Unit 1

Introduction

Prof. Keyur J Patel

Outline

- Understanding of Network and Internet,
- Network Topologies,
- The OSI Model,
- TCP/IP Protocol Suite,
- Guided and Unguided Transmission Media,
- Network Devices,
- Fundamental of Circuit-Switched and Packet-Switched Networks,
- Performance Metrics,
- Understanding of Delay,
- Loss and Throughput in the packet-switching network

Understanding of Network

- The term *telecommunication* means communication at a distance.
- The word *data* refers to information presented in whatever form is agreed upon by the parties creating and using the data.
- Data communications are the exchange of data between two devices via some form of transmission medium such as a wire cable.

Understanding of Network

- The effectiveness of a data communications system depends on four fundamental characteristics: delivery, accuracy, timeliness, and jitter.

1. **Delivery.** The system must deliver data to the correct destination. Data must be received by the intended device or user and only by that device or user.

2. **Accuracy.** The system must deliver the data accurately. Data that have been altered in transmission and left uncorrected are unusable.

Understanding of Network

3. **Timeliness.** The system must deliver data in a timely manner. Data delivered late are useless. In the case of video and audio, timely delivery means delivering data as they are produced, in the same order that they are produced, and without significant delay. This kind of delivery is called real-time transmission.

4. **Jitter.** Jitter refers to the variation in the packet arrival time. It is the uneven delay in the delivery of audio or video packets. For example, let us assume that video packets are sent every 3D ms. If some of the packets arrive with 3D-ms delay and others with 4D-ms delay, an uneven quality in the video is the result.

Understanding of Network

- Components of Data Communication

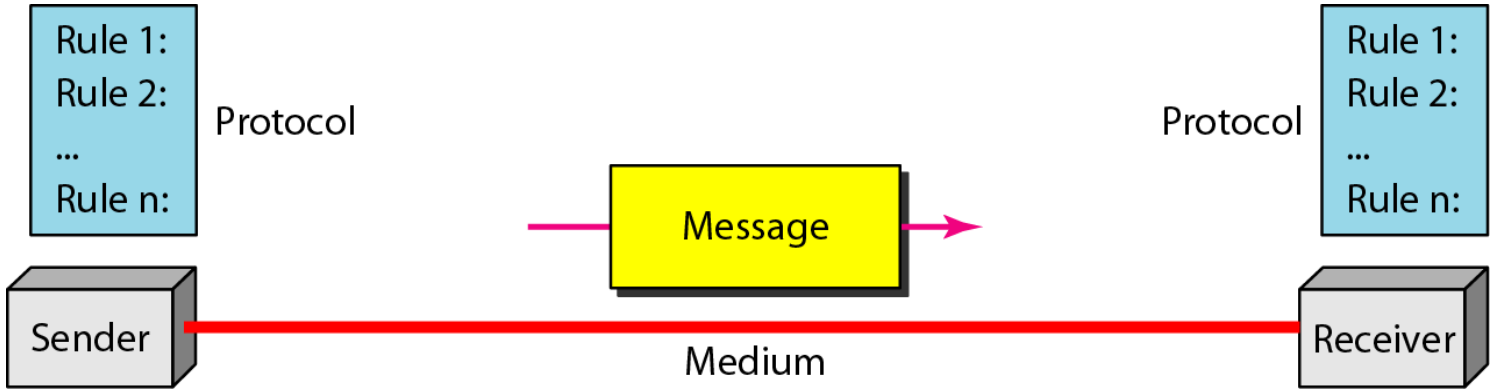


Figure **Five components of data communication**

Understanding of Network

1. **Message.** The message is the information (data) to be communicated. Popular forms of information include text, numbers, pictures, audio, and video.

2. **Sender.** The sender is the device that sends the data message. It can be a computer, workstation, telephone handset, video camera, and so on.

3. **Receiver.** The receiver is the device that receives the message. It can be a computer, workstation, telephone handset, television, and so on.

Understanding of Network

4. **Transmission medium.** The transmission medium is the physical path by which a message travels from sender to receiver. Some examples of transmission media include twisted-pair wire, coaxial cable, fiber-optic cable, and radio waves.

5. **Protocol.** A protocol is a set of rules that govern data communications. It represents an agreement between the communicating devices. Without a protocol, two devices may be connected but not communicating, just as a person speaking French cannot be understood by a person who speaks only Japanese.

Understanding of Network

Data Representation:

- Information today comes in different forms such as text, numbers, images, audio, and video.

Text.

- In data communications, text is represented as a bit pattern, a sequence of bits (0s or 1s). Different sets of bit patterns have been designed to represent text symbols. Each set is called a code, and the process of representing symbols is called coding.
- Today, the prevalent coding system is called Unicode, which uses 32 bits to represent a symbol or character used in any language in the world.

Understanding of Network

- The American Standard Code for Information Interchange (ASCII), developed some decades ago in the United States, now constitutes the first 127 characters in Unicode and is also referred to as Basic Latin.

Numbers

- Numbers are also represented by bit patterns. However, a code such as ASCII is not used to represent numbers; the number is directly converted to a binary number to simplify mathematical operations.

Understanding of Network

Images

- Images are also represented by bit patterns. In its simplest form, an image is composed of a matrix of pixels (picture elements), where each pixel is a small dot. The size of the pixel depends on the resolution.
- For example, an image can be divided into 1000 pixels or 10,000 pixels. In the second case, there is a better representation of the image (better resolution), but more memory is needed to store the image. After an image is divided into pixels, each pixel is assigned a bit pattern. The size and the value of the pattern depend on the image. For an image made of only black-and-white dots (e.g., a chessboard), a 1-bit pattern is enough to represent a pixel.

Understanding of Network

Audio

- Audio refers to the recording or broadcasting of sound or music. Audio is by nature different from text, numbers, or images. It is continuous, not discrete.
- Even when we use a microphone to change voice or music to an electric signal, we create a continuous signal.

Video

- Video refers to the recording or broadcasting of a picture or movie.
- Video can either be produced as a continuous entity (e.g., by a TV camera), or it can be a combination of images, each a discrete entity, arranged to convey the idea of motion.

Understanding of Network

Data Flow

- Communication between two devices can be simplex, half-duplex, or full-duplex as shown in Figure.

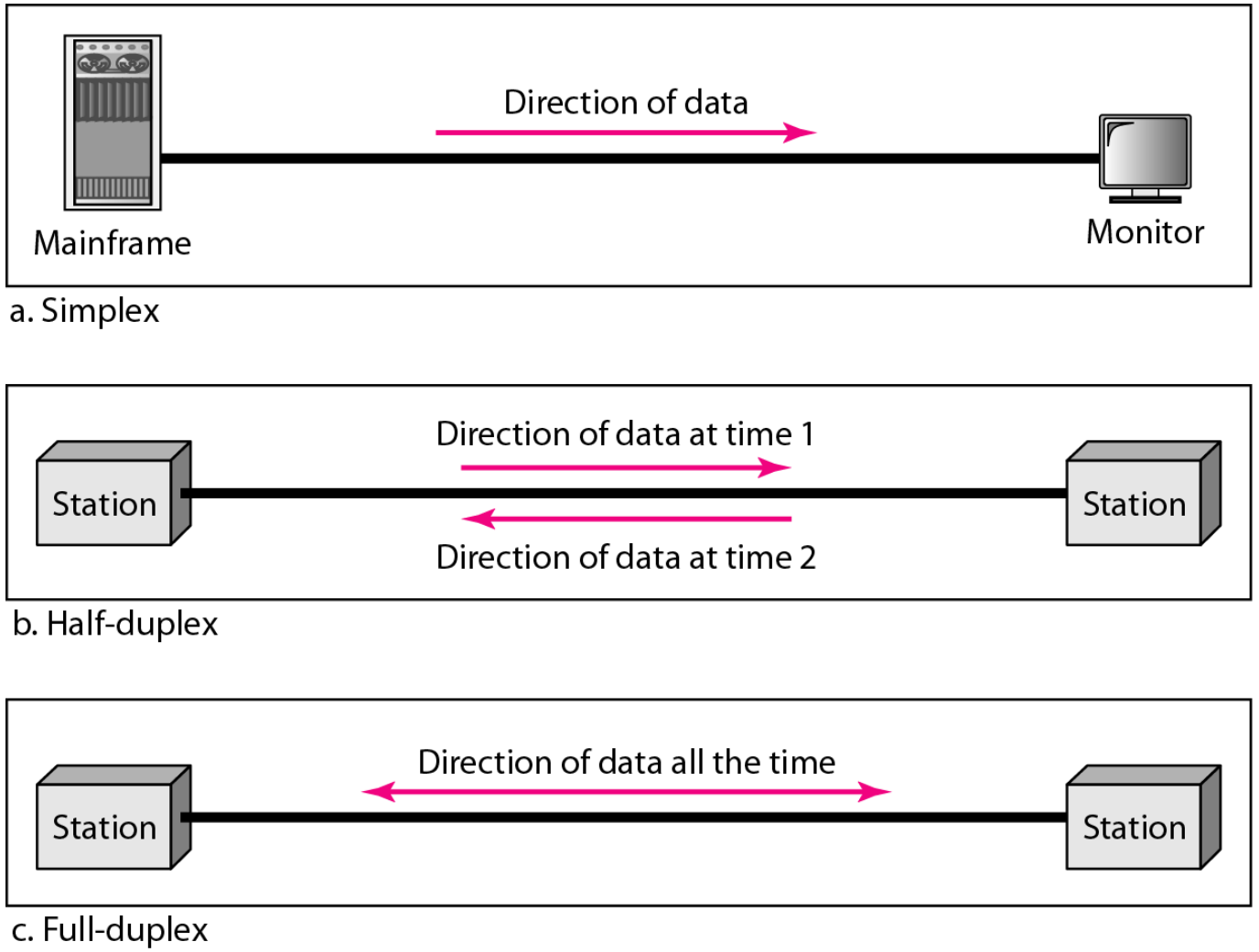


Figure **Data flow (simplex, half-duplex, and full-duplex)**

Understanding of Network

Simplex

- In simplex mode, the communication is unidirectional, as on a one-way street.
- Only one of the two devices on a link can transmit; the other can only receive (see Figure a).
- Keyboards and traditional monitors are examples of simplex devices. The keyboard can only introduce input; the monitor can only accept output. The simplex mode can use the entire capacity of the channel to send data in one direction.

Understanding of Network

Half-Duplex

- In half-duplex mode, each station can both transmit and receive, but not at the same time. When one device is sending, the other can only receive, and vice versa (see Figure b).
- The half-duplex mode is like a one-lane road with traffic allowed in both directions. When cars are traveling in one direction, cars going the other way must wait. In a half-duplex transmission, the entire capacity of a channel is taken over by whichever of the two devices is transmitting at the time.
- Walkie-talkies and CB (citizens band) radios are both half-duplex systems. The half-duplex mode is used in cases where there is no need for communication in both directions at the same time; the entire capacity of the channel can be utilized for each direction.

