UNIT-1
INTRODUCTION

NETWORKS:

A network is the interconnection of a set of devices capable of communication. In this definition, a device can be a host (or an end system as it is sometimes called) such as a large computer, desktop, laptop, workstation, cellular phone, or security system.

A device in this definition can also be a connecting device such as a router, which connects the network to other networks, a switch, which connects devices together, a modem (modulator-demodulator), which changes the form of data, and so on.

These devices in a network are connected using wired or wireless transmission media such as cable or air. When we connect two computers at home using a plug-and-play router, we have created a network, although very small.

NETWORK CRITERIA:

A network must be able to meet a certain number of criteria. The most important of these are ***performance***, ***reliability***, and ***security***.

Performance:

Performance can be measured in many ways, including ***transit time*** and ***response time***.

1. **Transit time** is the amount of time required for a message to travel from one device to another.

2. **Response time** is the elapsed time between an inquiry and a response.

The performance of a network depends on a number of factors, including the number of users, the type of transmission medium, the capabilities of the connected hardware, and the efficiency of the software.

Performance is often evaluated by two networking metrics: ***throughput*** and ***delay***. We often need more throughputs and less delay. However, these two criteria are often contradictory. If we try to send more data to the network, we may increase throughput but we increase the delay because of traffic congestion in the network.

Reliability: In addition to accuracy of delivery, network reliability is measured by the frequency of failure, the time it takes a link to recover from a failure, and the network's robustness in a catastrophe.

Security: Network security issues include protecting data from unauthorized access, protecting data from damage and development, and implementing policies and procedures for recovery from breaches and data losses.

Physical Structures:

Before discussing networks, we need to define some network attributes.

Type of Connection:

A network is two or more devices connected through links. A link is a communications pathway that transfers data from one device to another. For visualization purposes, it is simplest to imagine any link as a line drawn between two points. For communication to occur, two devices must be connected in some way to the same link at the same time.

There are two possible types of connections: ***point-to-point*** and ***multipoint***.

Point-to-Point: A point-to-point connection provides a dedicated link between two devices. The entire capacity of the link is reserved for transmission between those two devices. Most point-to-point connections use an actual length of wire or cable to connect the two ends, but other options, such as microwave or satellite links, are also possible (see Figure 1.1a). When we change television channels by infrared remote control, we are establishing a point-to-point connection between the remote control and the television's control system.

Multipoint: A multipoint **(al*so called multidrop***) connection is one in which more than two specific devices share a single link (see Figure 1.1b).

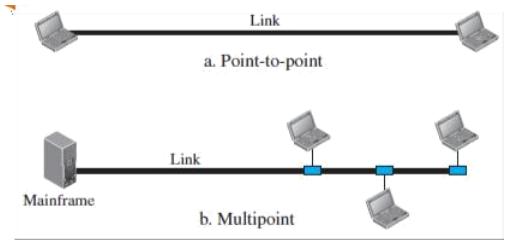


FIGURE 1.1 TYPES OF CONNECTIONS: POINT-TO-POINT AND MULTIPOINT

In a multipoint environment, the capacity of the channel is shared, either spatially or temporally. If several devices can use the link simultaneously, it is a spatially shared connection. If users must take turns, it is a timeshared connection.

Physical Topology:

The term ***physical topology*** refers to the way in which a network is laid out physically. Two or more devices connect to a link; two or more links form a

topology. The topology of a network is the geometric representation of the relationship of all the links and linking devices to one another.

There are four basic topologies possible: ***mesh***, ***star***, ***bus***, and ***ring***.

Mesh Topology:

In a **mesh topology,** every device has a dedicated point-to-point link to every other device. The term *dedicated* means that the link carries traffic only between the two devices it connects.

To find the number of physical links in a fully connected mesh network with *n* nodes, we first consider that each node must be connected to every other node. Node 1 must be connected to *n* – 1 nodes, node 2 must be connected to *n* – 1 nodes, and finally node *n* must be connected to *n* – 1 nodes. We need *n* (*n* – 1) physical links.

However, if each physical link allows communication in both directions (duplex mode) we can divide the number of links by 2. In other words, we can say that in a mesh topology, we need ***n* (*n* – 1)** / **2** duplex-mode links. To accommodate that many links, every device on the network must have *n* – 1 input/output (I/O) ports (see Figure 1.2) to be connected to the other *n* – 1 stations.

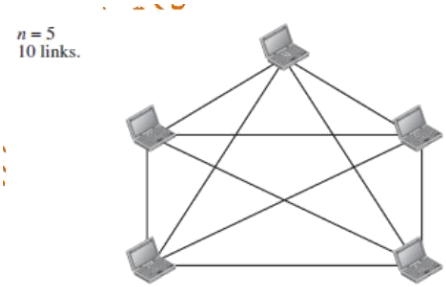


FIGURE 1.2: A FULLY CONNECTED MESH TOPOLOGY (FIVE DEVICES)

A mesh offers several *advantages* over other network topologies.

1. First, the use of dedicated links guarantees that each connection can carry its own data load, thus eliminating the traffic problems that can occur when links must be shared by multiple devices.

2. Second, a mesh topology is robust. If one link becomes unusable, it does not incapacitate the entire system.

3. Third, there is the advantage of privacy or security. When every message travels along a dedicated line, only the intended recipient sees it. Physical boundaries prevent other users from gaining access to messages.

4. Finally, point-to-point links make fault identification and fault isolation easy. Traffic can be routed to avoid links with suspected problems. This facility enables the network manager to discover the precise location of the fault and aids in finding its cause and solution.

The main *disadvantages* of a mesh are related to the amount of cabling and the number of I/O ports required.

1. First, because every device must be connected to every other device, installation and reconnection are difficult.
2. Second, the sheer bulk of the wiring can be greater than the available space (in walls, ceilings, or floors) can accommodate.
3. Finally, the hardware required to connect each link (I/O ports and cable) can be prohibitively expensive.

For these reasons a mesh topology is usually implemented in a limited fashion, for example, as a backbone connecting the main computers of a hybrid network that can include several other topologies.

Star Topology:

In a **star topology,** each device has a dedicated point-to-point link only to a central controller, usually called a ***hub*****.** The devices are not directly linked to one another. Unlike a mesh topology, a star topology does not allow direct traffic between devices. The controller acts as an exchange: If one device wants to send data to another, it sends the data to the controller, which then relays the data to the other connected device (see Figure 1.3).

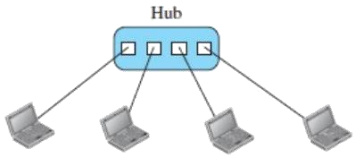


FIGURE 1.3: A STAR TOPOLOGY CONNECTING FOUR STATIONS

A star topology is *less expensive than* a mesh topology. In a star, each device needs only one link and one I/O port to connect it to any number of others. This factor also makes it easy to install and reconfigure. Far less cabling needs to be housed, and additions, moves, and deletions involve only one connection: between that device and the hub.

Other *advantages* include robustness. If one link fails, only that link is affected. All other links remain active. This factor also lends itself to easy fault identification and fault isolation. As long as the hub is working, it can be used to monitor link problems and bypass defective links.

One big *disadvantage* of a star topology is the dependency of the whole topology on one single point, the hub. If the hub goes down, the whole system is dead. Although a star requires far less cable than a mesh, each node must be linked to a central hub. For this reason, often more cabling is required in a star than in some other topologies (such as ring or bus).

The star topology is used in local-area networks (LANs). High-speed LANs often use a star topology with a central hub.

Bus Topology:

The preceding examples all describe point-to-point connections. A **bus topology,** on the other hand, is multipoint. One long cable acts as a **backbone** to link all the devices in a network (see Figure 1.4).

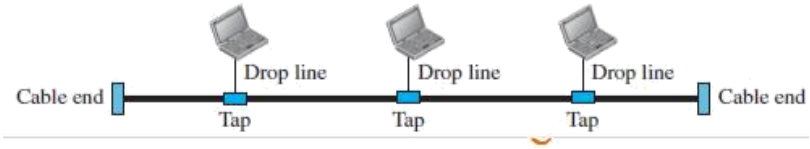


FIGURE 1.4: A BUS TOPOLOGY CONNECTING THREE STATIONS

Nodes are connected to the bus cable by drop lines and taps. A drop line is a connection running between the device and the main cable. A tap is a connector that either splices into the main cable or punctures the sheathing of a cable to create a contact with the metallic core.

As a signal travels along the backbone, some of its energy is transformed into heat. Therefore, it becomes weaker and weaker as it travels farther and farther. For this reason there is a limit on the number of taps a bus can support and on the distance between those taps.

Advantages of a bus topology include ease of installation. Backbone cable can be laid along the most efficient path, and then connected to the nodes by drop lines of various lengths. In this way, a bus uses less cabling than mesh or star topologies.

Disadvantages include difficult reconnection and fault isolation. A bus is usually designed to be optimally efficient at installation. It can therefore be difficult to add new devices. Signal reflection at the taps can cause degradation in quality. This degradation can be controlled by limiting the number and spacing of devices connected to a given length of cable. Adding new devices may therefore require modification or replacement of the backbone.

Bus topology was the *one of the first topologies* used in the design of early local area networks.

Ring Topology:

In a **ring topology,** each device has a dedicated point-to-point connection with only the two devices on either side of it. A signal is passed along the ring in one direction, from device to device, until it reaches its destination. Each device in the ring incorporates a repeater. When a device receives a signal intended for another device, its repeater regenerates the bits and passes them along (see Figure 1.5).

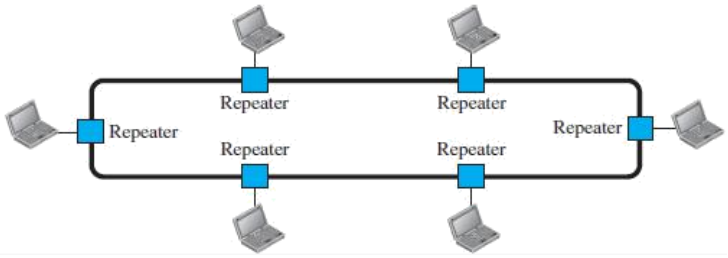


FIGURE 1.5: A RING TOPOLOGY CONNECTING SIX STATIONS

A ring is *relatively easy to install and reconfigure*. Each device is linked to only its immediate neighbors (either physically or logically). To add or delete a device requires changing only two connections. The only constraints are media and traffic considerations (maximum ring length and number of devices). In addition, fault isolation is simplified.

Generally, in a ring a signal is circulating at all times. If one device does not receive a signal within a specified period it can issue an alarm. The alarm alerts the network operator to the problem and its location.

However, unidirectional traffic can be a *disadvantage*. In a simple ring, a break in the ring (such as a disabled station) can disable the entire network. This weakness can be solved by using a dual ring or a switch capable of closing off the break.

Ring topology was prevalent when IBM introduced its local-area network, Token Ring. Today, the need for higher-speed LANs has made this topology less popular.

NETWORK TYPES:

We use a few criteria such as size, geographical coverage, and ownership to make this distinction. After discussing two types of networks, LANs and WANs, we define switching, which is used to connect networks to form an internetwork (a network of networks).

LOCAL AREA NETWORK:

A **local area network** (**LAN**) is usually privately owned and connects some hosts in a single office, building, or campus. Each host in a LAN has an identifier, an

address that uniquely defines the host in the LAN. A packet sent by a host to another host carries both the source host's and the destination host's addresses.

In the past, all hosts in a network were connected through a common cable, which meant that a packet sent from one host to another was received by all hosts. The intended recipient kept the packet; the others dropped the packet.

Today, most LANs use a smart connecting switch, which is able to recognize the destination address of the packet and guide the packet to its destination without sending it to all other hosts.

The switch alleviates the traffic in the LAN and allows more than one pair to communicate with each other at the same time if there is no common source and destination among them. Figure 1.6 shows a LAN using either a common cable or a switch.

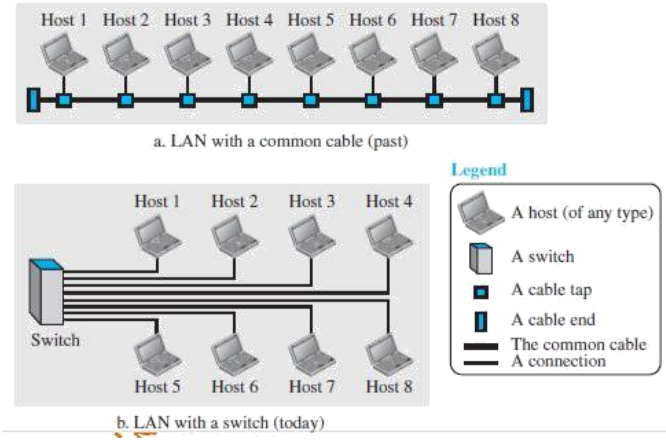


FIGURE 1.6: AN ISOLATED LAN IN THE PAST AND TODAY

WIDE AREA NETWORK:

A **wide area network (WAN)** is also an interconnection of devices capable of communication. However, there are some differences between a LAN and a WAN. A LAN is normally limited in size, spanning an office, a building, or a campus; a WAN has a wider geographical span, spanning a town, a state, a country, or even the world.

