allocated resources are freed. Resources remain allocated even if no data is flowing on a circuit, hereby wasting link capacity when a circuit does not carry as much traffic as the allocation permits. This is a major issue since frequencies (in FDM) or time slots (in TDM) are available in finite quantity on each link, and establishing a circuit consumes one of these frequencies or slots on each link of the circuit. As a result, establishing circuits for communications that carry less traffic than allocation permits can lead to resource exhaustion and network saturation, preventing further connections from being established. If no circuit can be established between a sender and a receiver because of a lack of resources, the connection is *blocked*.

A second characteristic of circuit switching is the time cost involved when establishing a connection. In a communication network, circuit-switched or not, nodes need to lookup in a *forwarding table* to determine on which link to send incoming data, and to actually send data from the input link to the output link. Performing a lookup in a forwarding table and sending the data on an incoming link is called *forwarding*. Building the forwarding tables is called *routing*. In circuit switching, routing must be performed for each communication, at circuit establishment time. During circuit establishment, the set of switches and links on the path between the sender and the receiver is determined and messages are exchanged on all the links between the two end hosts of the communication in order to make the resource allocation and build the routing tables. In circuit switching, forwarding tables are hardwired or implemented using fast hardware, making data forwarding at each switch almost instantaneous. Therefore, circuit switching is well suited for long-lasting connections where the initial circuit establishment time cost is balanced by the low forwarding time cost. The circuit identifier (a range of frequencies in FDM or a time slot position in a TDM frame) is changed by each switch at forwarding time so that switches do not

need to have a complete knowledge of all circuits established in the network but rather only local knowledge of available identifiers at a link. Using local identifiers instead of global identifiers for circuits also enables networks to handle a larger number of circuits.

Circuit switching relies on dedicated equipment especially built for the purpose, and is the dominant form of switching in telephone networks. Its main advantage lies in its predictable behavior: because it uses a dedicated circuit, it can offer a constant throughput with no noticeable delay in transfer of data. This property is important in telephone networks, where even a short delay in voice traffic can have disruptive effects.

Circuit switching's main weakness is its inflexibility in dealing with computer oriented data. A circuit uses a fixed amount of bandwidth, regardless of whether it is used or not. In case of voice traffic, the bandwidth is usually well used because most of the time one of the two parties in a telephone conversation is speaking. However, computers behave differently; they tend to go through long silent periods followed by a sudden burst of data transfer. This leads to significant underutilization of circuit bandwidth.

On the other hand, circuit switching networks are not reactive when a network topology change occurs. For instance, on a link failure, all circuits on a failed link are cut and communication is interrupted. Special mechanisms that handle such topological changes have been be devised. Traffic engineering can alleviate the consequences of a link failure by pre-planning failure recovery. A backup circuit can be established at the same time or after the primary circuit used for a communication is set up, and traffic can be rerouted from the failed circuit to the backup circuit if a link of the primary circuit fails. Circuit switching networks are intrinsically sensitive to link failures and rerouting must be performed by additional traffic engineering mechanisms.

Another disadvantage of circuit switching is that the network is only capable of supporting a limited number of simultaneous circuits. When this limit is reached, the network blocks further attempts for connection until some of the existing circuits are released.

1.8.4 Message Switching

An alternative switching strategy is **message switching**. When this form of switching is used, no physical path is established in advance between sender and receiver. Instead, when the sender has a block of data to be sent, it is stored in the first switching office (i.e., router) and then forwarded later, one hop at a time. Each block is received in its entirety, inspected for errors, and then retransmitted. A network using this technique is called a **store-and-forward** network.

The first electromechanical telecommunication systems used message switching, namely, for telegrams. The message was punched on paper tape (off-line) at the sending office, and then read in and transmitted over a communication line to the next office along the way, where it was punched out on paper tape. An operator there tore the tape off and read it in on one of the many tape readers, one reader per outgoing trunk. Such a switching office was called a torn tape office**.** Paper tape is long gone and message switching is not used any more.

1.8.5 Packet Switching

The alternative to circuit switching is packet switching. Conceived in the 1960's, *packet switching* is a more recent technology than circuit switching which addresses a disadvantage of circuit switching: the need to allocate resources for a circuit, thus incurring link capacity wastes when no data flows on a circuit. Packet switching introduces the idea of cutting data on a flow into packets which are transmitted over a network without any resource being allocated. If no data is available at the sender at some point during a communication, then no packet is transmitted over the network and no resources are wasted.