Understanding of Network

Full-Duplex

- In full-duplex mode (also called duplex), both stations can transmit and receive simultaneously. The full-duplex mode is like a two-way street with traffic flowing in both directions at the same time.
- In full-duplex mode, signals going in one direction share the capacity of the link: with signals going in the other direction.
- This sharing can occur in two ways: Either the link must contain two physically separate transmission paths, one for sending and the other for receiving; or the capacity of the channel is divided between signals traveling in both directions.
- One common example of full-duplex communication is the telephone network. When two people are communicating by a telephone line, both can talk and listen at the same time. The full-duplex mode is used when communication in both directions is required all the time. The capacity of the channel, however, must be divided between the two directions.

Understanding of Network

Network

- A network is a set of devices (often referred to as nodes) connected by communication links.
- A node can be a computer, printer, or any other device capable of sending and/or receiving data generated by other nodes on the network.

Understanding of Network

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

Understanding of 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.

Understanding of Network

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.

Understanding of Network

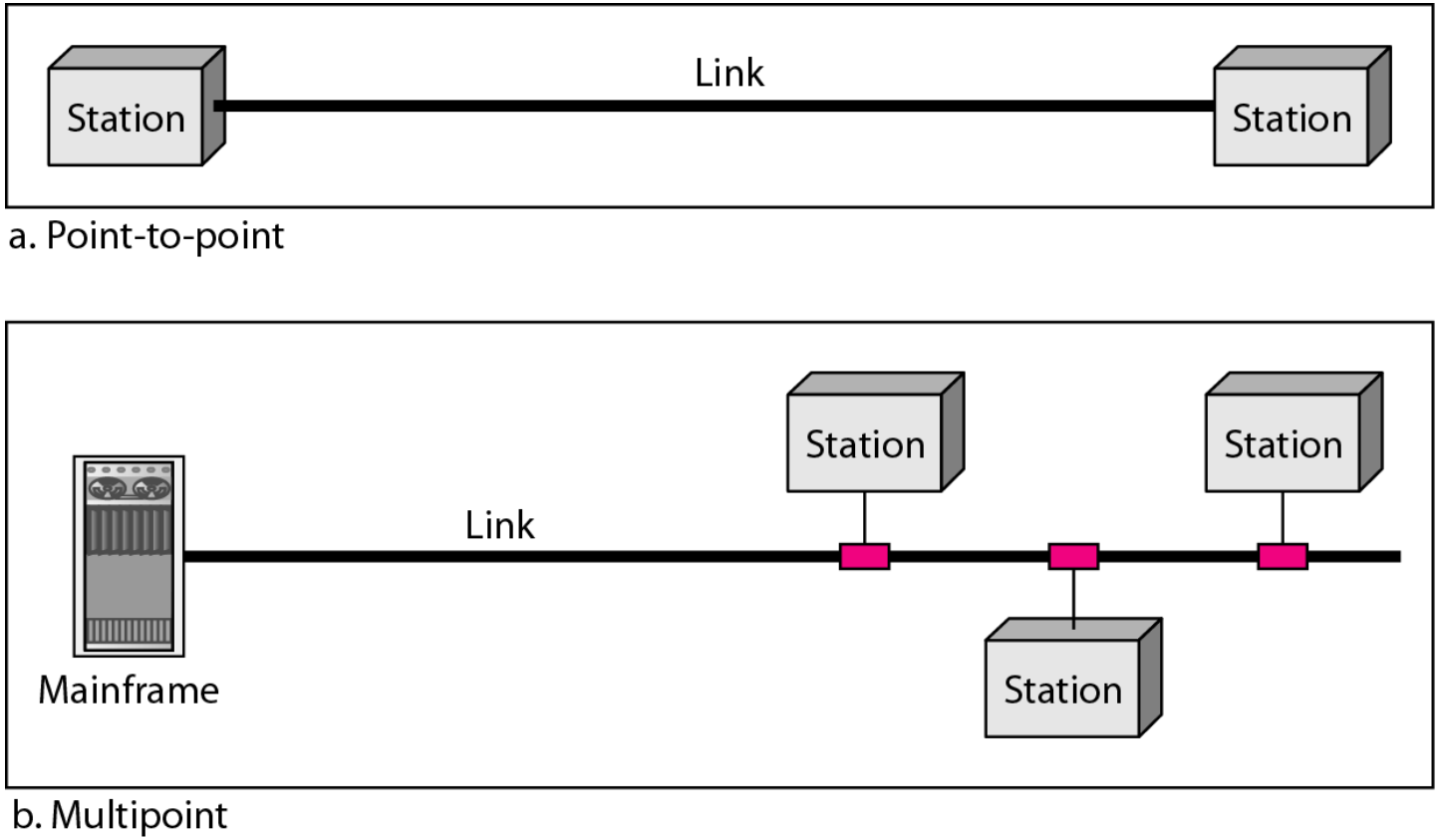


Figure **Types of connections: point-to-point and multipoint**

Understanding of Network

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 a).
- When you change television channels by infrared remote control, you are establishing a point-to-point connection between the remote control and the television's control system.

Understanding of Network

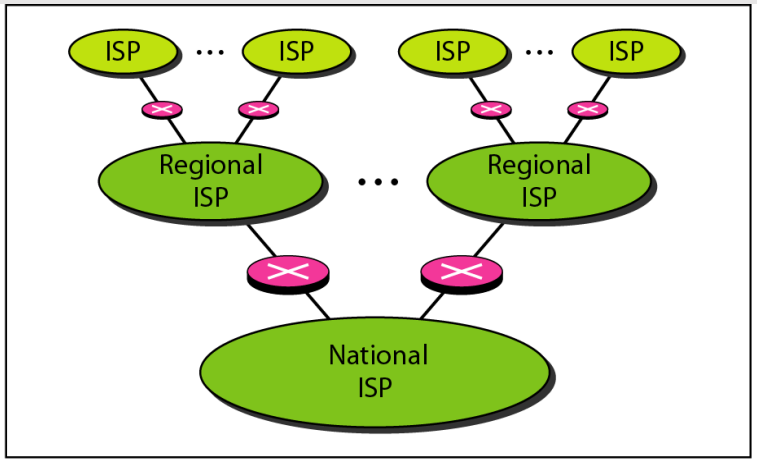
Multipoint

- A multipoint (also called multidrop) connection is one in which more than two specific devices share a single link (see Figure b).
- 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

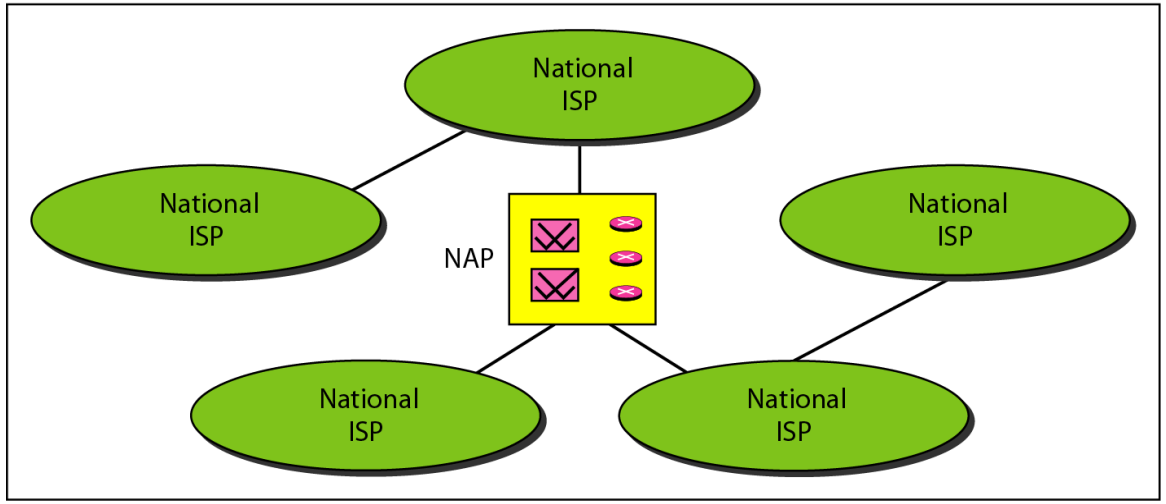
Internet

- It is very rare to see a LAN, a MAN, or a LAN in isolation; they are connected to one another. When two or more networks are connected, they become an **internetwork**, or **internet**.
- An internet is two or more networks that can communicate with each other. The most notable internet, a collaboration of more than hundreds of thousands of interconnected networks.
- Private individuals as well as various organizations such as government agencies, schools, research facilities, corporations, and libraries in more than 100 countries use the Internet. Millions of people are users.

Internet



a. Structure of a national ISP



b. Interconnection of national ISPs

Network Topologies

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 (usually called nodes) to one another. There are four basic topologies possible: mesh, star, bus, and ring

Network Topologies

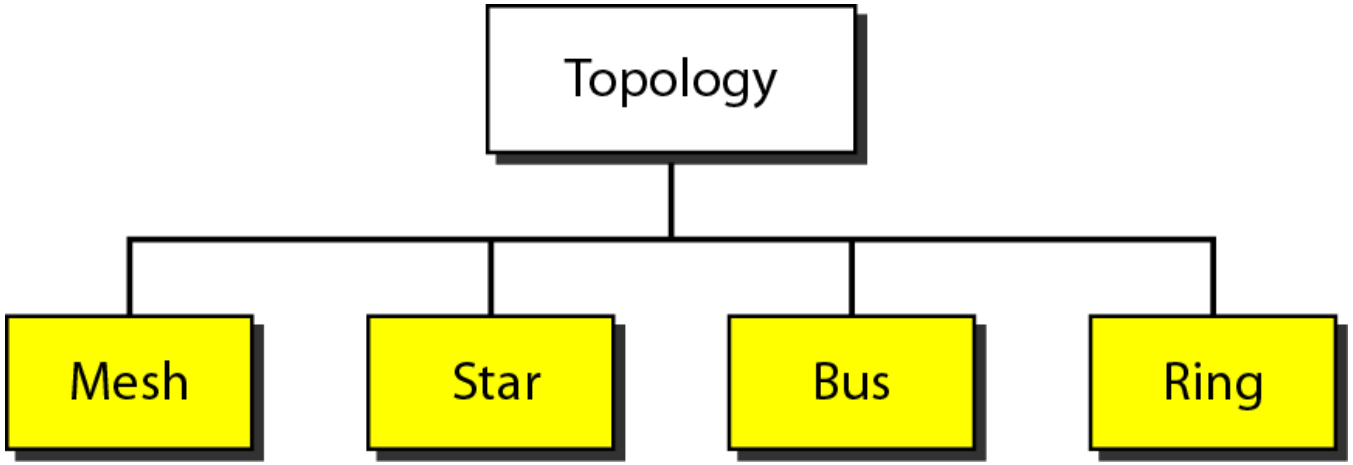


Figure Categories of topology

Network Topologies

Mesh

- 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.