A LAN interconnects hosts; a WAN interconnects connecting devices such as switches, routers, or modems. A LAN is normally privately owned by the organization that uses it; a WAN is normally created and run by communication companies and leased by an organization that uses it. We see two distinct examples of WANs today: point-to-point WANs and switched WANs.

Point-to-Point WAN: A point-to-point WAN is a network that connects two communicating devices through a transmission media (cable or air). Figure 1.7 shows an example of a point-to-point WAN.

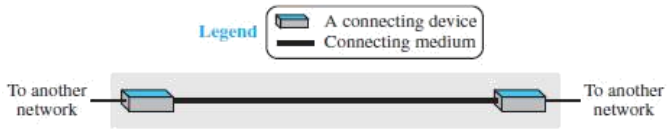


FIGURE 1.7: A POINT-TO-POINT WAN

Switched WAN: A switched WAN is a network with more than two ends. We can say that a switched WAN is a combination of several point-to-point WANs that are connected by switches. Figure 1.8 shows an example of a switched WAN.

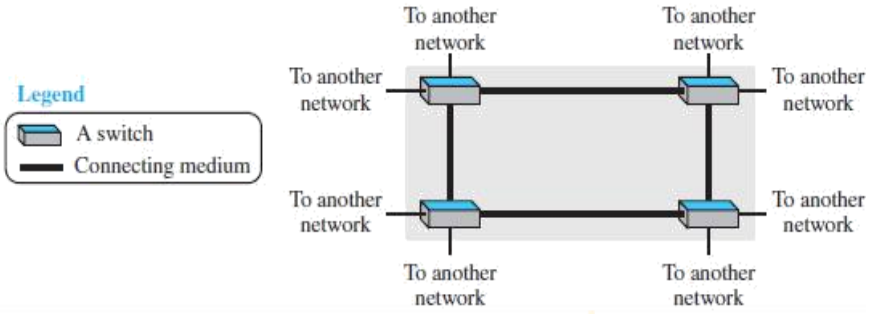


FIGURE 1.8: A SWITCHED WAN

INTERNETWORK:

Today, it is very rare to see a LAN or a WAN in isolation; they are connected to one another. When two or more networks are connected, they make an **internetwork,** or **internet.**

As an example, assume that an organization has two offices, one on the east coast and the other on the west coast. Now the company has an internetwork, or a private internet (with lowercase *i*). Communication between offices is now possible. Figure 1.9 shows this internet.

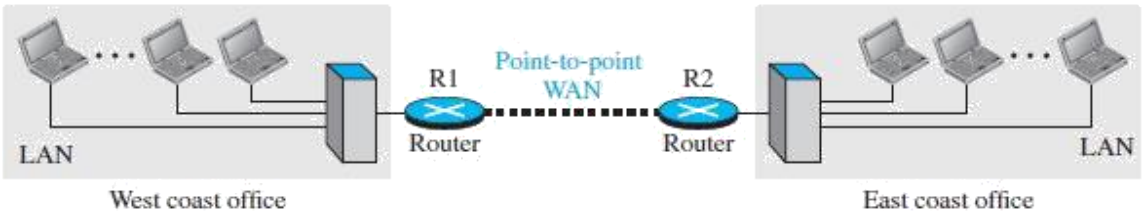


FIGURE 1.9: AN INTERNETWORK MADE OF TWO LANS AND ONE POINT-TO-POINT WAN

When a host in the west coast office sends a message to another host in the same office, the router blocks the message, but the switch directs the message to the destination.

SWITCHING:

An internet is a **switched network** in which a switch connects at least two links together. A switch needs to forward data from a network to another network when required. The two most common types of switched networks are ***circuit-switched*** and ***packet-switched networks***.

Circuit-Switched Network:

In a **circuit-switched network,** a dedicated connection, called a circuit, is always available between the two end systems; the switch can only make it active or inactive.

Figure 1.10 shows a very simple switched network that connects four telephones to each end. We have used telephone sets instead of computers as an end system because circuit switching was very common in telephone networks in the past, although part of the telephone network today is a packet-switched network.

In Figure 1.10, the four telephones at each side are connected to a switch. The switch connects a telephone set at one side to a telephone set at the other side. The thick line connecting two switches is a high-capacity communication line that can handle four voice communications at the same time; the capacity can be shared between all pairs of telephone sets. The switches used in this example have forwarding tasks but no storing capability.

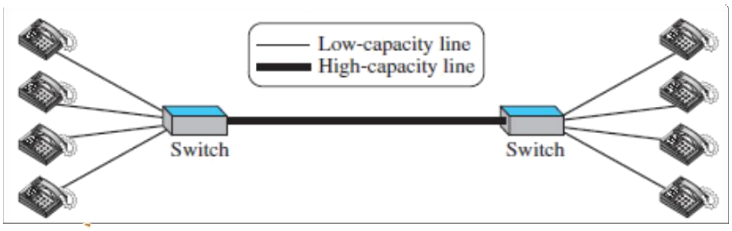


FIGURE 1.10: A CIRCUIT-SWITCHED NETWORK

Packet-Switched Network:

In a computer network, the communication between the two ends is done in blocks of data called **packets.** In other words, instead of the continuous communication we see between two telephone sets when they are being used, we see the exchange of individual data packets between the two computers.

This allows us to make the switches function for both storing and forwarding because a packet is an independent entity that can be stored and sent later. Figure 1.11 shows a small packet-switched network that connects four computers at one site to four computers at the other site.

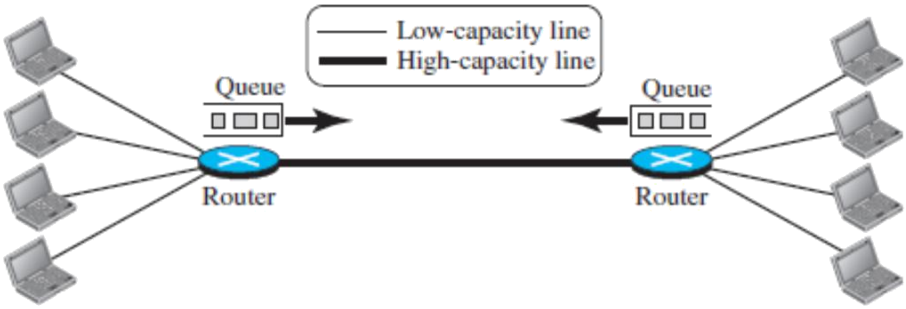


FIGURE 1.11: A PACKET-SWITCHED NETWORK

A router in a packet-switched network has a queue that can store and forward the packet. Now assume that the capacity of the thick line is only twice the capacity of the data line connecting the computers to the routers. If only two computers (one at each site) need to communicate with each other, there is no waiting for the packets.

However, if packets arrive at one router when the thick line is already working at its full capacity, the packets should be stored and forwarded in the order they arrived.

THE INTERNET: An internet (*note the lowercase i*) is two or more networks that can communicate with each other. The most notable internet is called the **Internet** (*uppercase I*), and is composed **of t**housands of interconnected networks.

Internet contains several backbones, provider networks, and customer networks.

At the top level, the *backbones* are large networks owned by some communication companies such as Sprint, Verizon (MCI), AT&T, and NTT. The backbone networks are connected through some complex switching systems, called *peering points*.

At the second level, there are smaller networks, called *provider networks* that use the services of the backbones for a fee. The provider networks are connected to backbones and sometimes to other provider networks.

The *customer networks* are networks at the edge of the Internet that actually use the services provided by the Internet. They pay fees to provider networks for receiving services.

Backbones and provider networks are also called **Internet Service Providers (ISPs).** The backbones are often referred to as *international ISPs;* the provider networks are often referred to as *national* or *regional ISPs*.

INTERNET HISTORY:

Early History:

There were some communication networks, such as telegraph and telephone networks, before 1960. These networks were suitable for constant-rate communication at that time, which means that after a connection was made between two users, the encoded message (telegraphy) or voice (telephony) could be exchanged.

A computer network, on the other hand, should be able to handle *bursty* data, which means data received at variable rates at different times. The world needed to wait for the packet-switched network to be invented.

Birth of Packet-Switched Networks: The theory of packet switching for bursty traffic was first presented by Leonard Kleinrock in 1961 at MIT. At the same time, two other researchers, Paul Baran at Rand Institute and Donald Davies at National Physical Laboratory in England, published some papers about packet-switched networks.

ARPANET: In the mid-1960s, mainframe computers in research organizations were stand-alone devices. Computers from different manufacturers were unable to communicate with one another.

The **Advanced Research Projects Agency (ARPA)** in the Department of Defense (DOD) was interested in finding a way to connect computers so that the researchers they funded could share their findings, thereby reducing costs and eliminating duplication of effort.

In 1967, at an Association for Computing Machinery (ACM) meeting, ARPA presented its ideas for the **Advanced Research Projects Agency Network (ARPANET),** a small network **of** connected computers.

The idea was that each host computer (not necessarily from the same manufacturer) would be attached to a specialized computer, called an *interface message processor* (IMP). The IMPs, in turn, would be connected to each other. Each IMP had to be able to communicate with other IMPs as well as with its own attached host. By 1969, ARPANET was a reality.

BIRTH OF THE INTERNET:

In 1972, Vint Cerf and Bob Kahn, both of whom were part of the core ARPANET group, collaborated on what they called the *Internetting Project*. They wanted to link dissimilar networks so that a host on one network could communicate with a host on another. There were many problems to overcome: diverse packet sizes, diverse interfaces, and diverse transmission rates, as well as differing reliability requirements. Cerf and Kahn devised the idea of a device called a *gateway* to serve as the intermediary hardware to transfer data from one network to another.

TCP/IP: Cerf and Kahn's landmark 1973 paper outlined the protocols to achieve end-to-end delivery of data. This was a new version of NCP (Network Control Protocol). This paper on transmission control protocol (TCP) included concepts such as encapsulation, the datagram, and the functions of a gateway.

A radical idea was the transfer of responsibility for error correction from the IMP to the host machine. This ARPA Internet now became the focus of the communication effort. Around this time, responsibility for the ARPANET was handed over to the Defense Communication Agency (DCA).

In October 1977, an internet consisting of three different networks (ARPANET, packet radio, and packet satellite) was successfully demonstrated. Communication between networks was now possible. TCP splits into two protocols: **Transmission Control Protocol (TCP)** and **Internet Protocol (IP).**

IP would handle datagram routing while TCP would be responsible for higher level functions such as segmentation, reassembly, and error detection. The new combination became known as TCP/IP.

MILNET: In 1983, ARPANET split into two networks: **Military Network (MILNET)** for military users and ARPANET for nonmilitary users.

CSNET: Another milestone in Internet history was the creation of CSNET in 1981. **Computer Science Network (CSNET)** was a network sponsored by the National Science Foundation (NSF). CSNET was a less expensive network; there were no redundant links and the transmission rate was slower.

NSFNET: With the success of CSNET, the NSF in 1986 sponsored the **National Science Foundation Network (NSFNET),** a backbone that connected five supercomputer centers located throughout the United States.

ANSNET: In 1991, the U.S. government decided that NSFNET was not capable of supporting the rapidly increasing Internet traffic. Three companies, IBM, Merit, and Verizon, filled the void by forming a nonprofit organization called Advanced Network & Services (ANS) to build a new, high-speed Internet backbone called **Advanced Network Services Network (ANSNET).**

Internet Today: Today, we witness a rapid growth both in the infrastructure and new applications. The Internet today is a set of pier networks that provide services to the whole world. What has made the Internet so popular is the invention of new applications.

World Wide Web: The 1990s saw the explosion of Internet applications due to the emergence of the World Wide Web (WWW). The Web was invented at CERN by Tim Berners-Lee. This invention has added the commercial applications to the Internet.

Multimedia: Recent developments in the multimedia applications such as voice over IP (telephony), video over IP (Skype), view sharing (YouTube), and television over IP (PPLive) has increased the number of users and the amount of time each user spends on the network.

Peer-to-Peer Applications: Peer-to-peer networking is also a new area of communication with a lot of potential.

STANDARDS AND ADMINISTRATION:

Internet standards:

An **Internet standard** is a thoroughly tested specification that is useful to and adhered to by those who work with the Internet. It is a formalized regulation that must be followed. A specification begins as an Internet draft.

An **Internet draft** is a working document (a work in progress) with no official status and a six-month lifetime. Upon recommendation from the Internet authorities, a draft may be published as a **Request for Comment (RFC).**

Each RFC is edited, assigned a number, and made available to all interested parties. RFCs go through maturity levels and are categorized according to their requirement level.

Maturity Levels:

An RFC, during its lifetime,falls into one of six *maturity levels:* proposed standard, draft standard, Internet standard, historic, experimental, and informational (see Figure 1.13)***.***

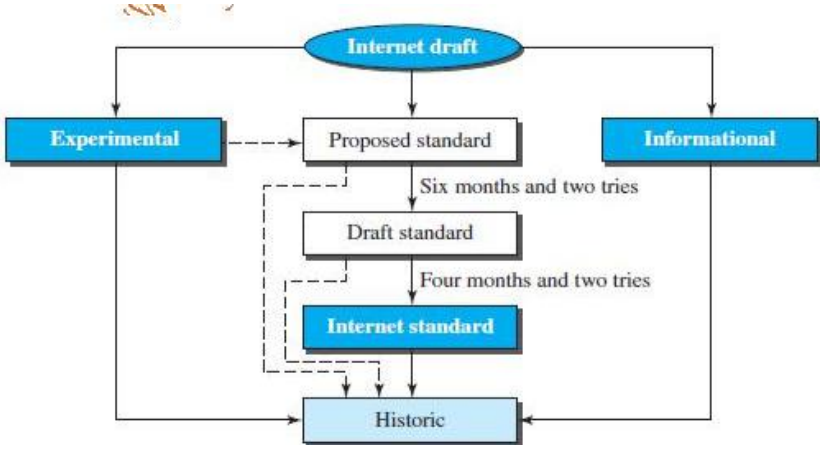


FIGURE 1.13: MATURITY LEVELS OF AN RFC

Proposed Standard: At this level, the specification is usually tested and implemented by several different groups.