Packet switching was designed to address the shortcomings of circuit switching in dealing with data communication. Unlike circuit switching where communication is continuous along a dedicated circuit, in packet switching, communication is discrete in form of packets. Each packet is of a limited size and can hold up to a certain number of octets of user data. Larger messages are broken into smaller chunks so that they can be fitted into packets. In addition to user data, each packet carries additional information (in form of a header) to enable the network to route it to its final destination. A packet-switched network has a much higher capacity for accepting further connections. Additional connections are usually not blocked but simply slow down existing connections, because they increase the overall number of packets in the network and hence increase the delivery time of each packet.

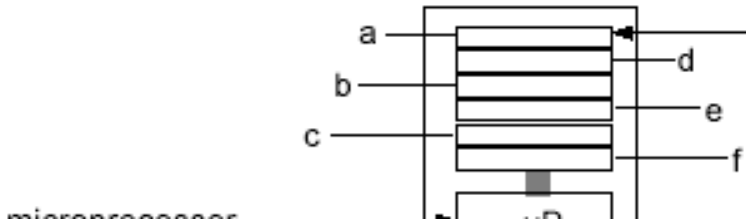


Figure3: A simple packet switch with six I/O channels (*a* through *f*).

Figure4 shows a simple packet switch with six I/O channels (*a* through *f*). Each channel has an associated buffer which it uses to store packets in transit. The operation of the switch is controlled by a microprocessor. A packet received on any of

the channels can be passed onto any of the other channels by the microprocessor moving it to the corresponding buffer.

Packet switching is the generic name for a set of two different techniques: datagram packet switching and virtual circuit packet switching.

1.8.5.1 Datagram Packet Switching

Different from circuit switching, datagram packet switching (also known as **connectionless**) does not require establishing circuits prior to transmission of data and terminating circuits after the transmission of data. The switches, called routers, have to make a lookup in the forwarding table, called *routing table*, for each incoming packet. A routing table contains a mapping between the possible final destinations of packets and the outgoing link on their path to the destination. Routing tables can be very large because they are indexed by possible destinations, making lookups and routing decisions computationally expensive and the full forwarding process relatively slow compared to circuit switching. In datagram packet switching networks, each packet must carry the address of the destination host and use the destination address to make a forwarding decision. Consequently, routers do not need to modify the destination addresses of packets when forwarding packets.

Since each packet is processed individually by a router, all packets sent by a host to another host are not guaranteed to use the same physical links. If the routing algorithm decides to change the routing tables of the network between the instants two packets are sent, then these packets will take different paths and can even arrive out of order. In Figure illustrates the datagram method. Note how the packets exercise different routes.