Network Topologies

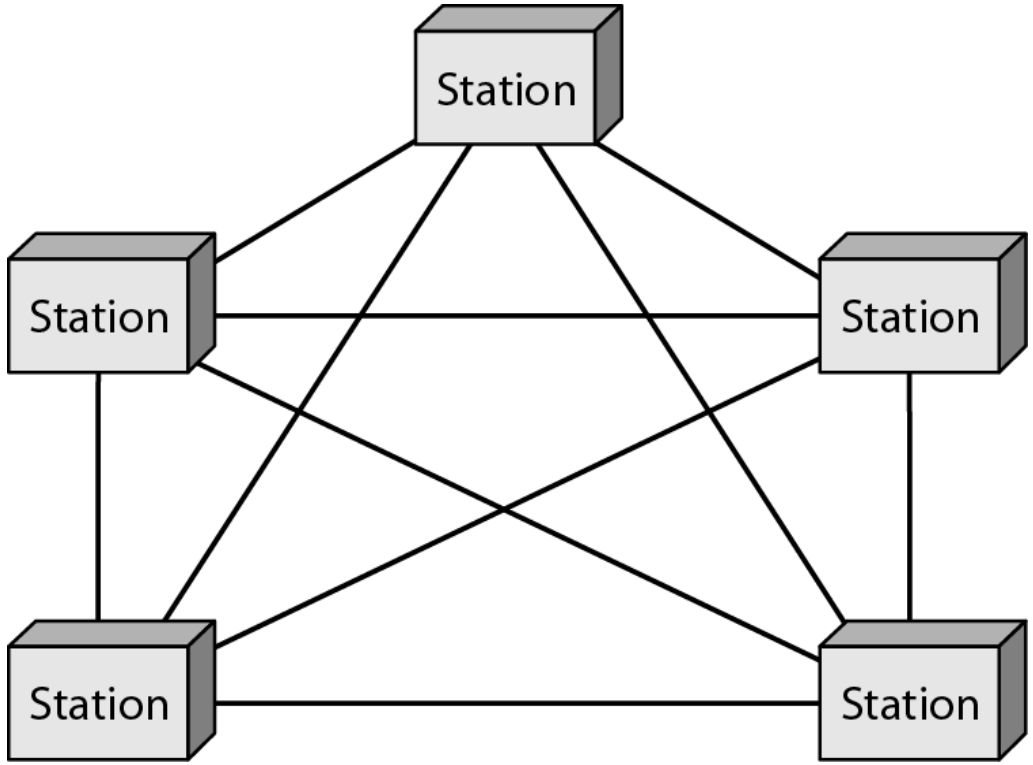


Figure A fully connected mesh topology (five devices)

Network Topologies

- A mesh offers several **advantages** over other network topologies.
- 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.
- Second, a mesh topology is robust. If one link becomes unusable, it does not incapacitate the entire system.
- 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

Network Topologies

- The main **disadvantages** of a mesh are related to the amount of cabling and the number of I/O ports required.
- First, because every device must be connected to every other device, installation and reconnection are difficult.
- Second, the sheer bulk of the wiring can be greater than the available space (in walls, ceilings, or floors) can accommodate.

- One practical **example** of a mesh topology is the connection of telephone regional offices in which each regional office needs to be connected to every other regional office.

Network 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) .

Network Topologies

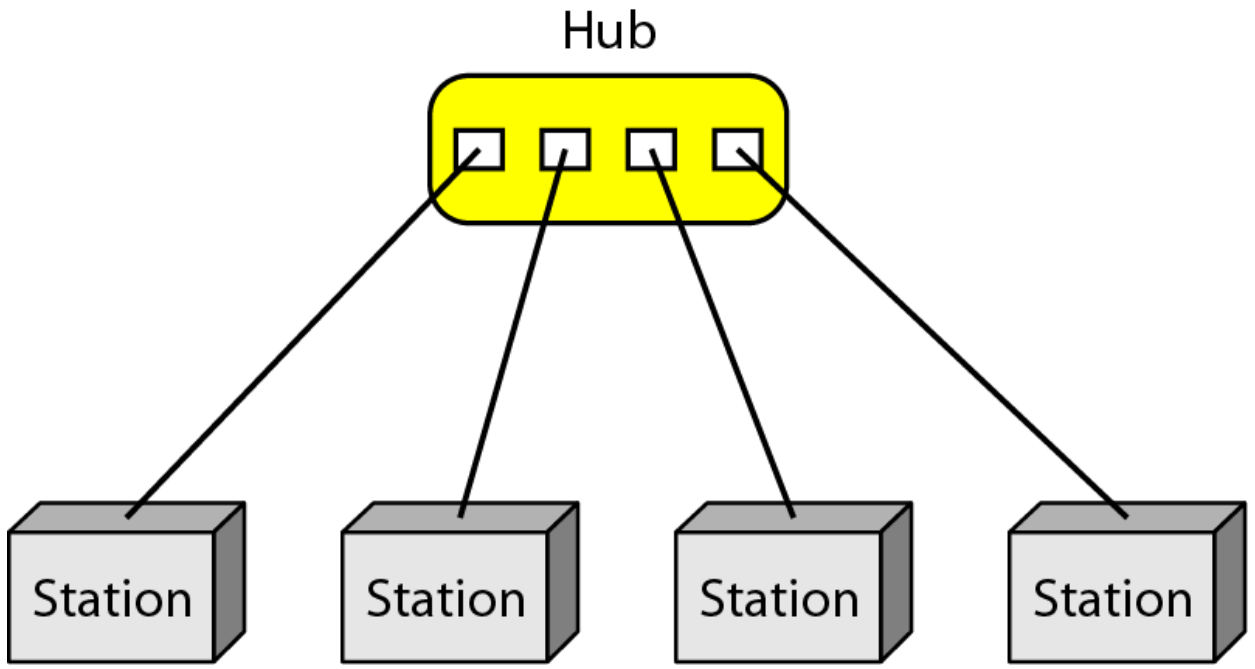


Figure **A star topology connecting four stations**

Network Topologies

- 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.

Network Topologies

- 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.

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

Network Topologies

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).

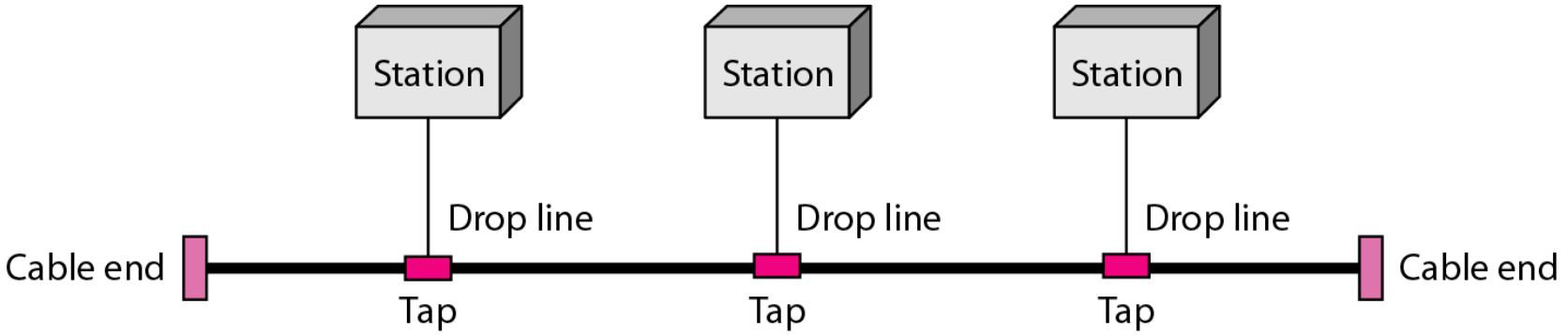


Figure A bus topology connecting three stations

Network Topologies

- 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.

Network Topologies

- **Advantages** of a bus topology include ease of installation.
- Backbone cable can be laid along the most efficient path, 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.

Network Topologies

- 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. Ethernet LANs can use a bus topology, but they are less popular now for reasons.

Network Topologies

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).

Network Topologies

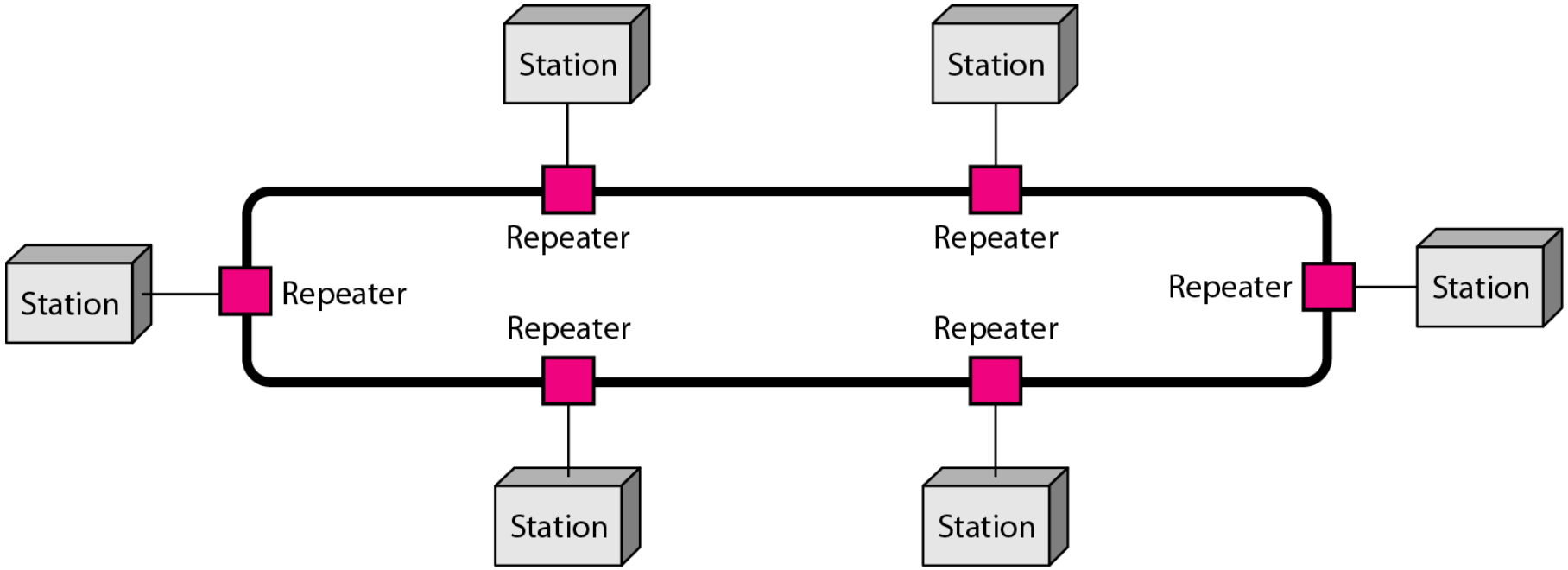


Figure A ring topology connecting six stations

Network Topologies

- 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.