Draft Standard: A proposed standard is elevated to draft standard status after at least two successful independent and interoperable

implementations. Barring difficulties, a draft standard, with modifications if specific problems are encountered, normally becomes an Internet standard.

Internet Standard: A draft standard reaches Internet standard status after demonstrations of successful implementation.

Historic: The historic RFCs are significant from a historical perspective. They either have been superseded by later specifications or have never passed the necessary maturity levels to become an Internet standard.

Experimental: An RFC classified as experimental describes work related to an experimental situation that does not affect the operation of the Internet. Such an RFC should not be implemented in any functional Internet service.

Informational: An RFC classified as informational contains general, historical, or tutorial information related to the Internet. It is usually written by someone in a non-Internet organization, such as a vendor.

Requirement Levels

RFCs are classified into five *requirement levels:* required, recommended, elective, limited use, and not recommended.

Required: An RFC is labeled *required* if it must be implemented by all Internet systems to achieve minimum conformance. For example, IP and ICMP are required protocols.

Recommended: An RFC labeled recommended is not required for minimum conformance; it is recommended because of its usefulness. For example, FTP and TELNET are recommended protocols.

Elective: An RFC labeled elective is not required and not recommended. However, a system can use it for its own benefit.

Limited Use: An RFC labeled limited use should be used only in limited situations. Most of the experimental RFCs fall under this category.

Not Recommended: An RFC labeled not recommended is inappropriate for general use. Normally a historic (deprecated) RFC may fall under this category.

INTERNET ADMINISTRATION:

The Internet, with its roots primarily in the research domain, has evolved and gained a broader user base with significant commercial activity. Various groups that coordinate Internet issues have guided this growth and development. Figure 1.14 shows the general organization of Internet administration.

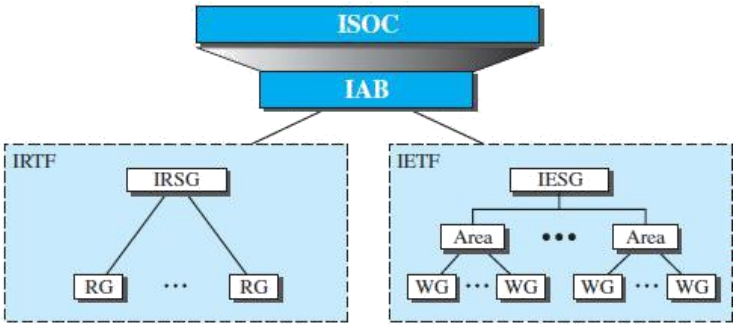


FIGURE 1.14: INTERNET ADMINISTRATION

ISOC: The **Internet Society (ISOC)** is an international, nonprofit organization formed in 1992 to provide support for the Internet standards process. ISOC accomplishes this through maintaining and supporting other Internet administrative bodies such as IAB, IETF, IRTF, etc... ISOC also promotes research and other scholarly activities relating to the Internet.

IAB: The **Internet Architecture Board (IAB)** is the technical advisor to the ISOC. The main purposes of the IAB are to oversee the continuing development of the TCP/IP Protocol Suite and to serve in a technical advisory capacity to research members of the Internet community. IAB accomplishes this through its two primary components, the Internet Engineering Task Force (IETF) and the Internet Research Task Force (IRTF).

Another responsibility of the IAB is the editorial management of the RFCs. IAB is also the external liaison (*meaning link / association*) between the Internet and other standards organizations and forums.

IETF: The **Internet Engineering Task Force (IETF)** is a forum of working groups managed by the Internet Engineering Steering Group (IESG). IETF is responsible for identifying operational problems and proposing solutions to these problems. IETF also develops and reviews specifications intended as Internet standards.

IRTF: The **Internet Research Task Force (IRTF)** is a forum of working groups managed by the Internet Research Steering Group (IRSG). IRTF focuses on long-term research topics related to Internet protocols, applications, architecture, and technology.

NETWORK MODELS

PROTOCOL LAYERING:

In data communication and networking, a protocol defines the rules that both the sender and receiver and all intermediate devices need to follow to be able to communicate effectively.

When communication is simple, we may need only one simple protocol.

When the communication is complex, we may need to divide the task between different layers, in which case we need a protocol at each layer, or **protocol layering.**

Scenarios:

Let us develop two simple scenarios to better understand the need for protocol layering.

First Scenario

In the first scenario, communication is so simple that it can occur in only one layer. Assume Maria and Ann are neighbors with a lot of common ideas. Communication between Maria and Ann takes place in one layer, face to face, in the same language, as shown in Figure 1.15.

FIGURE 1.15: A SINGLE-LAYER PROTOCOL

Second Scenario:

In the second scenario, we assume that Ann is offered a higher-level position in her company, but needs to move to another branch located in a city very far from Maria.

The two friends still want to continue their communication and exchange ideas because they have come up with an innovative project to start a new business when they both retire. They decide to continue their conversation using regular mail through the post office.

However, they do not want their ideas to be revealed by other people if the letters are intercepted. They agree on an encryption/decryption technique. The sender of the letter encrypts it to make it unreadable by an intruder; the receiver of the letter decrypts it to get the original letter.

For the moment we assume that Maria and Ann use one technique that makes it hard to decrypt the letter if one does not have the key for doing so. Now we can say that the communication between Maria and Ann takes place in three layers, as shown in Figure 1.16.

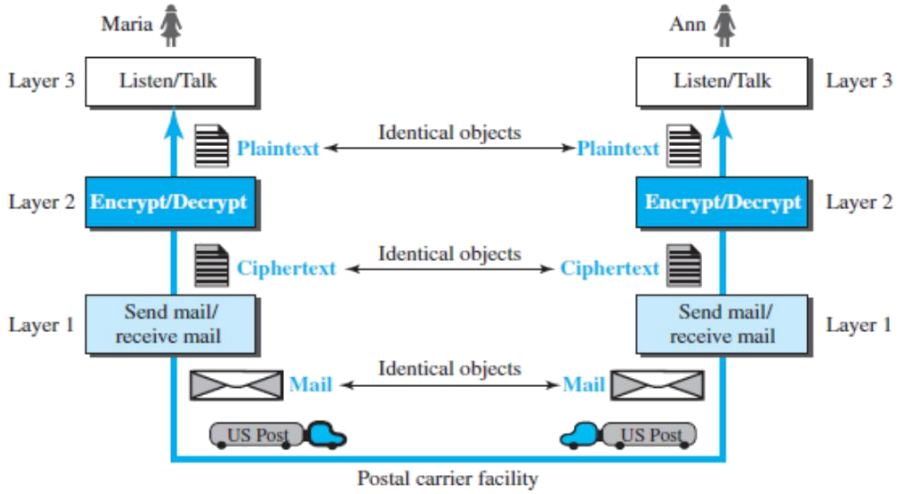


FIGURE 1.16: A THREE-LAYER PROTOCOL

Protocol layering enables us to divide a complex task into several smaller and simpler tasks. Advantage of protocol layering is that it allows us to separate the services from the implementation.

A layer needs to be able to receive a set of services from the lower layer and to give the services to the upper layer; we don't care about how the layer is implemented.

If we did not use protocol layering, we would have to make each intermediate system as complex as the end systems, which makes the whole system more expensive.

PRINCIPLES OF PROTOCOL LAYERING:

Let us discuss two principles of protocol layering.

First Principle: The first principle dictates that if we want bidirectional communication, we need to make each layer so that it is able to perform two opposite tasks, one in each direction.

Second Principle: The second principle that we need to follow in protocol layering is that the two objects under each layer at both sites should be identical.

Logical Connections:

After following the above two principles, we can think about logical connection between each layer as shown in Figure 1.17. This means that we have layer-to-layer communication.

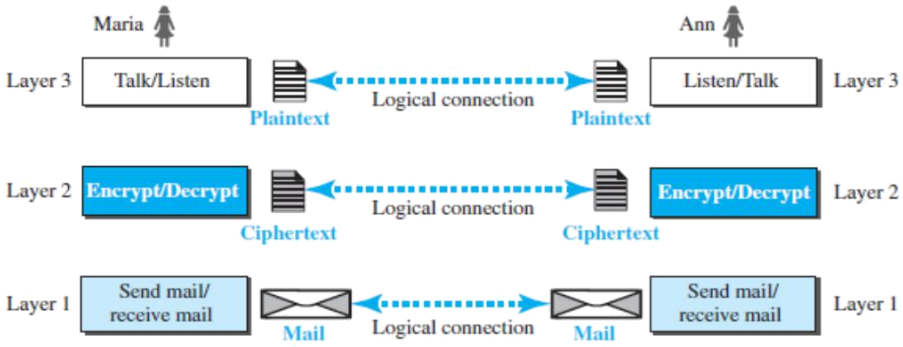


FIGURE 1.17: LOGICAL CONNECTION BETWEEN PEER LAYERS

TCP/IP PROTOCOL SUITE:

TCP/IP is a protocol suite (a set of protocols organized in different layers) used in the Internet today. It is a hierarchical protocol made up of interactive modules, each of which provides a specific functionality.

The term *hierarchical* means that each upper level protocol is supported by the services provided by one or more lower level protocols. The original TCP/IP protocol suite was defined as four software layers built upon the hardware. Today, however, TCP/IP is thought of as a five-layer model. Figure 1.18 shows both configurations.

LAYERED ARCHITECTURE:

To show how the layers in the TCP/IP protocol suite are involved in communication between two hosts, we assume that we want to use the suite in a small internet made up of three LANs (links), each with a link-layer switch. We also assume that the links are connected by one router, as shown in Figure 1.19.

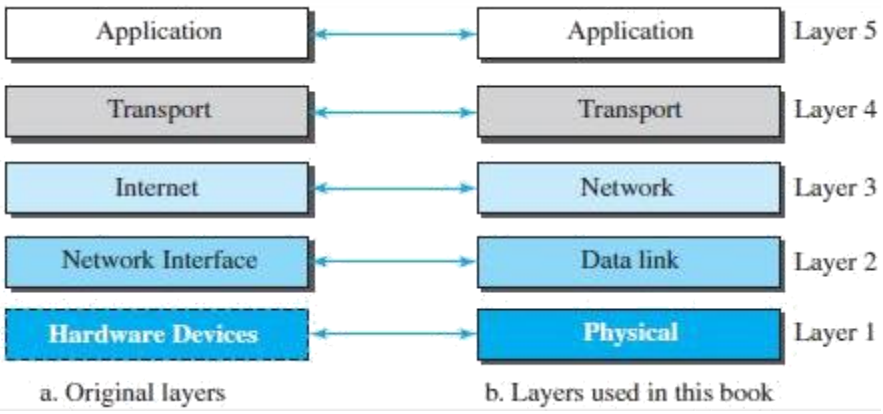


FIGURE 1.18: LAYERS IN THE TCP/IP PROTOCOL SUITE

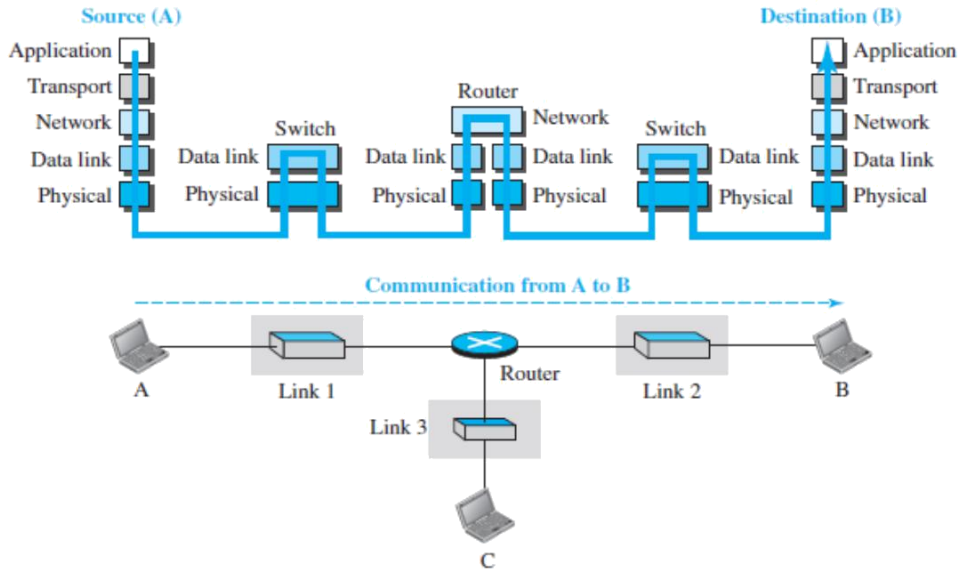


FIGURE 1.19: COMMUNICATION THROUGH AN INTERNET

Let us assume that computer A communicates with computer B. As the figure shows, we have five communicating devices in this communication: source host (computer A), the link-layer switch in link 1, the router, the link-layer switch in link 2, and the destination host computer B). Each device is involved with a set of layers depending on the role of the device in the internet.