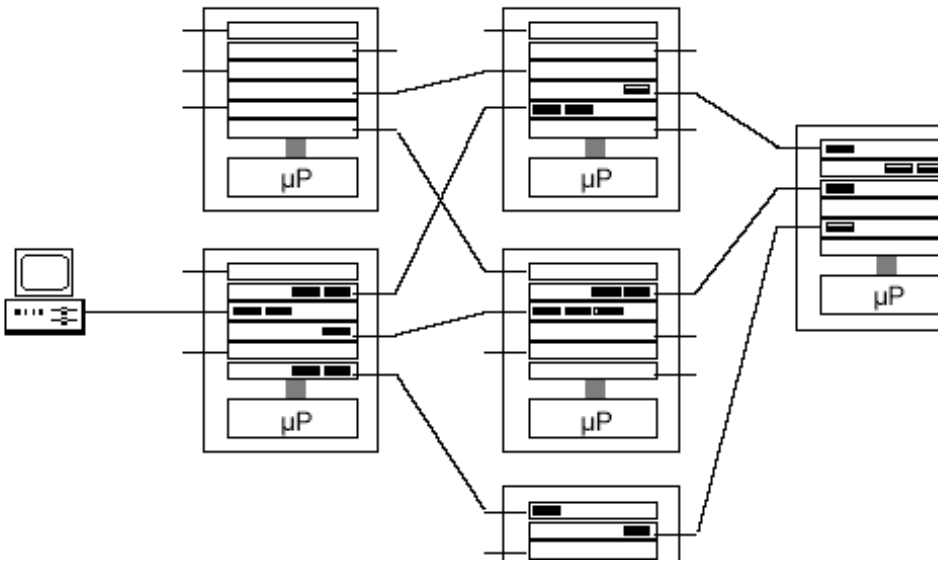


Figure4: A Datagram Packet Switching using the switch in Figure3

Second, on a network topology change such as a link failure, the routing protocol will automatically recomputed routing tables so as to take the new topology into account and avoid the failed link. As opposed to circuit switching, no additional traffic engineering algorithm is required to reroute traffic. Since routers make routing decisions locally for each packet, independently of the flow to which a packet belongs. Therefore, traffic engineering techniques, which heavily rely on controlling the route of traffic, are more difficult to implement with datagram packet switching than with circuit switching.

The advantage of the datagram approach is that because there is no circuit, congestion and faulty nodes can be avoided by choosing a different route. Also, connections can be established more quickly because of reduced overheads. This makes datagrams better suited than virtual circuits for brief connections. For example, database transactions in banking systems are of this nature, where each transaction involves only a few packets.

1.8.5.2 Virtual Circuit Packet Switching

Virtual circuit packet switching (VC-switching) is a packet switching technique which merges datagram packet switching and circuit switching to extract both of their advantages. VC-switching is a variation of datagram packet switching where packets flow on so-called logical circuits for which no physical resources like frequencies or time slots are allocated. Each packet carries a circuit identifier which is local to a link and updated by each switch on the path of the packet from its source to its destination. A virtual circuit is defined by the sequence of the mappings between a link taken by packets and the circuit identifier packets carry on this link. This sequence is set up at connection establishment time and identifiers are reclaimed during the circuit termination.

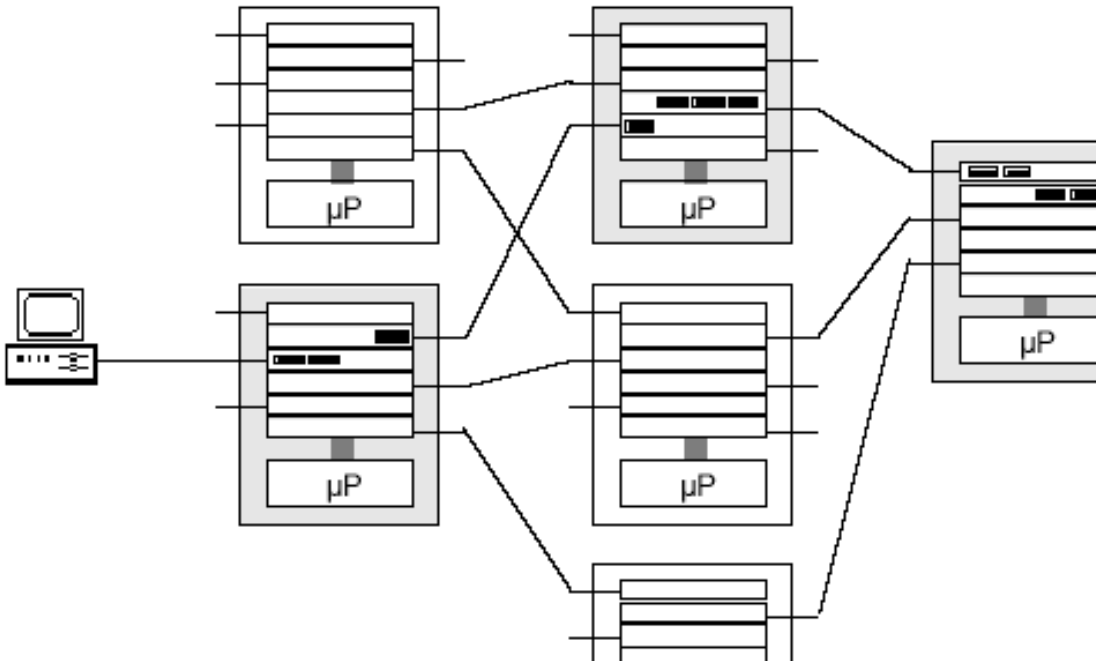


Figure5: A Virtual Circuit Packet Switching using the switch in Figure3

When the two hosts initiate a connection, the network layer establishes a virtual circuit (denoted by shaded switches) which is maintained for the duration of the connection. When the hosts disconnect, the network layer releases the circuit. The packets in transit are displayed as dark boxes within the buffers. These packets travel only along the designated virtual circuit.