Network Topologies

- 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 Topologies

Hybrid Topology

- A network can be hybrid.
- For example, we can have a main star topology with each branch connecting several stations in a bus top.

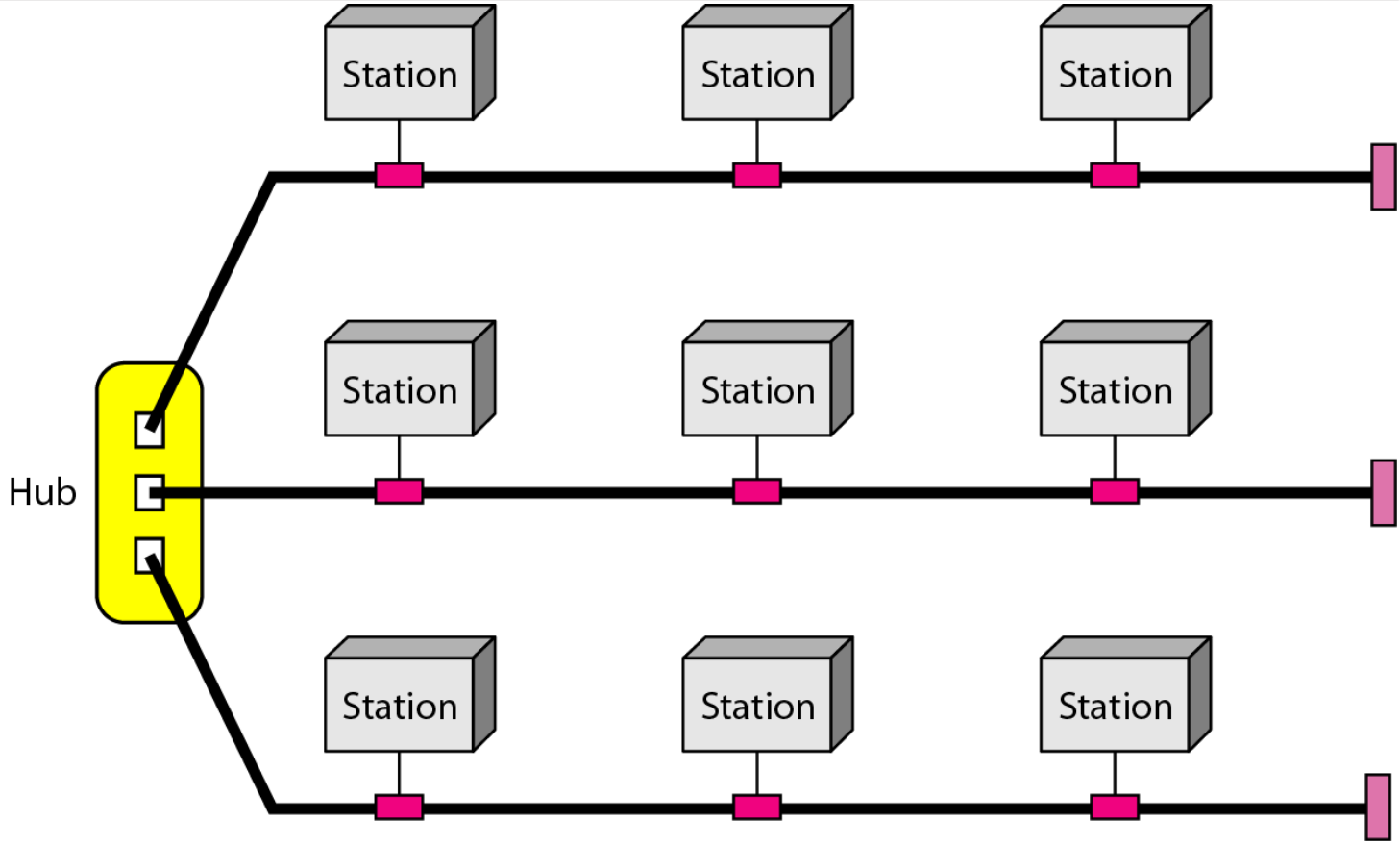


Figure A hybrid topology: a star backbone with three bus networks

Categories of Networks

Categories of Networks

- Today when we speak of networks, we are generally referring to two primary categories: local-area networks and wide-area networks.
- The category into which a network falls is determined by its size. A LAN normally covers an area less than 2 miles, a WAN can be worldwide.
- Networks of a size in between are normally referred to as metropolitan area networks and span tens of miles.

Categories of Networks

Local Area Network

- A local area network (LAN) is usually privately owned and links the devices in a single office, building, or campus (see Figure).
- Depending on the needs of an organization and the type of technology used, a LAN can be as simple as two PCs and a printer in someone's home office; or it can extend throughout a company and include audio and video peripherals.
- Currently, LAN size is limited to a few kilometers.