The two hosts are involved in all five layers; the source host needs to create a message in the application layer and send it down the layers so that it is physically sent to the destination host. The destination host needs to receive the communication at the physical layer and then deliver it through the other layers to the application layer.

The router is involved in only three layers; there is no transport or application layer in a router as long as the router is used only for routing. Although a router is always involved in one network layer, it is involved in *n* combinations of link and physical layers in which *n* is the number of links the router is connected to. The reason is that each link may use its own data-link or physical protocol.

A link-layer switch in a link, however, is involved only in two layers, data-link and physical. Although each switch in the above figure has two different connections, the connections are in the same link, which uses only one set of protocols. This means that, unlike a router, a link-layer switch is involved only in one data-link and one physical layer.

LAYERS IN THE TCP/IP PROTOCOL SUITE:

To better understand the duties of each layer, we need to think about the logical connections between layers. Figure 1.20 shows logical connections in our simple internet.

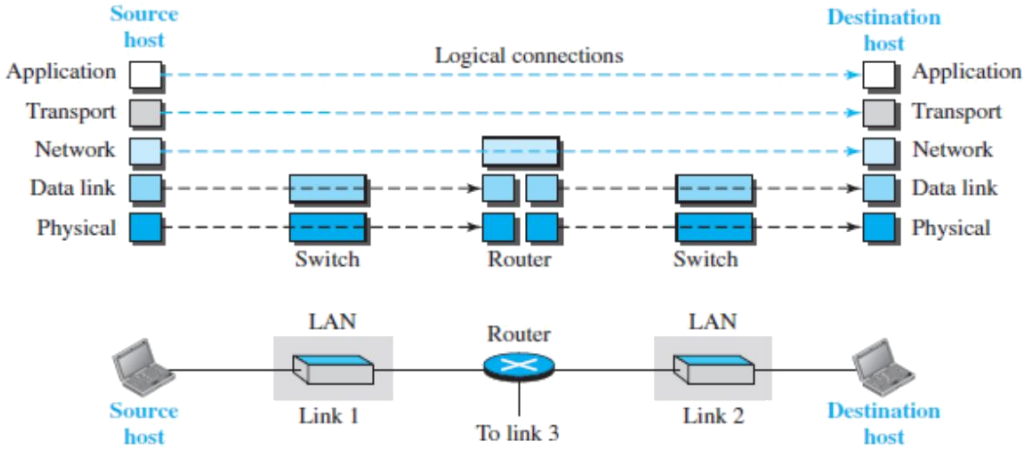


FIGURE 1.20: LOGICAL CONNECTIONS BETWEEN LAYERS OF THE TCP/IP PROTOCOL SUITE

Using logical connections makes it easier for us to think about the duty of each layer. As the figure shows, the duty of the application, transport, and network layers is end-to-end. However, the duty of the data-link and physical layers is hop-to-hop, in which a hop is a host or router. In other words, the domain of duty of the top three layers is the internet, and the domain of duty of the two lower layers is the link.

Another way of thinking of the logical connections is to think about the data unit created from each layer. In the top three layers, the data unit (packets) should not be changed by any router or link-layer switch.

In the bottom two layers,the packet created by the host is changed only by the routers, not by the link-layers switches. Figure 1.21 shows the second principle discussed previously for protocol layering.

Note that, although the logical connection at the network layer is between the two hosts, we can only say that identical objects exist between two hops in this case because a router may fragment the packet at the network layer and send more packets than received. Note that the link between two hops does not change the object.

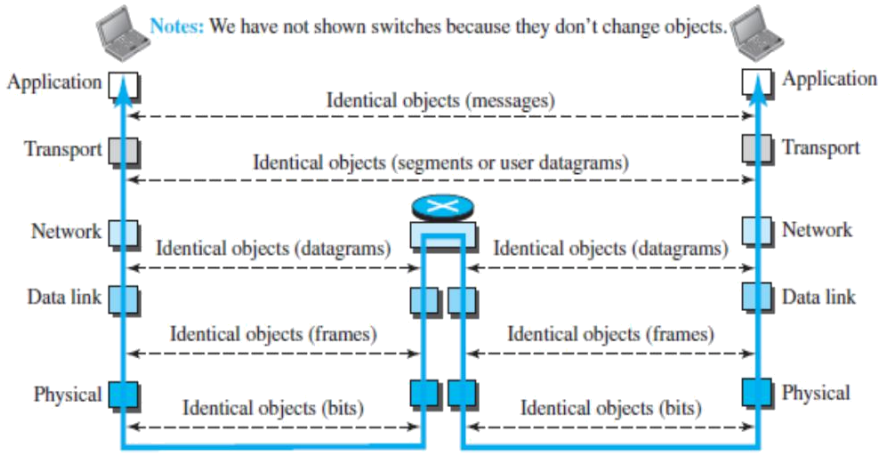


FIGURE 1.21: IDENTICAL OBJECTS IN THE TCP/IP PROTOCOL SUITE Description of each layer:

After understanding the concept of logical communication, we are ready to briefly discuss the duty of each layer.

Physical Layer:

We can say that the physical layer is responsible for carrying individual bits in a frame across the link. Although the physical layer is the lowest level in the TCP/IP protocol suite, the communication between two devices at the physical layer is still a logical communication because there is another, hidden layer, the transmission media, under the physical layer.

Data-link Layer:

The data-link layer is responsible for moving the packet through the link. TCP/IP does not define any specific protocol for the data-link layer. It supports all the standard and proprietary protocols.

Any protocol that can take the datagram and carry it through the link suffices for the network layer. The data-link layer takes a datagram and encapsulates it in a packet called a *frame*.

Network Layer:

The network layer is responsible for creating a connection between the source computer and the destination computer. The communication at the network layer is host-to-host.

However, since there can be several routers from the source to the destination, the routers in the path are responsible for choosing the best route for each packet. We can say that the network layer is responsible for host-to-host communication and routing the packet through possible routes.

The network layer in the Internet includes the main protocol, Internet Protocol (IP) that defines the format of the packet, called a datagram at the network layer. IP also defines the format and the structure of addresses used in this layer.

IP is also responsible for routing a packet from its source to its destination, which is achieved by each router forwarding the datagram to the next router in its path. IP is a connectionless protocol that provides no flow control, no error control, and no congestion control services.

This means that if any of these services is required for an application, the application should rely only on the transport-layer protocol. The network layer also includes unicast (one-to-one) and multicast (one-to-many) routing protocols.

A routing protocol does not take part in routing (it is the responsibility of IP), but it creates forwarding tables for routers to help them in the routing process. The network layer also has some auxiliary protocols that help IP in its delivery and routing tasks.

Transport Layer:

The logical connection at the transport layer is also end-to-end. The transport layer at the source host gets the message from the application layer, encapsulates it in a transport layer packet (called a *segment* or a *user datagram* in different protocols) and sends it, through the logical connection, to the transport layer at the destination host.

There are a few transport-layer protocols in the Internet, each designed for some specific task. The main protocol, Transmission Control Protocol (TCP), is a connection-oriented protocol that first establishes a logical connection between transport layers at two hosts before transferring data. It creates a logical pipe between two TCPs for transferring a stream of bytes.

TCP provides flow control (matching the sending data rate of the source host with the receiving data rate of the destination host to prevent overwhelming the destination), error control (to guarantee that the segments arrive at the destination without error and resending the corrupted ones), and congestion control to reduce the loss of segments due to congestion in the network.

The other common protocol, User Datagram Protocol (UDP), is a connectionless protocol that transmits user datagram’s without first creating a logical connection.

UDP is a simple protocol that does not provide flow, error, or congestion control. Its simplicity, which means small overhead, is attractive to an application

program that needs to send short messages and cannot afford the retransmission of the packets involved in TCP, when a packet is corrupted or lost.

Application Layer:

The logical connection between the two application layers is end-to-end. The two application layers exchange *messages* between each other as though there were a bridge between the two layers. However, we should know that the communication is done through all the layers.

Communication at the application layer is between two *processes* (two programs running at this layer). To communicate, a process sends a request to the other process and receives a response.

Process-to-process communication is the duty of the application layer. The application layer in the Internet includes many predefined protocols, but a user can also create a pair of processes to be run at the two hosts.

The Hypertext Transfer Protocol (HTTP) is a vehicle for accessing the World Wide Web (WWW). The Simple Mail Transfer Protocol (SMTP) is the main protocol used in electronic mail (e-mail) service. The File Transfer Protocol (FTP) is used for transferring files from one host to another.

The Terminal Network (TELNET) & Secure Shell (SSH) are used for accessing a site remotely. The Simple Network Management Protocol (SNMP) is used by an administrator to manage the Internet at global and local levels.

The Domain Name System (DNS) is used by other protocols to find the network-layer address of a computer. The Internet Group Management Protocol

(IGMP) is used to collect membership in a group.

ENCAPSULATION AND DECAPSULATION:

One of the important concepts in protocol layering in the Internet is encapsulation/ decapsulation. Figure 1.22 shows this concept for the small internet. We have not shown the layers for the link-layer switches because no encapsulation/decapsulation occurs in this device.

Encapsulation at the Source Host:

At the source, we have only encapsulation.

1. At the application layer, the data to be exchanged is referred to as a *message*. A message normally does not contain any header or trailer, but if it does, we refer to the whole as the message. The message is passed to the transport layer.

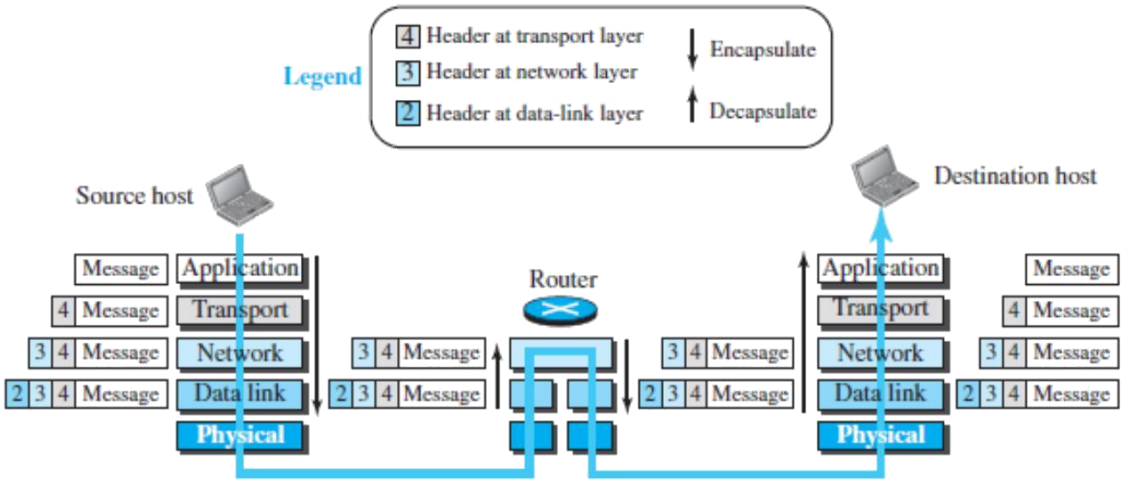


FIGURE 1.22: ENCAPSULATION/DECAPSULATION

2. The transport layer takes the message as the payload, the load that the transport layer should take care of. It adds the transport layer header to the payload, which contains the identifiers of the source, and destination application programs that want to communicate plus some more information that is needed for the end-to-end delivery of the message, such as information needed for flow, error control, or congestion control.

   a. The result is the transport-layer packet, which is called the *segment* (in TCP) and the *user datagram* (in UDP). The transport layer then passes the packet to the network layer.

3. The network layer takes the transport-layer packet as data or payload and adds its own header to the payload. The header contains the addresses of the source and destination hosts and some more information used for error checking of the header, fragmentation information, and so on. The result is the network-layer packet, called a *datagram*. The network layer then passes the packet to the data-link layer.

4. The data-link layer takes the network-layer packet as data or payload and adds its own header, which contains the link-layer addresses of the host or the next hop (the router). The result is the link-layer packet, which is called a *frame*. The frame is passed to the physical layer for transmission.

Decapsulation and Encapsulation at the Router:

At the router, we have both decapsulation and encapsulation because the router is connected to two or more links.

1. After the set of bits are delivered to the data-link layer, this layer decapsulates the datagram from the frame and passes it to the network layer.

2. The network layer only inspects the source and destination addresses in the datagram header and consults its forwarding table to find the next hop to which the datagram is to be delivered.

    a. The contents of the datagram should not be changed by the network layer in the router unless there is a need to fragment the datagram if it is too big to be passed through the next link. The datagram is then passed to the data-link layer of the next link.

3. The data-link layer of the next link encapsulates the datagram in a frame and passes it to the physical layer for transmission.

Decapsulation at the Destination Host:

At the destination host, each layer only decapsulates the packet received, removes the payload, and delivers the payload to the next-higher layer protocol until the message reaches the application layer. It is necessary to say that de capsulation in the host involves error checking.