We have seen the trade-off between connection establishment and forwarding time costs that exists in circuit switching and datagram packet switching. In VC-switching, routing is performed at circuit establishment time to keep packet forwarding fast. Other advantages of VC-switching include the traffic engineering capability of circuit switching, and the resources usage efficiency of datagram packet switching. Nevertheless, a main issue of VC-Switched networks is the behavior on a topology change. As opposed to Datagram Packet Switched networks which automatically recompute routing tables on a topology change like a link failure, in VC-switching all virtual circuits that pass through a failed link are interrupted. Hence, rerouting in VC-switching relies on traffic engineering techniques.

In practice, major implementations of VC-switching are X.25, Asynchronous Transfer Mode (ATM) and Multiprotocol Label Switching (MPLS). The Internet, today's most used computer network, is entirely built around the Internet Protocol (IP), which is responsible for routing packets from one host to another. Because of the central role of IP in the Internet, we now discuss how ATM and MPLS interact with IP.

The advantage of the virtual circuit approach is that because no separate routing is required for each packet, they are likely to reach their destination more quickly; this leads to improved throughput. Furthermore, packets always arrive in order. Virtual circuits are better suited to long connections that involve the transfer of large amounts of data (e.g., transfer of large files). Because packet switching is the more dominant form of switching for data communication, we will focus our attention on this form of switching from now on.

1.8.6 Comparison of Circuit-Switching and Packet-Switching Networks

Circuit switching and packet switching differ in many respects. To start with, circuit switching requires that a circuit be set up end to end before communication begins. Packet switching does not require any advance setup. The first packet can just be sent as soon as it is available.

The result of the connection setup with circuit switching is the reservation of bandwidth all the way from the sender to the receiver. All packets follow this path. Among other properties, having all packets follow the same path means that they cannot arrive out of order. With packet switching there is no path, so different packets can follow different paths, depending on network conditions at the time they are sent. They may arrive out of order.

Packet switching is more fault tolerant than circuit switching. In fact, that is why it was invented. If a switch goes down, all of the circuits using it are terminated and no more traffic can be sent on any of them. With packet switching, packets can be routed around dead switches. Setting up a path in advance also opens up the possibility of reserving bandwidth in advance. If bandwidth is reserved, then when a packet arrives, it can be sent out immediately over the reserved bandwidth. With packet switching, no bandwidth is reserved, so packets may have to wait their turn to be forwarded.

Having bandwidth reserved in advance means that no congestion can occur when a packet shows up (unless more packets show up than expected). On the other hand, when an attempt is made to establish a circuit, the attempt can fail due to congestion. Thus, congestion can occur at different times with circuit switching (at setup time) and packet switching (when packets are sent). If a circuit has been reserved for a particular user and there is no traffic to send, the bandwidth of that circuit is wasted. It cannot be used for other traffic. Packet switching does not waste bandwidth and thus is more efficient from a system-wide perspective. Understanding this trade-off is crucial for comprehending the difference between circuit switching and packet switching. The trade-off is between guaranteed service and wasting resources versus not guaranteeing service and not wasting resources.

Packet switching uses store-and-forward transmission. A packet is accumulated in a router's memory, then sent on to the next router. With circuit

switching, the bits just flow through the wire continuously. The store-and-forward technique adds delay.

Another difference is that circuit switching is completely transparent. The sender and receiver can use any bit rate, format, or framing method they want to. The carrier does not know or care. With packet switching, the carrier determines the basic parameters. A rough analogy is a road versus a railroad. In the former, the user determines the size, speed, and nature of the vehicle; in the latter, the carrier does. It is this transparency that allows voice, data, and fax to coexist within the phone system.