Categories of Networks

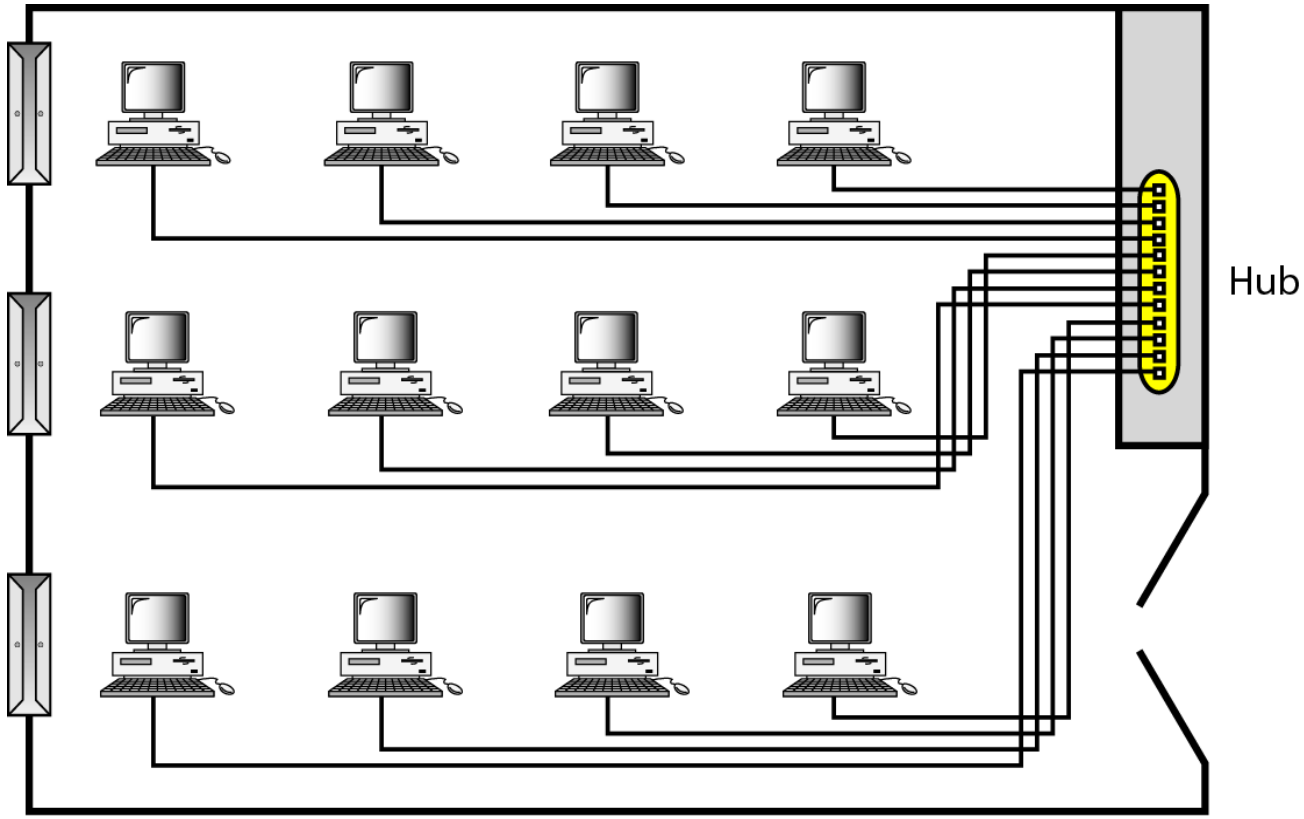


Figure Local Area Network

Categories of Networks

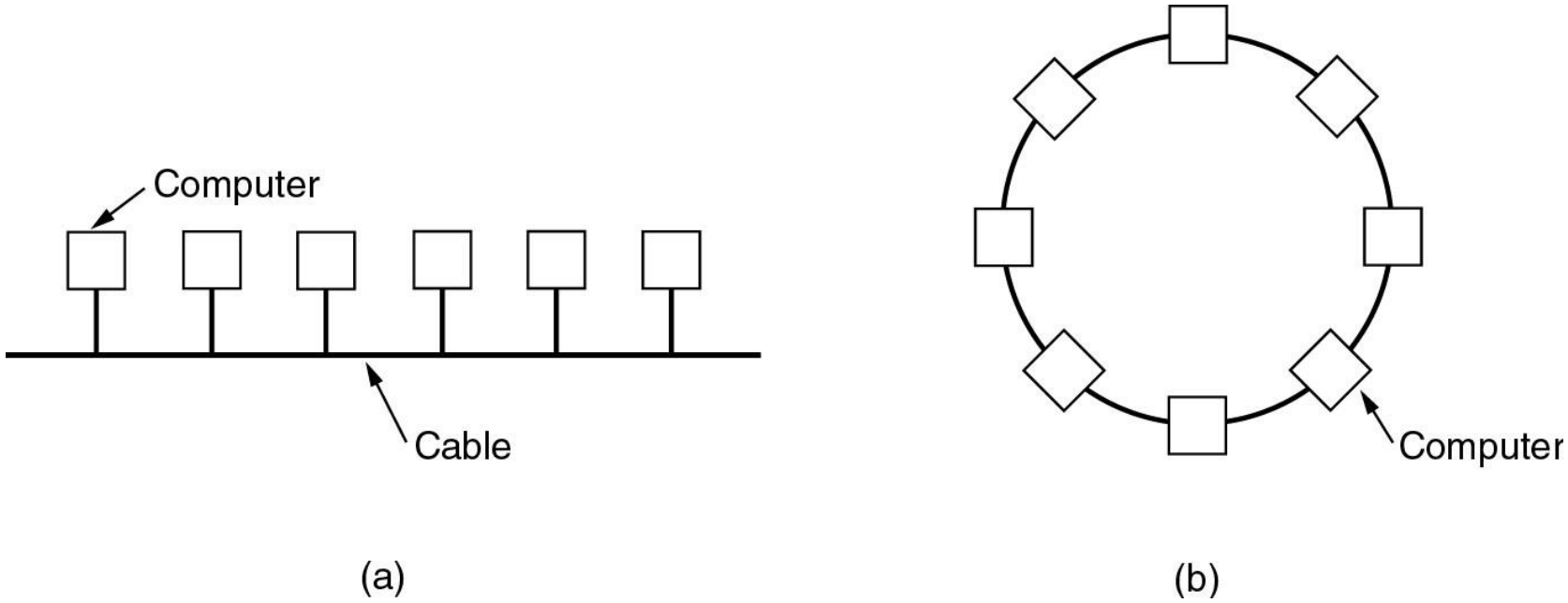


Figure Local Area Network

Categories of Networks

- LANs are designed to allow resources to be shared between personal computers or workstations. The resources to be shared can include hardware (e.g., a printer), software (e.g., an application program), or data.
- A common example of a LAN, found in many business environments, links a workgroup of task-related computers, for example, engineering workstations or accounting PCs.

Categories of Networks

Metropolitan Area Network

- A metropolitan area network (MAN) is a network with a size between a LAN and a WAN. It normally covers the area inside a town or a city. It is designed for customers who need a high-speed connectivity, normally to the Internet, and have endpoints spread over a city or part of city.
- A good example of a MAN is the part of the telephone company network that can provide a high-speed DSL line to the customer.
- Another example is the cable TV network that originally was designed for cable TV, but today can also be used for high-speed data connection to the Internet.

Categories of Networks

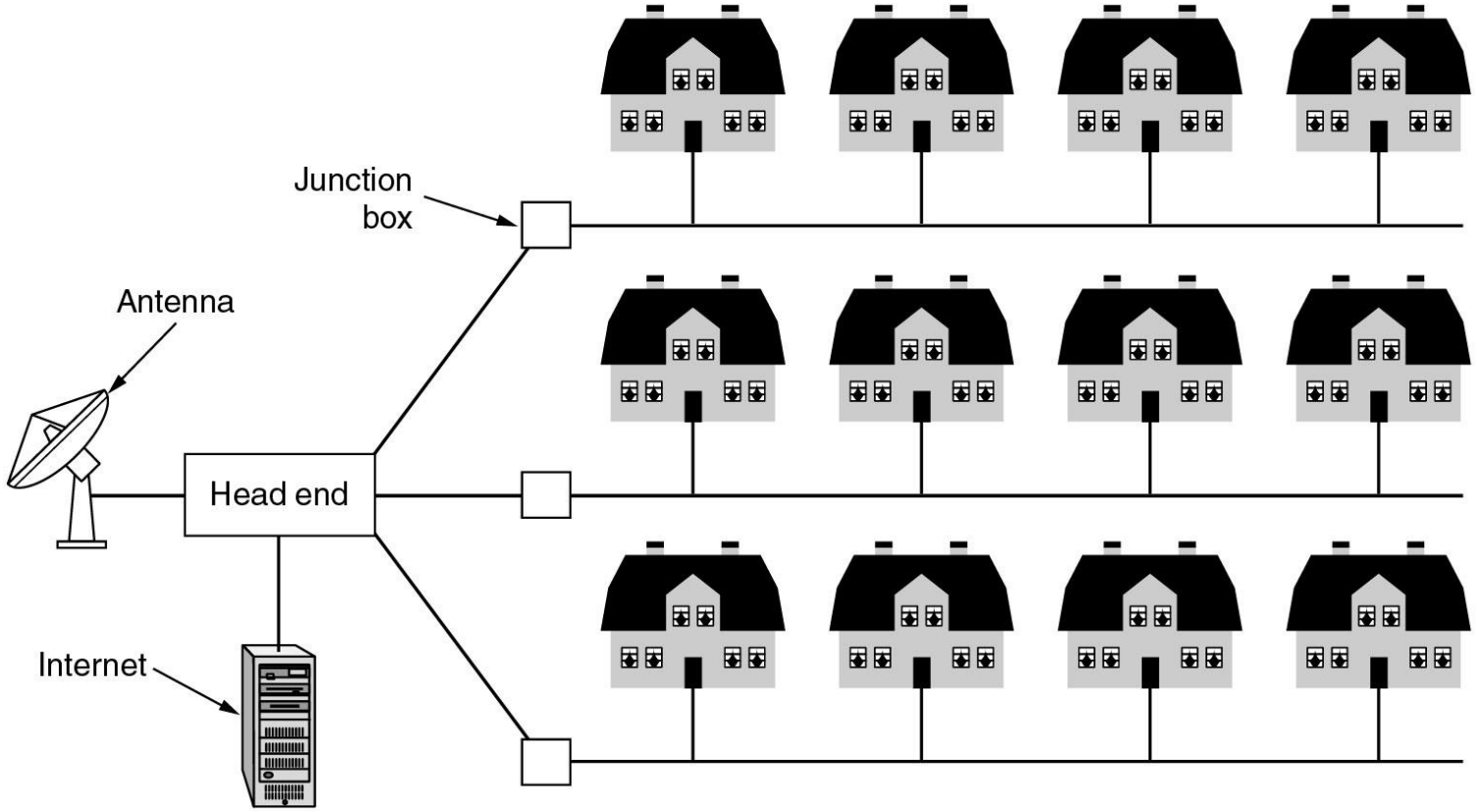


Figure Metropolitan Area Networks

Categories of Networks

Wide Area Network

- A wide area network (WAN) provides long-distance transmission of data, image, audio, and video information over large geographic areas that may comprise a country, a continent, or even the whole world. A WAN can be as complex as the backbones that connect the Internet or as simple as a dial-up line that connects a home computer to the Internet.
- The point-to-point WAN is normally a line leased from a telephone or cable TV provider that connects a home computer or a small LAN to an Internet service provider (ISP). This type of WAN is often used to provide Internet access.
- Example of WANs is the wireless WAN that is becoming more and more popular.