Addressing: It is worth mentioning another concept related to protocol layering in the Internet, *addressing*. As we discussed before, we have logical communication between pairs of layers in this model.

Any communication that involves two parties needs two addresses: source address and destination address. Although it looks as if we need five pairs of addresses, one pair per layer, we normally have only four because the physical layer does not need addresses; the unit of data exchange at the physical layer is a bit, which definitely cannot have an address. Figure 1.23 shows the addressing at each layer.

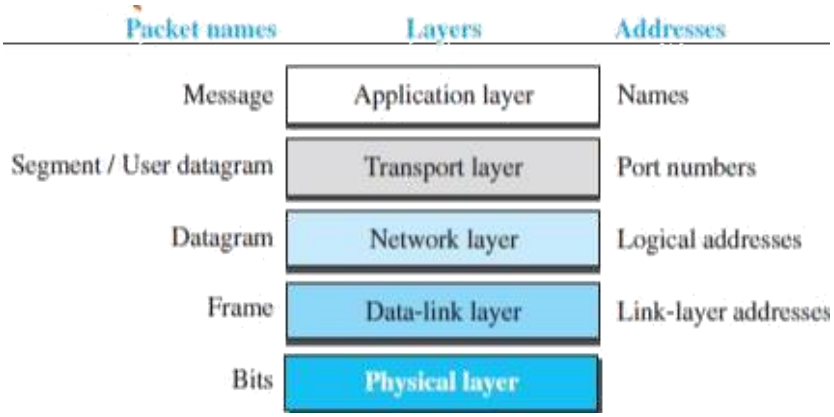


FIGURE 1.23: ADDRESSING IN THE TCP/IP PROTOCOL SUITE

As the figure shows, there is a *relationship between the layer*, *the address used in that layer*, and *the packet name at that layer*.

☑ At the application layer, we normally use names to define the site that provides services, such as *someorg.com*, or the e-mail address, such as *somebody@coldmail.com.*

- ☑ At the transport layer, addresses are called port numbers, and these define the application-layer programs at the source and destination. Port numbers are local addresses that distinguish between several programs running at the same time.

- ☑ At the network-layer, the addresses are global, with the whole Internet as the scope. A network-layer address uniquely defines the connection of a device to the Internet.

- ☑ The link-layer addresses, sometimes called MAC addresses, are locally defined addresses, each of which defines a specific host or router in a network (LAN or WAN).

Multiplexing and Demultiplexing:

Since the TCP/IP protocol suite uses several protocols at some layers, we can say that we have multiplexing at the source and demultiplexing at the destination.

Multiplexing in this case means that a protocol at a layer can encapsulate a packet from several next-higher layer protocols (one at a time); demultiplexing means that a protocol can decapsulate and deliver a packet to several next-higher layer protocols (one at a time).

Figure 1.24 shows the concept of multiplexing and demultiplexing at the three upper layers.

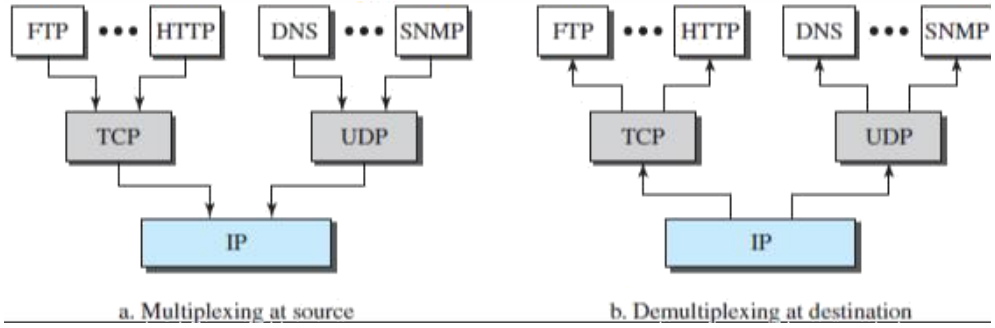


FIGURE 1.24: MULTIPLEXING AND DEMULTIPLEXING

To be able to multiplex and demultiplex, a protocol needs to have a field in its header to identify to which protocol the encapsulated packets belong.

- ☑ At the transport layer, either UDP or TCP can accept a message from several application-layer protocols.

- ☑ At the network layer, IP can accept a segment from TCP or a user datagram from UDP. IP can also accept a packet from other protocols such as ICMP, IGMP, and so on.

- ☑ At the data-link layer, a frame may carry the payload coming from IP or other protocols such as ARP.

THE OSI MODEL:

Established in 1947, the **International Organization for Standardization (ISO)** is a multinational body dedicated to worldwide agreement on international standards. Almost three-fourths of the countries in the world are represented in the ISO. An ISO standard that covers all aspects of network communications is the **Open Systems Interconnection (OSI) model.** It was first introduced in the late 1970s.

An *open system* is a set of protocols that allows any two different systems to communicate regardless of their underlying architecture.

- ☑ The purpose of the OSI model is to show how to facilitate communication between different systems without requiring changes to the logic of the underlying hardware and software.
- ☑ The OSI model is not a protocol; it is a model for understanding and designing a network architecture that is flexible, robust, and interoperable
- ☑ The OSI model was intended to be the basis for the creation of the protocols in the OSI stack.

The OSI model is a layered framework for the design of network systems that allows communication between all types of computer systems. It consists of seven separate but related layers, each of which defines a part of the process of moving information across a network (see Figure 1.25).

OSI versus TCP/IP:

When we compare the two models, we find that two layers, session and presentation, are missing from the TCP/IP protocol suite. These two layers were not added to the TCP/IP protocol suite after the publication of the OSI model. The application layer in the suite is usually considered to be the combination of three layers in the OSI model, as shown in Figure 1.26.

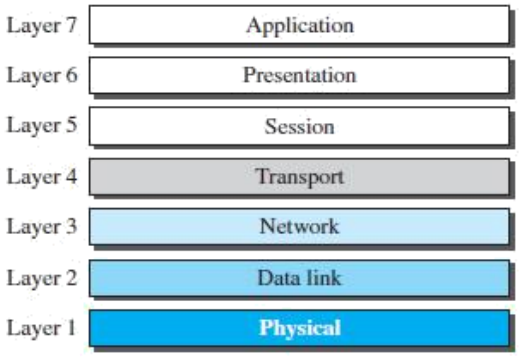


FIGURE 1.25: THE OSI MODEL

Two reasons were mentioned for this decision. *First*, TCP/IP has more than one transport-layer protocol. Some of the functionalities of the session layer are available in some of the transport-layer protocols.

Second, the application layer is not only one piece of software. Many applications can be developed at this layer. If some of the functionalities mentioned in the session and presentation layers are needed for a particular application, they can be included in the development of that piece of software.

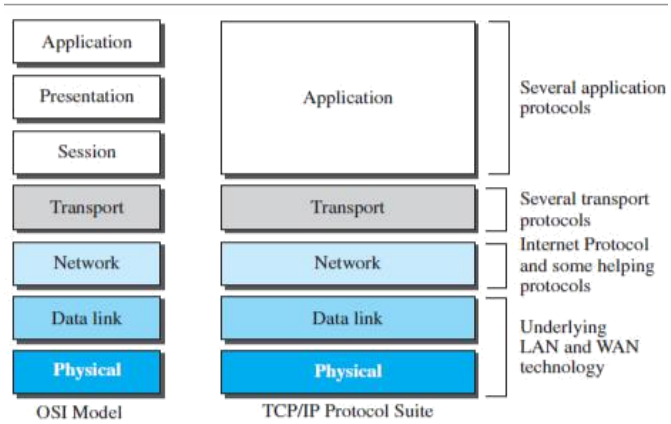


FIGURE 1.26: TCP/IP AND OSI MODEL

Lack of OSI model's success:

The OSI model appeared after the TCP/IP protocol suite. Most experts were at first excited and thought that the TCP/IP protocol would be fully replaced by the OSI model. This did not happen for several reasons, but we describe only three, which are agreed upon by all experts in the field.

First, OSI was completed when TCP/IP was fully in place and a lot of time and money had been spent on the suite; changing it would cost a lot.

Second, some layers in the OSI model were never fully defined

Third, when OSI was implemented by an organization in a different application, it did not show a high enough level of performance to entice (*meaning*

attract) the Internet authority to switch from the TCP/IP protocol suite to the OSI model.

INTRODUCTION TO PHYSICAL LAYER

DATA AND SIGNALS:

Analog and Digital Data:

Data can be analog or digital. The term **analog data** refers to information that is continuous; **digital data** refers to information that has discrete states.

Analog data, such as the sounds made by a human voice, take on continuous values. When someone speaks, an analog wave is created in the air. This can be captured by a microphone and converted to an analog signal or sampled and converted to a digital signal.

Digital data take on discrete values. For example, data are stored in computer memory in the form of 0s and 1s. They can be converted to a digital signal or modulated into an analog signal for transmission across a medium.

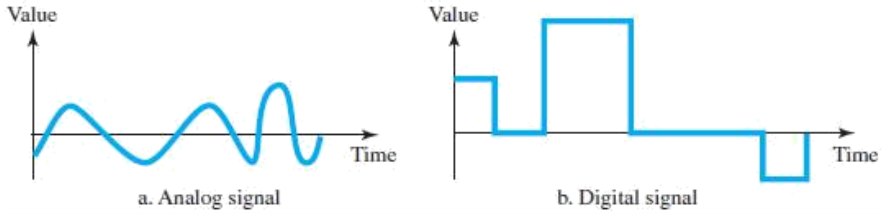
ANALOG AND DIGITAL SIGNALS: Like the data they represent, **signals** can be either analog or digital.

An **analog signal** has infinitely many levels of intensity (*meaning strength/power*) over a period of time. As the wave moves from value *A* to value *B*, it passes through and includes an infinite number of values along its path.

A **digital signal,** on the other hand, can have only a limited number of defined values. Although each value can be any number, it is often as simple as 1 and 0.

The simplest way to show signals is by plotting them on a pair of perpendicular axes. The vertical axis represents the value or strength of a signal. The horizontal axis represents time.

Figure 1.27 illustrates an analog signal and a digital signal. The curve representing the analog signal passes through an infinite number of points. The vertical lines of the digital signal, however, demonstrate the sudden jump that the signal makes from value to value.



a. Analog signal b. Digital signal

FIGURE 1.27: COMPARISON OF ANALOG AND DIGITAL SIGNALS

Periodic and Nonperiodic: Both analog and digital signals can take one of two forms: *periodic* or *nonperiodic* (Sometimes referred to as *aperiodic;* the prefix *a* in Greek means "non").

A **periodic signal** completes a pattern within a measurable time frame, called a **period,** and repeats that pattern over subsequent identical periods. The completion of one full pattern is called a **cycle.**

A **nonperiodic signal** changes without exhibiting a pattern or cycle that repeats over time. Both analog and digital signals can be periodic or nonperiodic.

TRANSMISSION IMPAIRMENT:

Signals travel through transmission media, which are not perfect. The imperfection causes signal impairment. This means that the signal at the beginning of the medium is not the same as the signal at the end of the medium. What is sent is not what is received. Three causes of impairment are ***attenuation***, ***distortion***, and ***noise*** (see Figure 1.28).

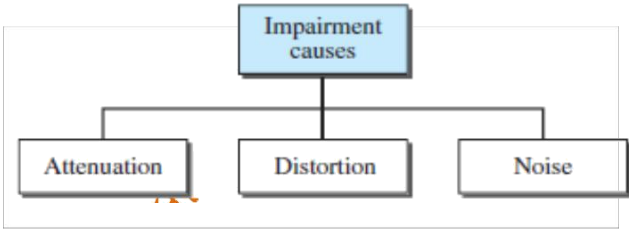


FIGURE 1.28: CAUSES OF IMPAIRMENT

Attenuation: It means a loss of energy. When a signal, simple or composite, travels through a medium, it loses some of its energy in overcoming the resistance of the medium. That is why a wire carrying electric signals gets warm, if not hot, after a while. Some of the electrical energy in the signal is converted to heat. To compensate for this loss, amplifiers are used to amplify (*meaning enlarge on/go into detail/develop/expand/clarify/add details to*) the signal.

Decibel: To show that a signal has lost or gained strength, engineers use the unit of the decibel. The **decibel (dB)** measures the relative strengths of two signals or one signal at two different points. Note that the decibel is negative if a signal is attenuated and positive if a signal is amplified.

Distortion: It means that the signal changes its form or shape. Distortion can occur in a composite signal made of different frequencies. Each signal component has its own propagation speed through a medium and, therefore, its own delay in arriving at the final destination. Differences in delay may create a difference in phase if the delay is not exactly the same as the period duration.

Noise: It is another cause of impairment. Several types of noise, such as ***thermal noise***, ***induced noise***, ***crosstalk***, and ***impulse noise***, may corrupt the signal.

Thermal noise is the random motion of electrons in a wire, which creates an extra signal not originally sent by the transmitter.

Induced noise comes from sources such as motors and appliances. These devices act as a sending antenna, and the transmission medium acts as the receiving antenna.

Crosstalk is the effect of one wire on the other. One wire acts as a sending antenna and the other as the receiving antenna.

Impulse noise is a spike (a signal with high energy in a very short time) that comes from power lines, lightning, and so on.