A final difference between circuit and packet switching is the charging algorithm. With circuit switching, charging has historically been based on distance and time. For mobile phones, distance usually does not play a role, except for international calls, and time plays only a minor role (e.g., a calling plan with 2000 free minutes costs more than one with 1000 free minutes and sometimes night or weekend calls are cheaper than normal). With packet switching, connect time is not an issue, but the volume of traffic sometimes is. For home users, ISPs usually charge a flat monthly rate because it is less work for them and their customers can understand this model easily, but backbone carriers charge regional networks based on the volume of their traffic. The differences are summarized in Table1.

Description	Circuit switched	Packet switched
Call setup	Required	Not needed
Dedicated physical	Yes	No
Each packet follows	Yes	No
Packets arrive in order	Yes	No
Is a switch crash fatal	Yes	No
Bandwidth available	Fixed	Dynamic
Time of possible	At setup time	On every packet
Potentially wasted	Yes	No
Transparency	Yes	No
Charging	Per minute	Per packet

Table1: A comparison of circuit-switched and packet-switched networks.

1.8.7 Summary

One of the most important functions of the network layer is to employ the switching capability of the nodes in order to route messages across the network. Switching involves the nodes in the network utilizing their direct communication lines to other nodes so that a path is established in a piecewise fashion. Each node has the capability to 'switch' to a neighboring node to further stretch the path until it is completed. *Circuit switching* is the transmission technology that has been used since

the first communication networks in the nineteenth century. In circuit switching, two communicating stations are connected by a *dedicated* communication path which consists of intermediate nodes in the network and the links that connect these nodes. In **message switching**, no physical path is established in advance between sender and receiver and message switching is not used any more. In packet switching technology, individual packets are sent as need be, with no dedicated path being set up in advance. It is up to each packet to find its way to the destination on its own. The **datagram** method (also known as **connectionless**) does not rely on a pre established route, instead each packet is treated independently. Therefore, it is possible for different packets to travel along different routes in the network to reach the same final destination. As a result, packets may arrive out of order, or even never arrive (due to node failure). It is up to the network user to deal with lost packets, and to rearrange packets to their original order. Because of the absence of a pre established circuit, each packet must carry enough information in its header to enable the nodes to route it correctly. The **virtual circuit** method (also known as **connection-oriented**) is closer to circuit switching. Here a complete route is worked out prior to sending data packets. The route is established by sending a connection request packet along the route to the intended destination. This packet informs the intermediate nodes about the connection and the established route so that they will know how to route subsequent packets. The result is a circuit somewhat similar to those in circuit switching, except that it uses packets as its basic unit of communication. Hence it is called a virtual circuit.

Each packet carries a virtual circuit identifier which enables a node to determine to which virtual circuit it belongs and hence how it should be handled. (The virtual circuit identifier is essential because multiple virtual circuits may pass through the same node at the same time.) Because the route is fixed for the duration of the call, the nodes spend no effort in determining how to route packets.

1.8.8 Self-Assessment Questions

11. Discuss the concept of switching as it relates to the problems involved in the connection of devices.
12. What are the different switching methods? Discuss briefly.
13. Which is more efficient, circuit switching or virtual circuit switching? Why?
14. What are the two popular approaches of packet switching?
15. Compare circuit switching and packet switching on various parameters.

6. **Multiple Choice Questions**

i) Which type of switching uses entire capacity of a dedicated link?

a) Circuit switching b) Datagram Packet switching
c) Virtual circuit packet switching d) Message switching

ii) In which type of switching do all the datagrams of a message follow the same channels of a path?

a) Circuit switching b) Message switching

c) Virtual circuit packet switching d) Datagram Packet switching

iii) A switched virtual circuit involves----------.

a) Connection establishment

b) Connection Release

c) Data transfer

d) All of the above

iv) In ----------, each packet of a message need not follow the same channels of a path?

a) Circuit switching

b) Message switching

c) Virtual circuit approach of packet switching

d) Datagram approach Packet switching

1.8.9 References and Suggested Readings

- Forouzan, B.A. 2003, " Data Communication and Networking", Tata McGraw Hill.
- Sinha, Pradeep. K. 2003, "Foundations of Computing", BPB Publications
- Tanenbaum, Andrew.S. 2003, "Computer Networks." Pearson Education Inc.
- Laudon, Kenneth, C. and Traver, carol, Guercio, 2003, "E-commerce", Pearson Education Inc.