Categories of Networks

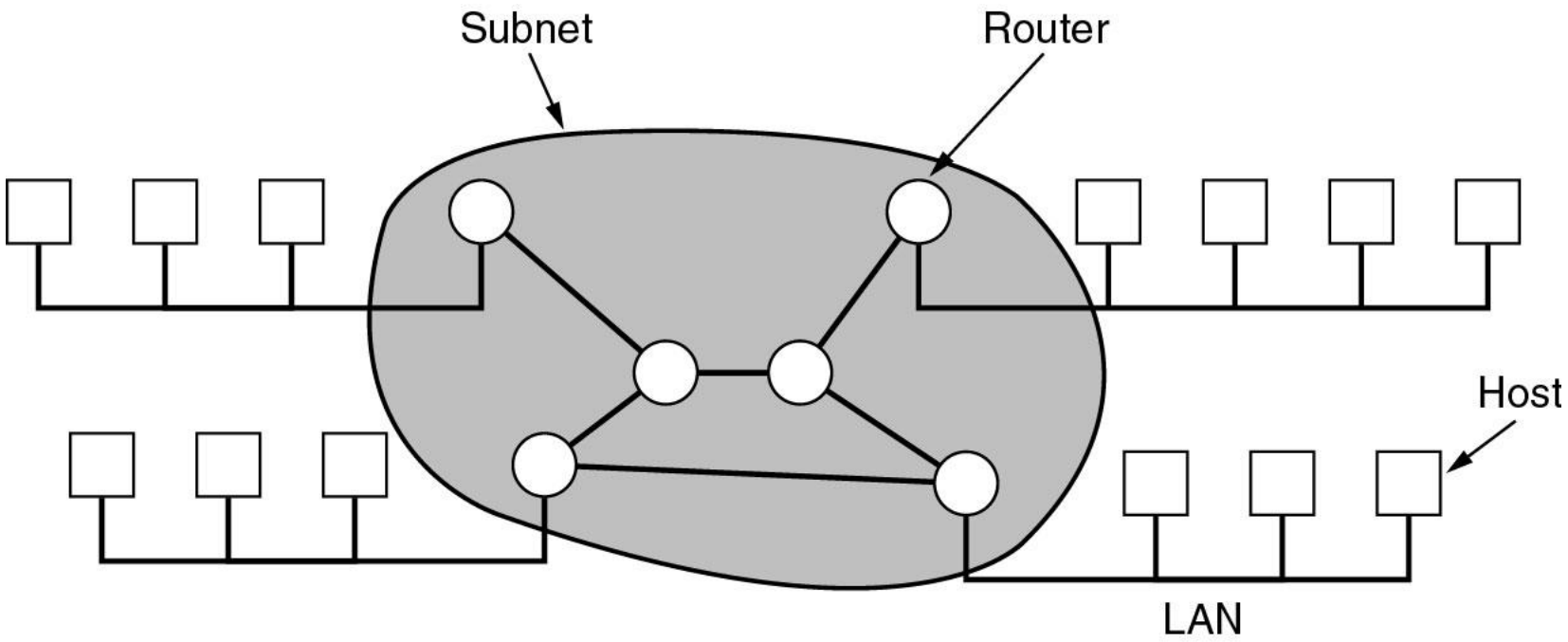


Figure Wide Area Networks

Categories of Networks

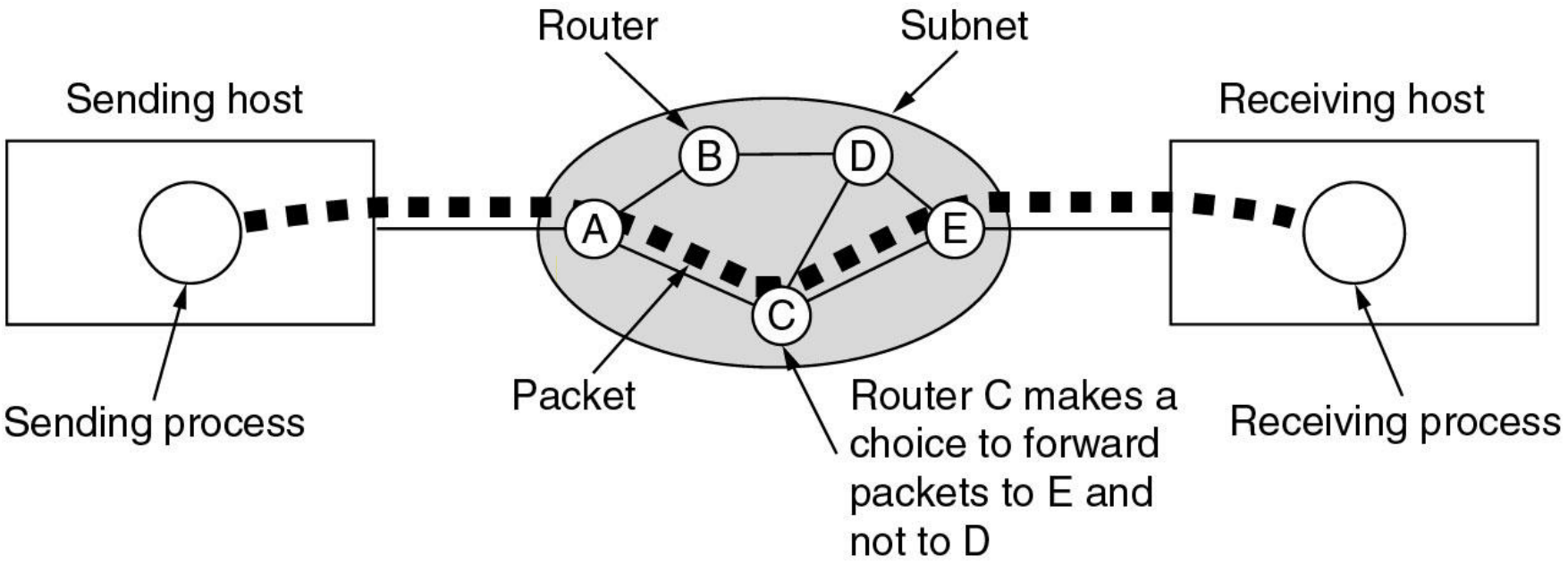


Figure Wide Area Networks

The OSI Model

- the International Standards Organization (ISO) is a multinational body dedicated to worldwide agreement on international standards.
- An ISO standard that covers all aspects of network communications is the Open Systems Interconnection 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

- 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).

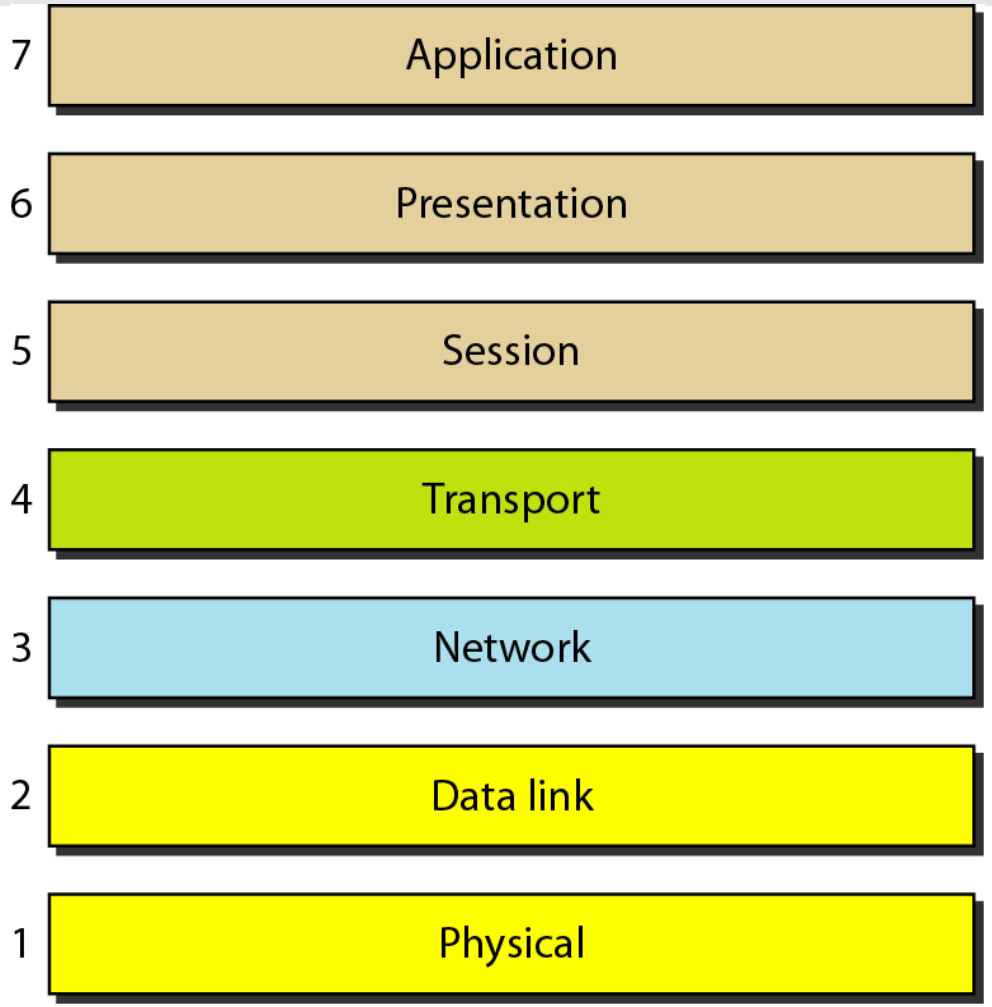


Figure Seven layers of the OSI model

The OSI Model

Layered Architecture

- The OSI model is composed of seven ordered layers: physical (layer 1), data link (layer 2), network (layer 3), transport (layer 4), session (layer 5), presentation (layer 6), and application (layer 7).
- Figure shows the layers involved when a message is sent from device *A* to device *B*. As the message travels from *A* to *B*, it may pass through many intermediate nodes.
- These intermediate nodes usually involve only the first three layers of the OSI model.

The OSI Model

Peer-to-Peer Processes

- At the physical layer, communication is direct: In Figure, device *A* sends a stream of bits to device *B* (through intermediate nodes). At the higher layers, however, communication must move down through the layers on device *A*, over to device *B*, and then back up through the layers.
- Each layer in the sending device adds its own information to the message it receives from the layer just above it and passes the whole package to the layer just below it.

The OSI Model

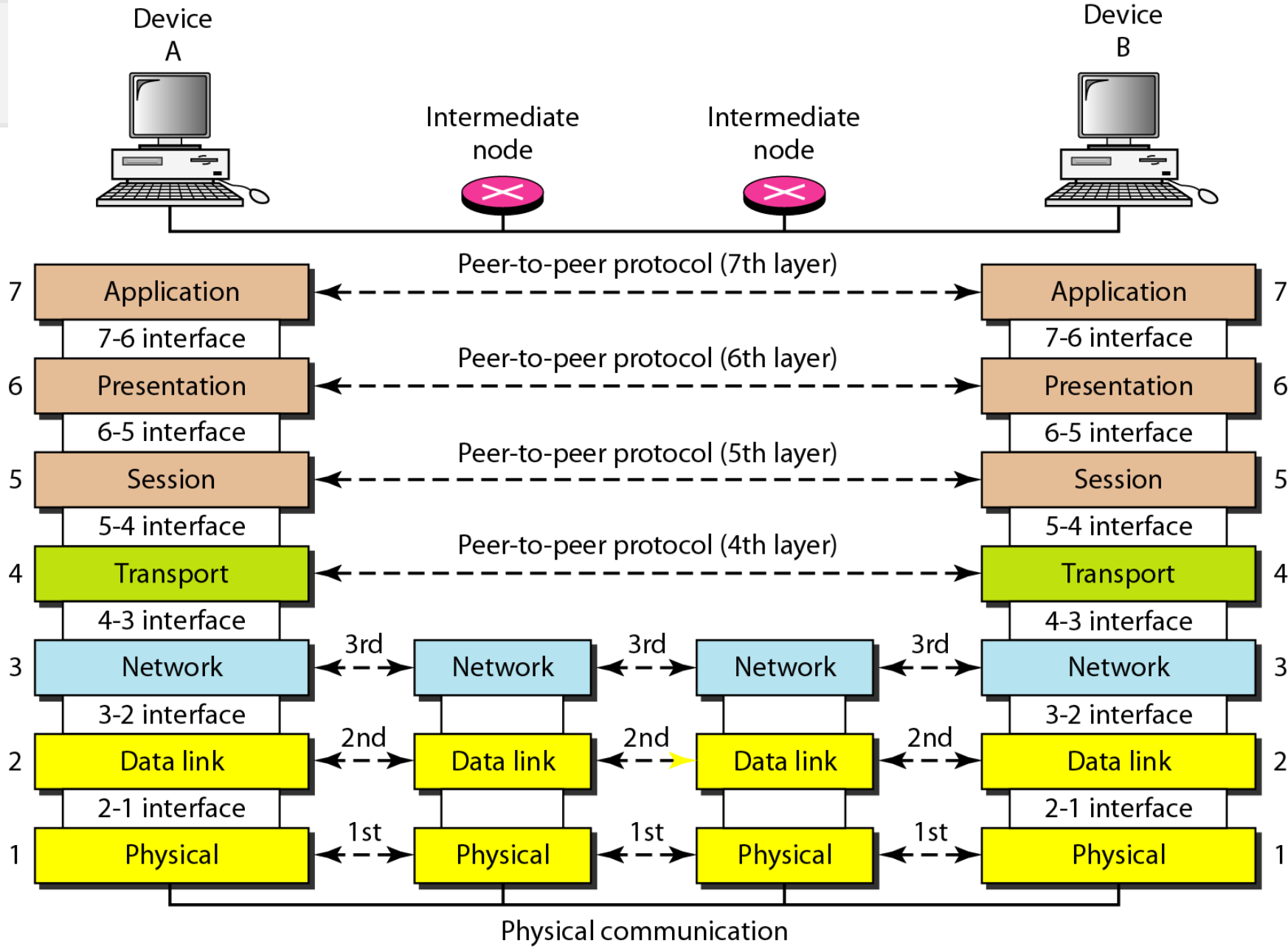


Figure The interaction between layers in the OSI model

Interfaces between Layers

- The passing of the data and network information down through the layers of the sending device and back up through the layers of the receiving device is made possible by an interface between each pair of adjacent layers.
- Each interface defines the information and services a layer must provide for the layer above it. Well-defined interfaces and layer functions provide modularity to a network.