DATA RATE LIMITS:

A very important consideration in data communications is how fast we can send data, in bits per second, over a channel. Data rate depends on three factors:

1. The bandwidth available
2. The level of the signals we use
3. The quality of the channel (the level of noise)

Two theoretical formulas were developed to calculate the data rate: one by ***Nyquist for a noiseless channel***, another by ***Shannon for a noisy channel***.

Noiseless Channel: Nyquist Bit Rate:

For a noiseless channel, the **Nyquist bit rate** formula defines the theoretical maximum bit rate **BitRate = 2 x bandwidth x $\log_2$L**

In this formula, bandwidth is the bandwidth of the channel, *L* is the number of signal levels used to represent data, and BitRate is the bit rate in bits per second.

Noisy Channel: Shannon Capacity:

In reality, we cannot have a noiseless channel; the channel is always noisy. In 1944, Claude Shannon introduced a formula, called the **Shannon capacity,** to determine the theoretical highest data rate for a noisy channel: **Capacity = bandwidth x $\log_2$ (1 + SNR)**

In this formula, bandwidth is the bandwidth of the channel, SNR is the signal-to-noise ratio, and capacity is the capacity of the channel in bits per second.

PERFORMANCE:

One important issue in networking is the performance of the network—how good is it? There are certain characteristics that measure the network performance which are given as follows:

Bandwidth: One characteristic that measures network performance is bandwidth. However, the term can be used in two different contexts with two different measuring values: bandwidth in hertz and bandwidth in bits per second.

Bandwidth in Hertz: Bandwidth in hertz is the range of frequencies contained in a composite signal or the range of frequencies a channel can pass. For example, we can say the bandwidth of a subscriber telephone line is 4 kHz.

Bandwidth in Bits per Seconds: The term *bandwidth* can also refer to the number of bits per second that a channel, a link, or even a network can transmit. For example, one can say the bandwidth of a Fast Ethernet network is a maximum of 100 Mbps.

Relationship: An increase in bandwidth in hertz means an increase in bandwidth in bits per second.

THROUGHPUT:

The **throughput** is a measure of how fast we can actually send data through a network. Although, at first glance, bandwidth in bits per second and throughput seem the same, they are different. A link may have a bandwidth of *B* bps, but we can only send *T* bps through this link with *T* always less than *B*.

For example, we may have a link with a bandwidth of 1 Mbps, but the devices connected to the end of the link may handle only 200 kbps. This means that we cannot send more than 200 kbps through this link.

LATENCY (DELAY):

The **latency** or delay defines how long it takes for an entire message to completely arrive at the destination from the time the first bit is sent out from the source. We can say that latency is made of four components: ***propagation time***, ***transmission time***, ***queuing time*** and ***processing delay***.

Latency = **propagation time + transmission time + queuing time + processing delay**

Propagation Time: **Propagation time** measures the time required for a bit to travel from the source to the destination. The propagation time is calculated by dividing the distance by the propagation speed.

Propagation time = Distance / (Propagation Speed)

Transmission Time: The time between the first bit leaving the sender and the last bit arriving at the receiver.

The first bit leaves earlier and arrives earlier; the last bit leaves later and arrives later. The **transmission time** of a message depends on the size of the message and the bandwidth of the channel.

Transmission time = (Message size) / Bandwidth

Queuing Time: The third component in latency is the **queuing time**, the time needed for each intermediate or end device to hold the message before it can be processed. The queuing time is not a fixed factor; it changes with the load imposed on the network.

JITTER:

Another performance issue that is related to delay is **jitter.** We can roughly say that jitter is a problem if different packets of data encounter different delays and the application using the data at the receiver site is time-sensitive (audio and video data, for example).

TRANSMISSION MEDIA

INTRODUCTION:

Transmission media are actually located below the physical layer and are directly controlled by the physical layer. We could say that transmission media belong to layer zero. Figure 1.29 shows the position of transmission media in relation to the physical layer.

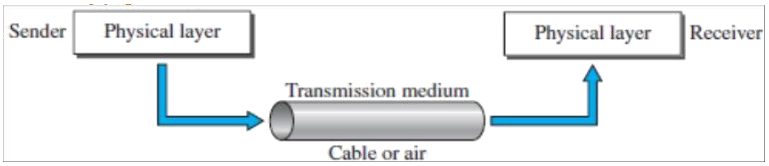


FIGURE 1.29: TRANSMISSION MEDIUM AND PHYSICAL LAYER

- ☑ A **transmission medium** can be broadly defined as anything that can carry information from a source to a destination. For example, the transmission medium for two people having a dinner conversation is the air.
- ☑ The air can also be used to convey the message in a smoke signal or semaphore.
- ☑ For a written message, the transmission medium might be a mail carrier, a truck, or an airplane.

- ☑ In data communications the definition of the information and the transmission medium is more specific. The transmission medium is usually free space, metallic cable, or fiber-optic cable.
- ☑ The information is usually a signal that is the result of a conversion of data from another form.
- ☑ The use of long-distance communication using electric signals started with the invention of the telegraph by Morse in the 19th century.
- ☑ Communication by telegraph was slow and dependent on a metallic medium. Extending the range of the human voice became possible when the *telephone was invented in 1869*.
- ☑ Telephone communication at that time also needed a metallic medium to carry the electric signals that were the result of a conversion from the human voice.
- ☑ The communication was, however, unreliable due to the poor quality of the wires. The lines were often noisy and the technology was unsophisticated.
- ☑ *Wireless communication started in 1895* when Hertz was able to send high frequency signals. Later, Marconi devised a method to send telegraph-type messages over the Atlantic Ocean.
- ☑ We have come a long way. Better metallic media have been invented (twisted-pair and coaxial cables, for example).

The use of optical fibers has increased the data rate incredibly. Free space (air, vacuum, and water) is used more efficiently, in part due to the technologies (such as modulation and multiplexing).

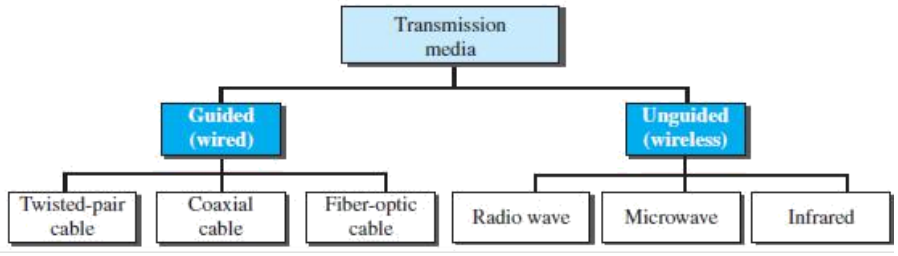


FIGURE 1.30: CLASSES OF TRANSMISSION MEDIA

GUIDED MEDIA:

In telecommunications, transmission media can be divided into two broad categories: guided and unguided.

Guided media are those that provide a conduit (*meaning medium*) from one device to another. Guided media include twisted-pair cable, coaxial cable, and fiber-optic cable. Unguided medium is free space. Figure 1.30 shows this taxonomy.

A signal traveling along any of these media is directed and contained by the physical limits of the medium.

TWISTED-PAIR CABLE:

A twisted pair consists of two conductors (normally copper), each with its own plastic insulation, twisted together, as shown in Figure 1.31.

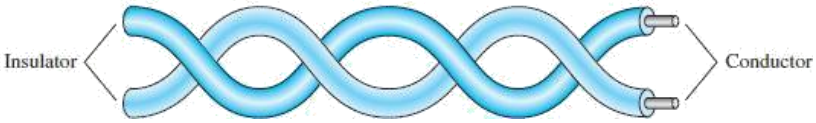


FIGURE 1.31: TWISTED-PAIR CABLE

One of the wires is used to carry signals to the receiver, and the other is used only as a ground reference. The receiver uses the difference between the two. In addition to the signal sent by the sender on one of the wires, interference (noise) and crosstalk may affect both wires and create unwanted signals.

If the two wires are parallel, the effect of these unwanted signals is not the same in both wires because they are at different locations relative to the noise or crosstalk sources (e.g., one is closer and the other is farther). This results in a difference at the receiver.

By twisting the pairs, a balance is maintained. Twisting makes it probable that both wires are equally affected by external influences (noise or crosstalk). This means that the receiver, which calculates the difference between the two, receives no unwanted signals.

Unshielded Versus Shielded Twisted-Pair Cable The most common twisted-pair cable used in communications is referred to as ***unshielded twisted-pair* (UTP).** IBM has also produced a version of twisted-pair cable for its use, called ***shielded twisted-pair* (STP);** STP cable has a metal foil or braided mesh covering that encases each pair of insulated conductors.

Although metal casing improves the quality of cable by preventing the penetration of noise or crosstalk, it is bulkier and more expensive. Figure 1.32 shows the difference between UTP and STP.

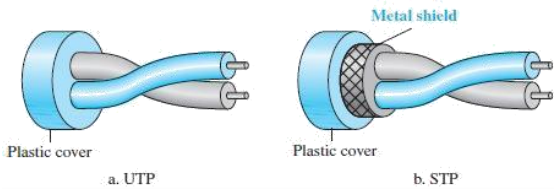


FIGURE 1.32: UTP AND STP CABLES

Connectors: The most common UTP connector is **RJ45** (RJ stands for registered jack), as shown in Figure 1.33. The RJ45 is a keyed connector, meaning the connector can be inserted in only one way.

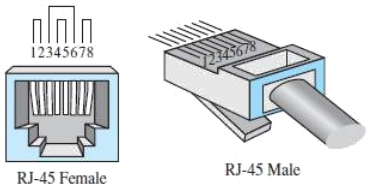


FIGURE 1.33: UTP CONNECTOR

Performance: One way to measure the performance of twisted-pair cable is to compare attenuation versus frequency and distance. A twisted-pair cable can pass a wide range of frequencies.

Applications: Twisted-pair cables are used in telephone lines to provide voice and data channels. The local loop—the line that connects subscribers to the central telephone office— commonly consists of unshielded twisted-pair cables.

The DSL lines that are used by the telephone companies to provide high-data-rate connections also use the high-bandwidth capability of unshielded twisted-pair cables.

COAXIAL CABLE:

Coaxial cable (or *coax*) carries signals of higher frequency ranges than those in twisted pair cable, in part because the two media are constructed quite differently. Instead of having two wires, coax has a central core conductor of solid or stranded wire (usually copper) enclosed in an insulating sheath, which is, in turn, encased in an outer conductor of metal foil, braid, or a combination of the two.

The outer metallic wrapping serves both as a shield against noise and as the second conductor, which completes the circuit; this outer conductor is also enclosed in an insulating sheath, and the whole cable is protected by a plastic cover (see Figure 1.34).

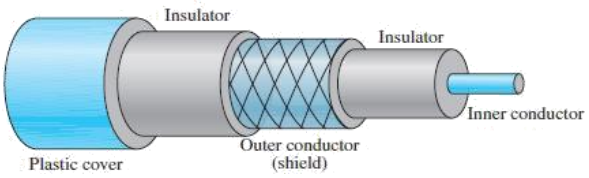


FIGURE 1.34: COAXIAL CABLE

Coaxial Cable Standards: Coaxial cables are categorized by their **Radio Government (RG)** ratings. Each RG number denotes a unique set of physical

specifications, including the wire gauge of the inner conductor, the thickness and type of the inner insulator, the construction of the shield, and the size and type of the outer casing.

Each cable defined by an RG rating is adapted for a specialized function, as shown in Table 1.1.

Category	Impedance	Use
RG-59	75 Ω	Cable TV
RG-58	50 Ω	Thin Ethernet
RG-11	50 Ω	Thick Ethernet

TABLE 1.1: CATEGORIES OF COAXIAL CABLES

Coaxial Cable Connectors: The most common type of connector used today is the **Bayonet Neill-Concelman (BNC)** connector.

Figure 1.35 shows three popular types of these connectors: the BNC connector, the BNC T connector, and the BNC terminator.

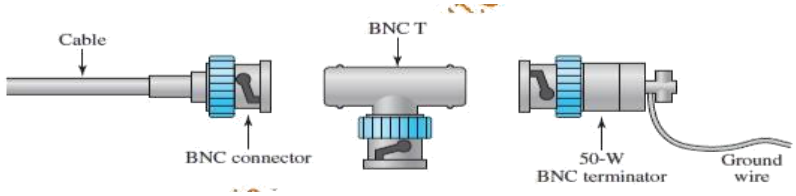


FIGURE 1.35: BNC CONNECTORS

The BNC connector is used to connect the end of the cable to a device, such as a TV set. The BNC T connector is used in Ethernet networks to branch out to a connection to a computer or other device. The BNC terminator is used at the end of the cable to prevent the reflection of the signal.

FIBER-OPTIC CABLE:

A fiber-optic cable is made of glass or plastic and transmits signals in the form of light. Optical fibers use reflection to guide light through a channel.

A glass or plastic **core** is surrounded by a **cladding** of less dense glass or plastic. The difference in density of the two materials must be such that a beam of light moving through the core is reflected off the cladding instead of being refracted into it. See Figure 1.36.

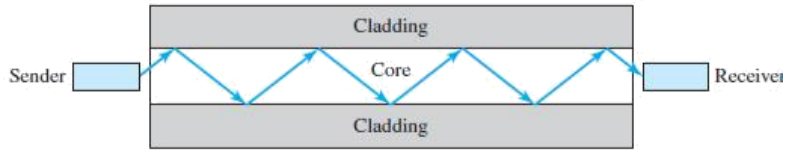


FIGURE 1.36: OPTICAL FIBER

Propagation Modes: Current technology supports two modes multimode and single mode for propagating light along optical channels, each requiring fiber with different physical characteristics. Multimode can be implemented in two forms: step-index or graded-index.