The OSI Model

Organization of the Layers

- The seven layers can be thought of as belonging to three subgroups.
- Layers 1, 2, and 3-physical, data link, and network-are the network support layers; they deal with the physical aspects of moving data from one device to another (such as electrical specifications, physical connections, physical addressing, and transport timing and reliability).
- Layers 5, 6, and 7-session, presentation, and application-can be thought of as the user support layers; they allow interoperability among unrelated software systems.

The OSI Model

- Layer 4, the transport layer, links the two subgroups and ensures that what the lower layers have transmitted is in a form that the upper layers can use.
- The upper OSI layers are almost always implemented in software; lower layers are a combination of hardware and software, except for the physical layer, which is mostly hardware.

The OSI Model

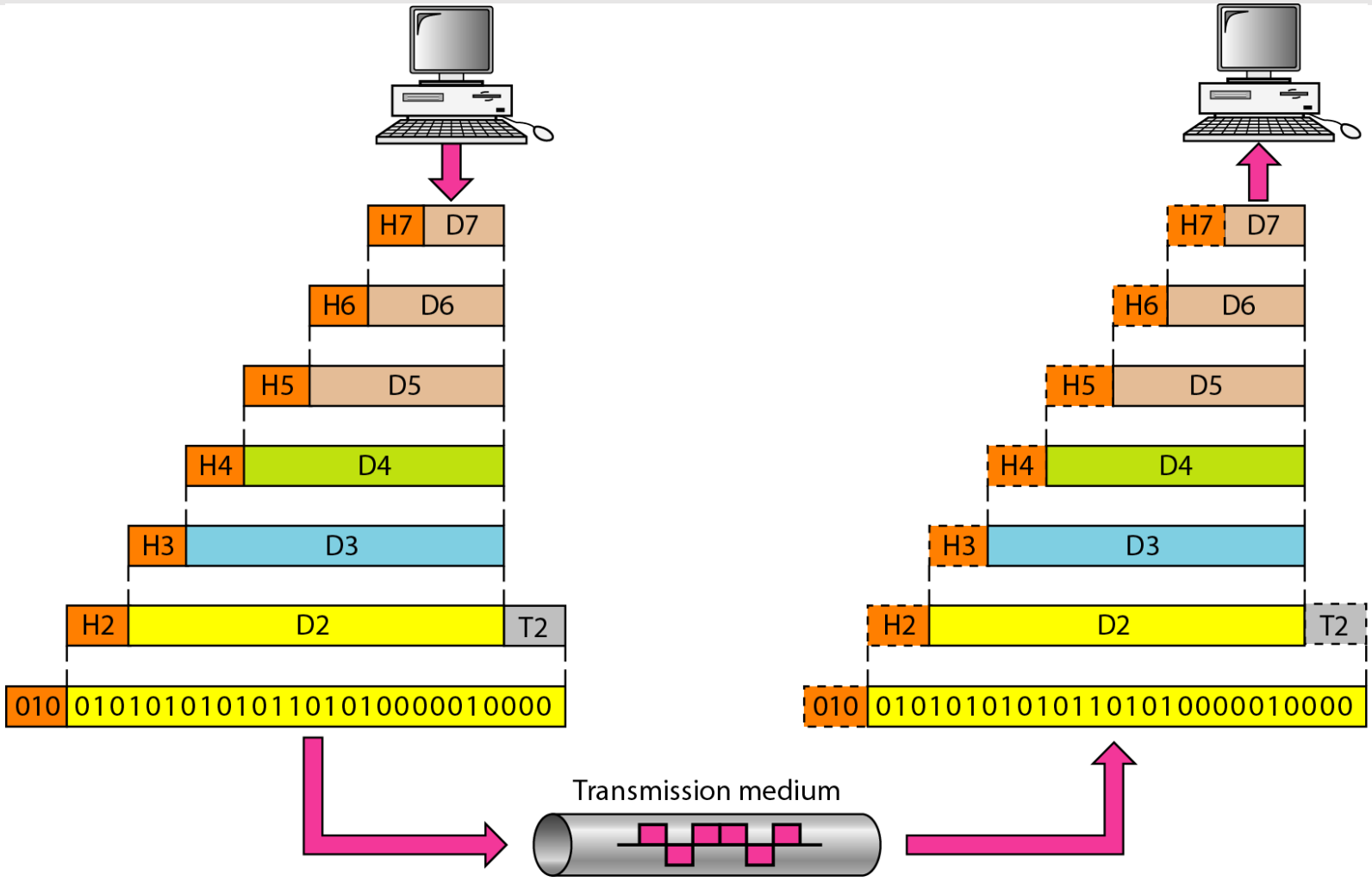


Figure An exchange using the OSI model

The OSI Model

- In Figure, which gives an overall view of the OSI layers, D7 means the data
- unit at layer 7, D6 means the data unit at layer 6, and so on. The process starts at layer 7 (the application layer), then moves from layer to layer in descending, sequential order. At each layer, a header, or possibly a trailer, can be added to the data unit.
- Commonly, the trailer is added only at layer 2. When the formatted data unit passes through the physical layer (layer 1), it is changed into an electromagnetic signal and transported along a physical link

The OSI Model

Physical Layer

- The physical layer coordinates the functions required to carry a bit stream over a physical medium.
- It deals with the mechanical and electrical specifications of the interface and transmission medium. It also defines the procedures and functions that physical devices and interfaces have to perform for transmission to Occur.
- Figure shows the position of the physical layer with respect to the transmission medium and the data link layer.

The OSI Model

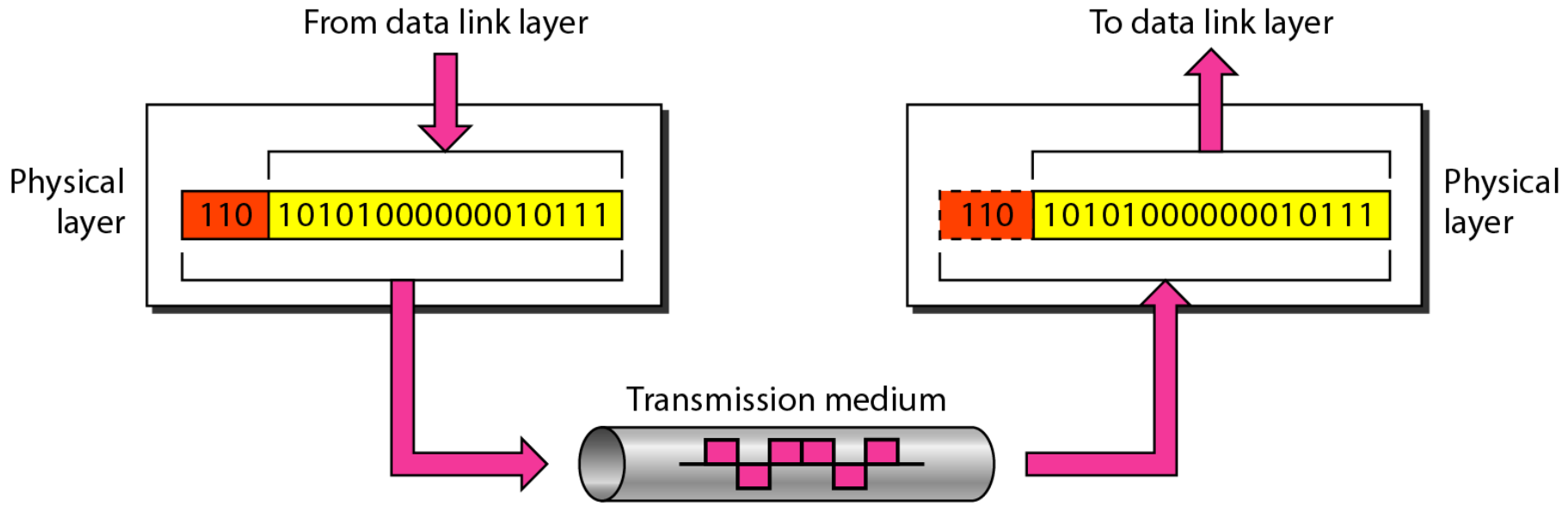


Figure Physical layer

The OSI Model

- The physical layer is also concerned with the following:
- **Physical characteristics of interfaces and medium.** The physical layer defines the characteristics of the interface between the devices and the transmission medium. It also defines the type of transmission medium.
- **Representation of bits.** The physical layer data consists of a stream of bits (sequence of 0s or 1s) with no interpretation. To be transmitted, bits must be encoded into signals--electrical or optical. The physical layer defines the type of encoding (how 0s and 1s are changed to signals).
- **Data rate.** The transmission rate-the number of bits sent each second-is also defined by the physical layer. In other words, the physical layer defines the duration of a bit, which is how long it lasts.

The OSI Model

- **Synchronization of bits.** The sender and receiver not only must use the same bit rate but also must be synchronized at the bit level. In other words, the sender and the receiver clocks must be synchronized.
- **Line configuration.** The physical layer is concerned with the connection of devices to the media. In a point-to-point configuration, two devices are connected through a dedicated link. In a multipoint configuration, a link is shared among several devices.

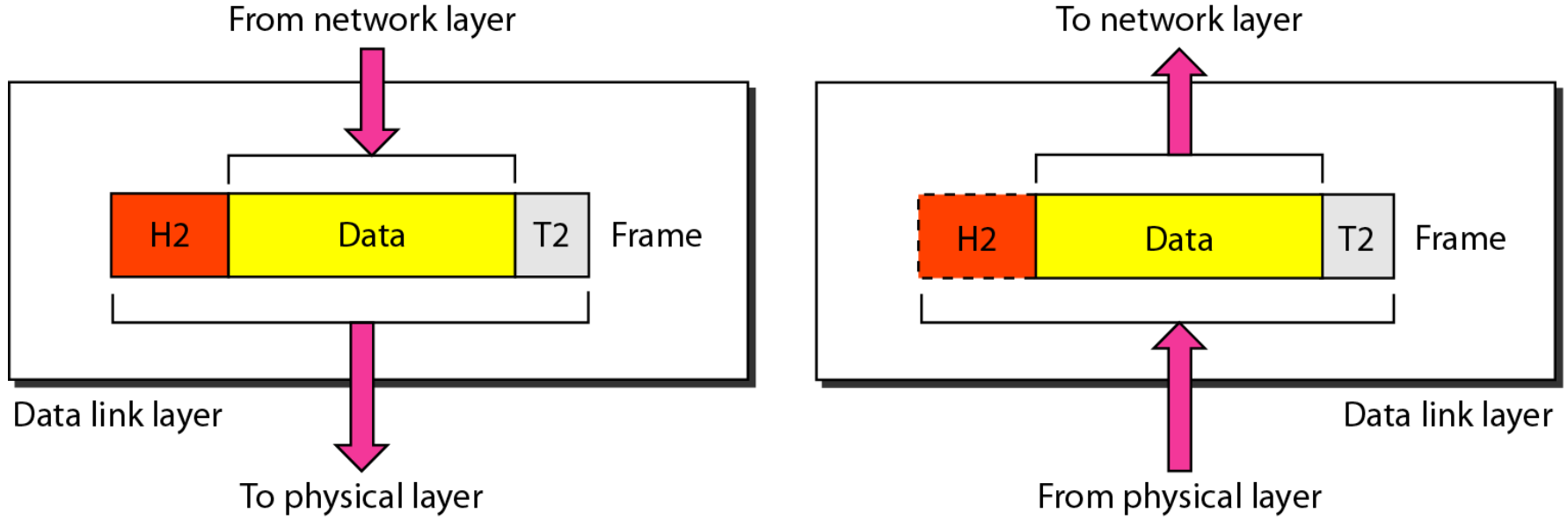
The OSI Model

- **Physical topology.** The physical topology defines how devices are connected to make a network. Devices can be connected by using a mesh topology (every device is connected to every other device), a star topology (devices are connected through a central device), a ring topology (each device is connected to the next, forming a ring), a bus topology (every device is on a common link), or a hybrid topology (this is a combination of two or more topologies).
- **Transmission mode.** The physical layer also defines the direction of transmission between two devices: simplex, half-duplex, or full-duplex. In simplex mode, only one device can send; the other can only receive. The simplex mode is a one-way communication. In the half-duplex mode, two devices can send and receive, but not at the same time. In a full-duplex (or simply duplex) mode, two devices can send and receive at the same time.