Multimode: Multimode is so named because multiple beams from a light source move through the core in different paths. How these beams move within the cable depends on the structure of the core.

Single-Mode: Single-mode uses step-index fiber and a highly focused source of light that limits beams to a small range of angles, all close to the horizontal.

Fiber Sizes: Optical fibers are defined by the ratio of the diameter of their core to the diameter of their cladding, both expressed in micrometers. The common sizes are shown in Table 1.2. Note that the last size listed is for single-mode only.

Type	Core (μm)	Cladding (μm)	Mode
50/125	50.0	125	Multi mode, Graded index
62.5/125	62.5	125	Multi mode, Graded index
100/125	100.0	125	Multi mode, Graded index
7/125	7.0	125	Single mode

TABLE 1.2: FIBER TYPES

Cable Composition

Figure 1.37 shows the composition of a typical fiber-optic cable. The outer jacket is made of either PVC or Teflon. Inside the jacket are Kevlar strands to strengthen the cable.

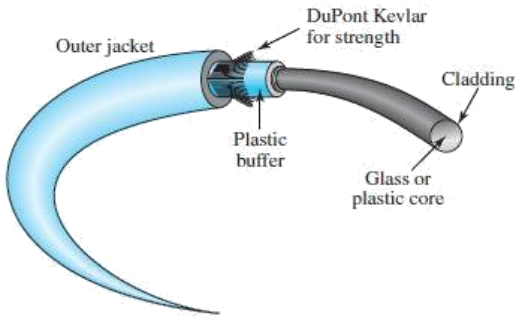


FIGURE 1.37: FIBER CONSTRUCTION

Kevlar is a strong material used in the fabrication of bulletproof vests. Below the Kevlar is another plastic coating to cushion the fiber. The fiber is at the center of the cable, and it consists of cladding and core.

Fiber-Optic Cable Connectors:

There are three types of connectors for fiber-optic cables, as shown in Figure 1.38. Subscriber channel (SC) connector is used for cable TV. The straight-tip (ST) connector is used for connecting cable to networking devices. It uses a bayonet locking system and is more reliable than SC. MT-RJ is a connector that is the same size as RJ45.

Performance: The performance is such that we need fewer (actually one tenth as many) repeaters when we use fiber-optic cable.

Applications: Fiber-optic cable is often found in backbone networks because its wide bandwidth is cost-effective. Some cable TV companies use a combination of optical fiber and coaxial cable, thus creating a hybrid network.

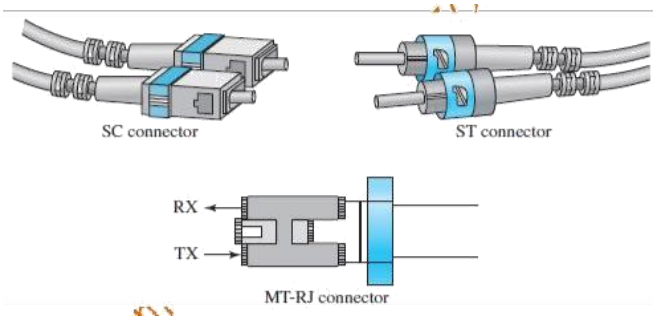


FIGURE 1.38: FIBER-OPTIC CABLE CONNECTORS

ADVANTAGES: Fiber-optic cable has several advantages over metallic cable (twisted-pair or coaxial).

Higher bandwidth. Fiber-optic cable can support dramatically higher bandwidths (and hence data rates) than either twisted-pair or coaxial cable.

Less signal attenuation. Fiber-optic transmission distance is significantly greater than that of other guided media. A signal can run for 50 km without requiring regeneration. We need repeaters every 5 km for coaxial or twisted-pair cable.

Immunity to electromagnetic interference. Electromagnetic noise cannot affect fiber-optic cables.

Resistance to corrosive materials. Glass is more resistant to corrosive materials than copper.

Light weight. Fiber-optic cables are much lighter than copper cables.

Greater immunity to tapping. Fiber-optic cables are more immune to tapping than copper cables. Copper cables create antenna effects that can easily be tapped.

Disadvantages: There are some disadvantages in the use of optical fiber. **Installation and maintenance:** Fiber-optic cable is a relatively new technology. Its installation and maintenance require expertise that is not yet available everywhere.

UNGUIDED MEDIA: WIRELESS:

Unguided medium transport electromagnetic waves without using a physical conductor. This type of communication is often referred to as ***wireless communication*****.** Signals are normally broadcast through free space and thus are available to anyone who has a device capable of receiving them.

Unguided signals can travel from the source to the destination in several ways: ground propagation, sky propagation, and line-of-sight propagation.

In **ground propagation,** radio waves travel through the lowest portion of the atmosphere, hugging the earth.

In **sky propagation,** higher-frequency radio waves radiate upward into the ionosphere (the layer of atmosphere where particles exist as ions) where they are reflected back to earth. This type of transmission allows for greater distances with lower output power.

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.

We can divide wireless transmission into three broad groups: radio waves, microwaves, and infrared waves.

RADIO WAVES:

Although there is no clear-cut demarcation between radio waves and microwaves, electromagnetic waves ranging in frequencies between 3 kHz and 1 GHz are normally called **radio waves;** waves ranging in frequencies between 1 and 300 GHz are called **microwaves.**

However, the behavior of the waves, rather than the frequencies, is a better criterion for classification. Radio waves, for the most part, are omnidirectional. When an antenna transmits radio waves, they are propagated in all directions.

Radio waves, particularly those waves that propagate in the sky mode, can travel long distances. This makes radio waves a good candidate for long-distance broadcasting such as AM radio.

Radio waves, particularly those of low and medium frequencies, can penetrate walls. This characteristic can be both an advantage and a disadvantage.

- It is an advantage because, for example, an AM radio can receive signals inside a building.
- It is a disadvantage because we cannot isolate a communication to just inside or outside a building.

Omnidirectional Antenna: Radio waves use **omnidirectional antennas** that send out signals in all directions. Based on the wavelength, strength, and the purpose of transmission, we can have several types of antennas. Figure 1.39 shows an omnidirectional antenna.

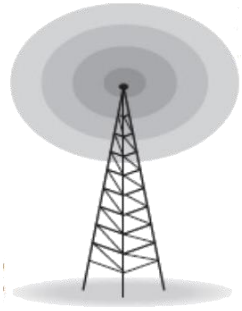
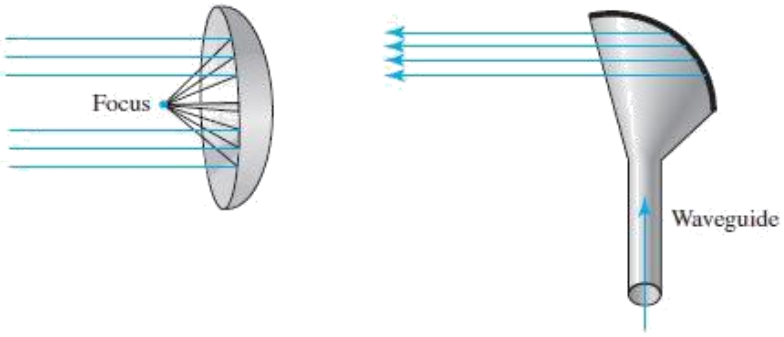


FIGURE 1.39: OMNIDIRECTIONAL ANTENNA

Applications: The **omnidirectional** characteristics of radio waves make them useful for multicasting, in which there is one sender but many receivers. AM and FM radio, television, maritime radio, cordless phones, and paging are examples of multicasting.

Microwaves: Electromagnetic waves having frequencies between 1 and 300 GHz are called microwaves. Microwaves are unidirectional. When an antenna transmits microwaves, they can be narrowly focused. This means that the sending and receiving antennas need to be aligned. The unidirectional property has an obvious advantage. A pair of antennas can be aligned without interfering with another pair of aligned antennas.

Unidirectional Antenna: Microwaves need **unidirectional antennas** that send out signals in one direction. Two types of antennas are used for microwave communications: the parabolic dish and the horn (see Figure 1.40).



a. Parabolic dish antenna b. Horn antenna

FIGURE 1.40: UNIDIRECTIONAL ANTENNAS

Applications: Microwaves, due to their unidirectional properties, are very useful when unicast (one-to- one) communication is needed between the sender and the receiver. They are used in cellular phones, satellite networks and wireless LANs.

Infrared: Infrared waves, with frequencies from 300 GHz to 400 THz (wavelengths from 1 mm to 770 nm), can be used for short-range communication. Infrared waves, having high frequencies, cannot penetrate walls.

This advantageous characteristic prevents interference between one system and another; a short-range communication system in one room cannot be affected by another system in the next room.

Applications: The infrared band, almost 400 THz, has an excellent potential for data transmission. Such a wide bandwidth can be used to transmit digital data with a very high data rate.

SWITCHING:

We have switching at the physical layer, at the data-link layer, at the network layer, and even logically at the application layer (message switching).

A switched network consists of a series of interlinked nodes, called ***switches*****.** Switches are devices capable of creating temporary connections between two or more devices linked to the switch. In a switched network, some of these nodes are connected to the end systems (computers or telephones, for example). Others are used only for routing. Figure 1.41 shows a switched network.

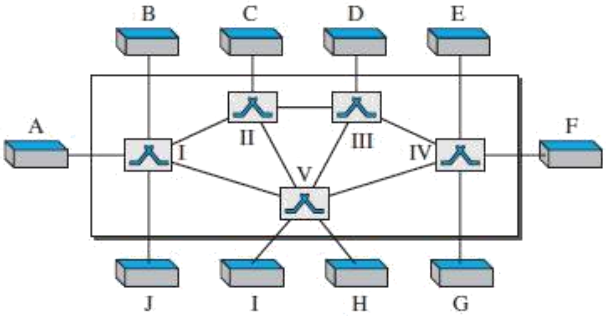


FIGURE 1.41: SWITCHED NETWORK

The **end systems** (communicating devices) are labeled A, B, C, D, and so on, and the switches are labeled I, II, III, IV, and V. Each switch is connected to multiple links.

Three methods of switching:

Traditionally, three methods of switching have been discussed: circuit switching, packet switching, and message switching. The first two are commonly used today. The third has been phased out in general communications but still has networking applications. Packet switching can further be divided into two subcategories—virtual circuit approach and datagram approach.

Switching and TCP/IP Layers:

Switching can happen at several layers of the TCP/IP protocol suite.

Switching at Physical Layer: At the physical layer, we can have only circuit switching. There are no packets exchanged at the physical layer. The switches at the physical layer allow signals to travel in one path or another.

Switching at Data-Link Layer: At the data-link layer, we can have packet switching. However, the term *packet* in this case means *frames* or *cells*. Packet switching at the data-link layer is normally done using a virtual-circuit approach.

Switching at Network Layer: At the network layer, we can have packet switching. In this case, either a virtual-circuit approach or a datagram approach can be used. Currently the Internet uses a datagram approach, but the tendency is to move to a virtual-circuit approach.

Switching at Application Layer: At the application layer, we can have only message switching. The communication at the application layer occurs by exchanging messages.

Conceptually, we can say that communication using e-mail is a kind of message-switched communication, but we do not see any network that actually can be called a message-switched network.

CIRCUIT-SWITCHED NETWORKS:

A **circuit-switched network** consists of a set of switches connected by physical links. A connection between two stations is a dedicated path made of one or more links. However, each connection uses only one dedicated channel on each link. Each link is normally divided into *n* channels by using FDM or TDM.

Figure 1.42 shows a trivial circuit-switched network with four switches and four links. Each link is divided into *n* (*n* is 3 in the figure) channels by using FDM or TDM.

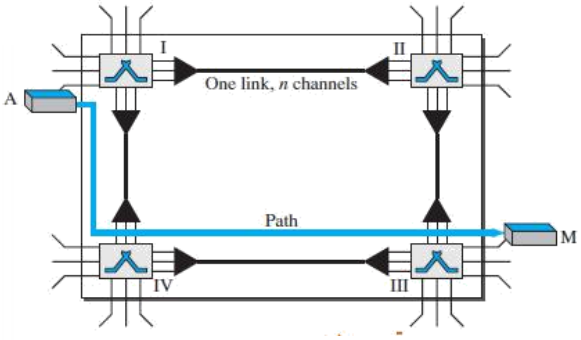


FIGURE 1.42: A TRIVIAL CIRCUIT-SWITCHED NETWORK

We have explicitly shown the multiplexing symbols to emphasize the division of the link into channels even though multiplexing can be implicitly included in the switch fabric. The end systems, such as computers or telephones, are directly connected to a switch. We have shown only two end systems for simplicity.

When end system A needs to communicate with end system M, system A needs to request a connection to M that must be accepted by all switches as well as by M itself. This is called the **setup phase;** a circuit (channel) is reserved on each link, and the combination of circuits or channels defines the dedicated path.

After the dedicated path made of connected circuits (channels) is established, the **data-transfer phase** can take place. After all data have been transferred, the circuits are torn down.

Three Phases:

The actual communication in a circuit-switched network requires three phases: connection setup, data transfer, and connection teardown.

Setup Phase: Before the two parties (or multiple parties in a conference call) can communicate, a dedicated circuit needs to be established. The end systems are normally connected through dedicated lines to the switches, so connection setup means creating dedicated channels between the switches.

Data Transfer Phase: After the establishment of the dedicated circuit (channels), the two parties can transfer data.

Teardown Phase: When one of the parties needs to disconnect, a signal is sent to each switch to release the resources.

Efficiency:

It can be argued that circuit-switched networks are not as efficient as the other two types of networks because resources are allocated during the entire duration of the connection. These resources are unavailable to other connections. In a telephone network, people normally terminate the communication when they have finished their conversation.

However, in computer networks, a computer can be connected to another computer even if there is no activity for a long time. In this case, allowing resources to be dedicated means that other connections are deprived.

Delay:

Although a circuit-switched network normally has low efficiency, the delay in this type of network is minimal. During data transfer the data are not delayed at each switch; the resources are allocated for the duration of the connection.

PACKET SWITCHING:

In data communications, we need to send messages from one end system to another. If the message is going to pass through a **packet-switched network**, it needs to be divided into packets of fixed or variable size.

The size of the packet is determined by the network and the governing protocol. In packet switching, there is no resource allocation for a packet. This means that there is no reserved bandwidth on the links, and there is no scheduled processing time for each packet. Resources are allocated on demand. The allocation is done on a first come, first-served basis.

When a switch receives a packet, no matter what the source or destination is, the packet must wait if there are other packets being processed. As with other systems in our daily life, this lack of reservation may create delay. For example, if we do not have a reservation at a restaurant, we might have to wait.

We can have two types of packet-switched networks: datagram networks and virtual circuit networks.

DATAGRAM NETWORKS:

In a **datagram network**, each packet is treated independently of all others. Even if a packet is part of a multipacket transmission, the network treats it as

though it existed alone. Packets in this approach are referred to as ***datagrams***. Datagram switching is normally done at the network layer.

Figure 1.43 shows how the datagram approach is used to deliver four packets from station A to station X. The switches in a datagram network are traditionally referred to as routers. That is why we use a different symbol for the switches in the figure.

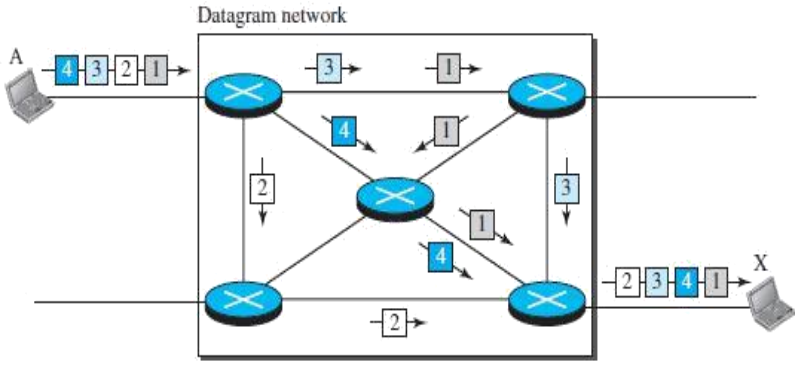


FIGURE 1.43: A DATAGRAM NETWOR WITH FOUR SWITCHES (ROUTERS)

The datagram networks are sometimes referred to as *connectionless networks*. The term *connectionless* here means that the switch (packet switch) does not keep information about the connection state. There are no setup or teardown phases. Each packet is treated the same by a switch regardless of its source or destination.

ROUTING TABLE:

A switch in a datagram network uses a routing table that is based on the destination address. The routing tables are dynamic and are updated periodically. The destination addresses and the corresponding forwarding output ports are recorded in the tables.

Destination Address: Every packet in a datagram network carries a header that contains, among other information, the destination address of the packet.

When the switch receives the packet, this destination address is examined; the routing table is consulted to find the corresponding port through which the packet should be forwarded. This address, unlike the address in a virtual-circuit network, remains the same during the entire journey of the packet.

Efficiency: The efficiency of a datagram network is better than that of a circuit-switched network; resources are allocated only when there are packets to be transferred. If a source sends a packet and there is a delay of a few minutes before another packet can be sent, the resources can be reallocated during these minutes for other packets from other sources.

Delay: There may be greater delay in a datagram network than in a virtual-circuit network. Although there are no setup and teardown phases, each packet may experience a wait at a switch before it is forwarded.

VIRTUAL-CIRCUIT NETWORKS:

A **virtual-circuit network** is a cross between a circuit-switched network and a datagram network. It has some characteristics of both.

As in a circuit-switched network, there are setup and teardown phases in addition to the data transfer phase. Resources can be allocated during the setup phase, as in a circuit-switched network, or on demand, as in a datagram network.

As in a datagram network, data are packetized and each packet carries an address in the header.

As in a circuit-switched network, all packets follow the same path established during the connection. A virtual-circuit network is normally implemented in the data-link layer, while a circuit-switched network is implemented in the physical layer and a datagram network in the network layer. But this may change in the future.

Figure 1.44 is an example of a virtual-circuit network. The network has switches that allow traffic from sources to destinations, can be a computer, packet switch, bridge, or any other device that connects other networks.

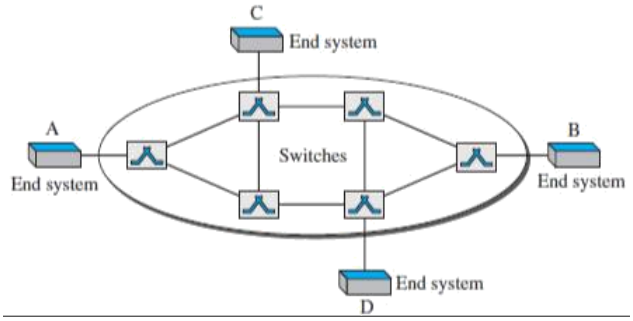


FIGURE 1.44: VIRTUAL-CIRCUIT NETWORK

Addressing: In a virtual-circuit network, two types of addressing are involved: global and local (virtual-circuit identifier).

Global Addressing: A source or a destination needs to have a global address—an address that can be unique in the scope of the network or internationally if the network is part of an international network. However, we will see that a global address in virtual-circuit networks is used only to create a virtual-circuit identifier.

Virtual-Circuit Identifier:

The identifier that is actually used for data transfer is called the ***virtual-circuit identifier*** **(VCI)** or the ***label*****.** A VCI, unlike a global address, is a small number that has only switch scope; it is used by a frame between two switches. When a frame arrives at a switch, it has a VCI; when it leaves, it has a different VCI.

Figure 1.45 shows how the VCI in a data frame changes from one switch to another. Note that a VCI does not need to be a large number since each switch can use its own unique set of VCIs.

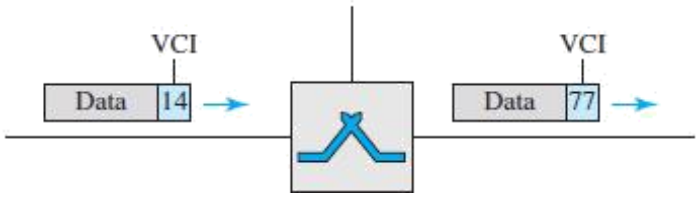


FIGURE 1.45: VIRTUAL-CIRCUIT IDENTIFIER

Three Phases:

As in a circuit-switched network, a source and destination need to go through three phases in a virtual-circuit network: setup, data transfer, and teardown.

Data-Transfer Phase: To transfer a frame from a source to its destination, all switches need to have a table entry for this virtual circuit. The table, in its simplest form, has four columns. This means that the switch holds four pieces of information for each virtual circuit that is already set up. Figure 1.46 shows such a switch and its corresponding table.

Setup Phase: In the setup phase, a switch creates an entry for a virtual circuit. Two steps are required: the ***setup request*** and the ***acknowledgment***.

Setup Request: A setup request frame is sent from the source to the destination. Figure 1.47 shows the process.

Acknowledgment: A special frame, called the *acknowledgment frame,* completes the entries in the switching tables. Figure 1.48 shows the process.

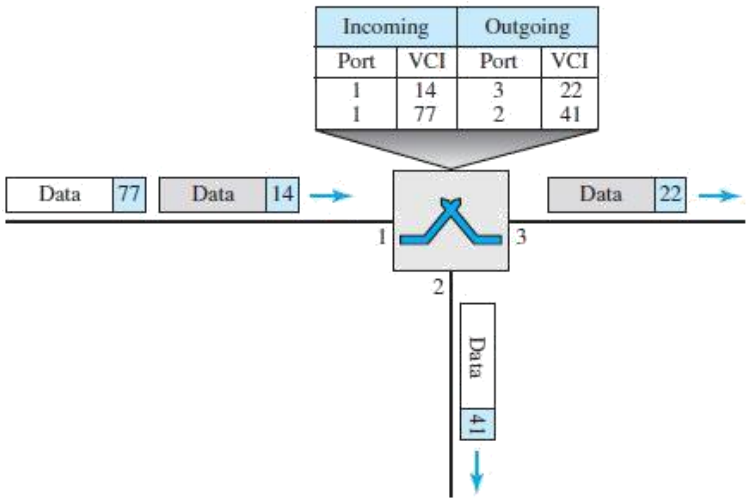


FIGURE 1.46: SWITCH AND TABLES IN A VIRTUAL-CIRCUIT NETWORK

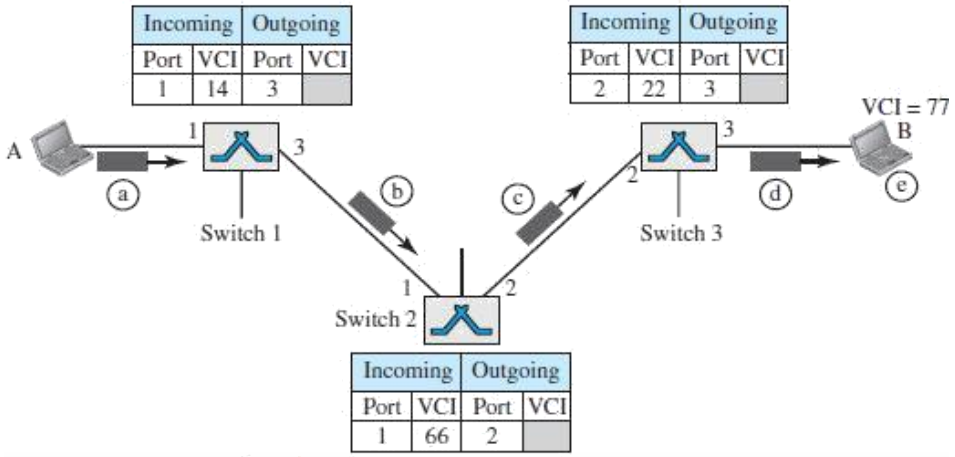


FIGURE 1.47: SETUP REQUEST IN A VIRTUAL-CIRCUIT NETWORK

Teardown Phase: In this phase, all switches delete the corresponding entry from their tables after sending all frames from one to another.

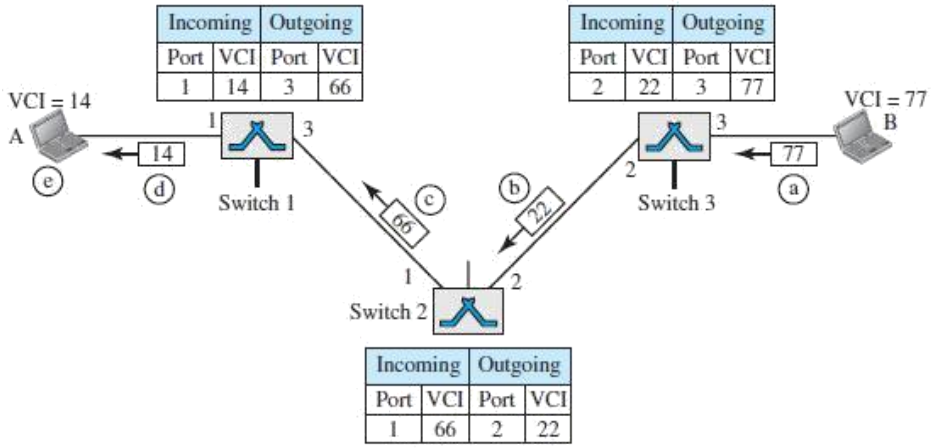


FIGURE 1.48: SETUP ACKNOWLEDGMENT IN A VIRTUAL-CIRCUIT NETWORK

Efficiency: As we said before, resource reservation in a virtual-circuit network can be made during the setup or can be on demand during the data-transfer phase.

In the first case, the delay for each packet is the same; in the second case, each packet may encounter different delays.

There is one big advantage in a virtual-circuit network even if resource allocation is on demand. The source can check the availability of the resources, without actually reserving it.

Delay in Virtual-Circuit Networks:

In a virtual-circuit network, there is a one-time delay for setup and a one-time delay for teardown. If resources are allocated during the setup phase, there is no wait time for individual packets.

Circuit-Switched Technology in WANs: virtual-circuit networks are used in switched WANs such as ATM networks. The data-link layer of these technologies is well suited to the virtual circuit technology.

Switching at the data-link layer in a switched WAN is normally implemented by using virtual-circuit techniques.