The OSI Model

Data Link Layer

- The data link layer transforms the physical layer, a raw transmission facility, to a reliable link. It makes the physical layer appear error-free to the upper layer (network layer).
- Figure shows the relationship of the data link layer to the network and physical layers.



The OSI Model

- Other responsibilities of the data link layer include the following:
- **Framing.** The data link layer divides the stream of bits received from the network layer into manageable data units called frames.
- **Physical addressing.** If frames are to be distributed to different systems on the network, the data link layer adds a header to the frame to define the sender and/or receiver of the frame. If the frame is intended for a system outside the sender's network, the receiver address is the address of the device that connects the network to the next one.

The OSI Model

- **Flow control.** If the rate at which the data are absorbed by the receiver is less than the rate at which data are produced in the sender, the data link layer imposes a flow control mechanism to avoid overwhelming the receiver.
- **Error control.** The data link layer adds reliability to the physical layer by adding mechanisms to detect and retransmit damaged or lost frames. It also uses a mechanism to recognize duplicate frames. Error control is normally achieved through a trailer added to the end of the frame.
- **Access control.** When two or more devices are connected to the same link, data link layer protocols are necessary to determine which device has control over the link at any given time.

The OSI Model

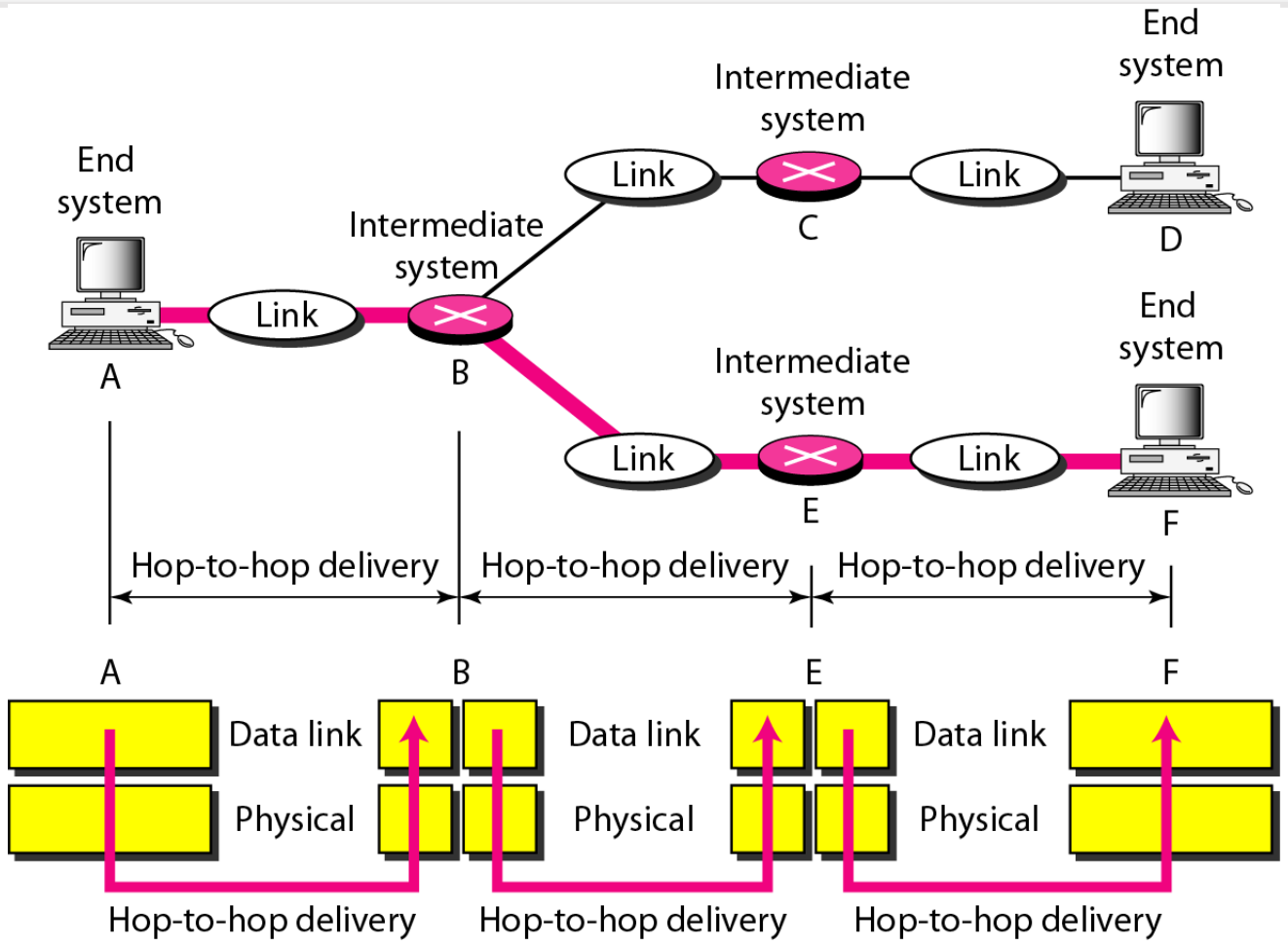


Figure Hop-to-hop delivery

The OSI Model

- As the figure shows, communication at the data link layer occurs between two adjacent nodes. To send data from *A* to *F*, three partial deliveries are made.
- First, the data link layer at *A* sends a frame to the data link layer at *B* (a router).
- Second, the data link layer at *B* sends a new frame to the data link layer at *E*. Finally, the data link layer at *E* sends a new frame to the data link layer at *F*.

The OSI Model

Network Layer

- The network layer is responsible for the *source-to-destination* delivery of a packet, possibly across multiple networks (links). Whereas the data link layer oversees the delivery of the packet between two systems on the same network (links), the network layer ensures that each packet gets from its point of origin to its final destination.
- If two systems are connected to the same link, there is usually no need for a network layer. However, if the two systems are attached to different networks (links) with connecting devices between the networks (links), there is often a need for the network layer to accomplish source-to-destination delivery.
- Figure shows the relationship of the network layer to the data link and transport layers.

The OSI Model

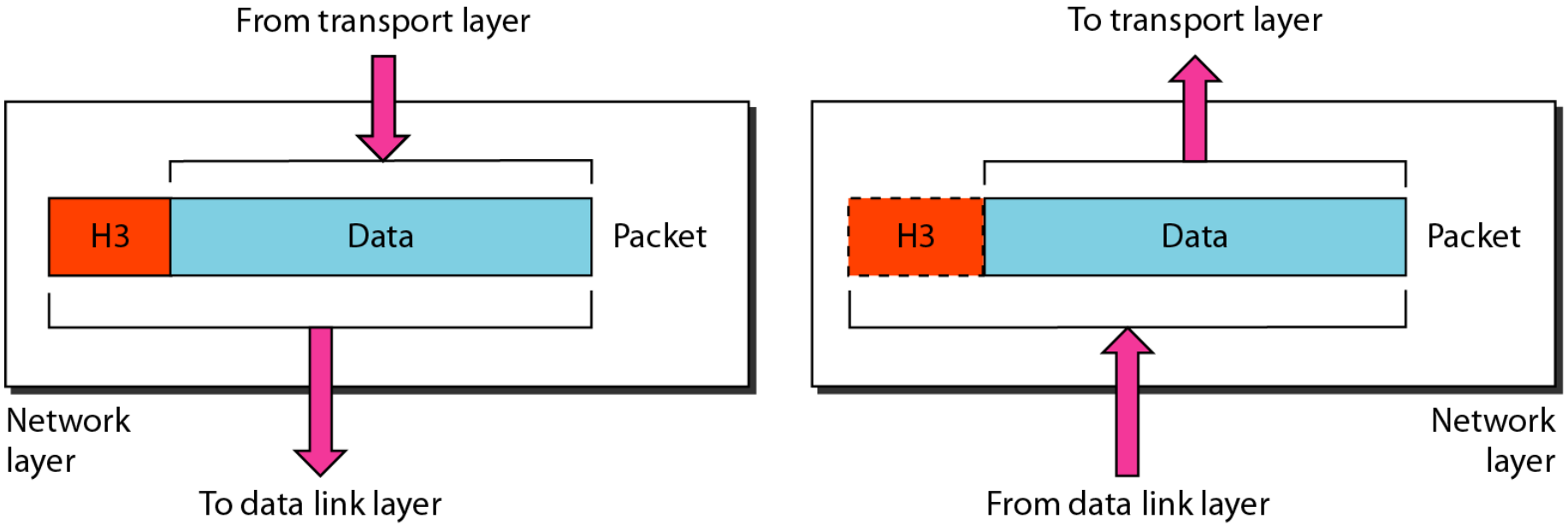


Figure Network layer

The OSI Model

- Other responsibilities of the network layer include the following:
- **Logical addressing.** The physical addressing implemented by the data link layer handles the addressing problem locally. If a packet passes the network boundary, we need another addressing system to help distinguish the source and destination systems. The network layer adds a header to the packet coming from the upper layer that, among other things, includes the logical addresses of the sender and receiver.
- **Routing.** When independent networks or links are connected to create internetworks (network of networks) or a large network, the connecting devices (called routers or switches) route or switch the packets to their final destination. One of the functions of the network layer is to provide this mechanism.

The OSI Model

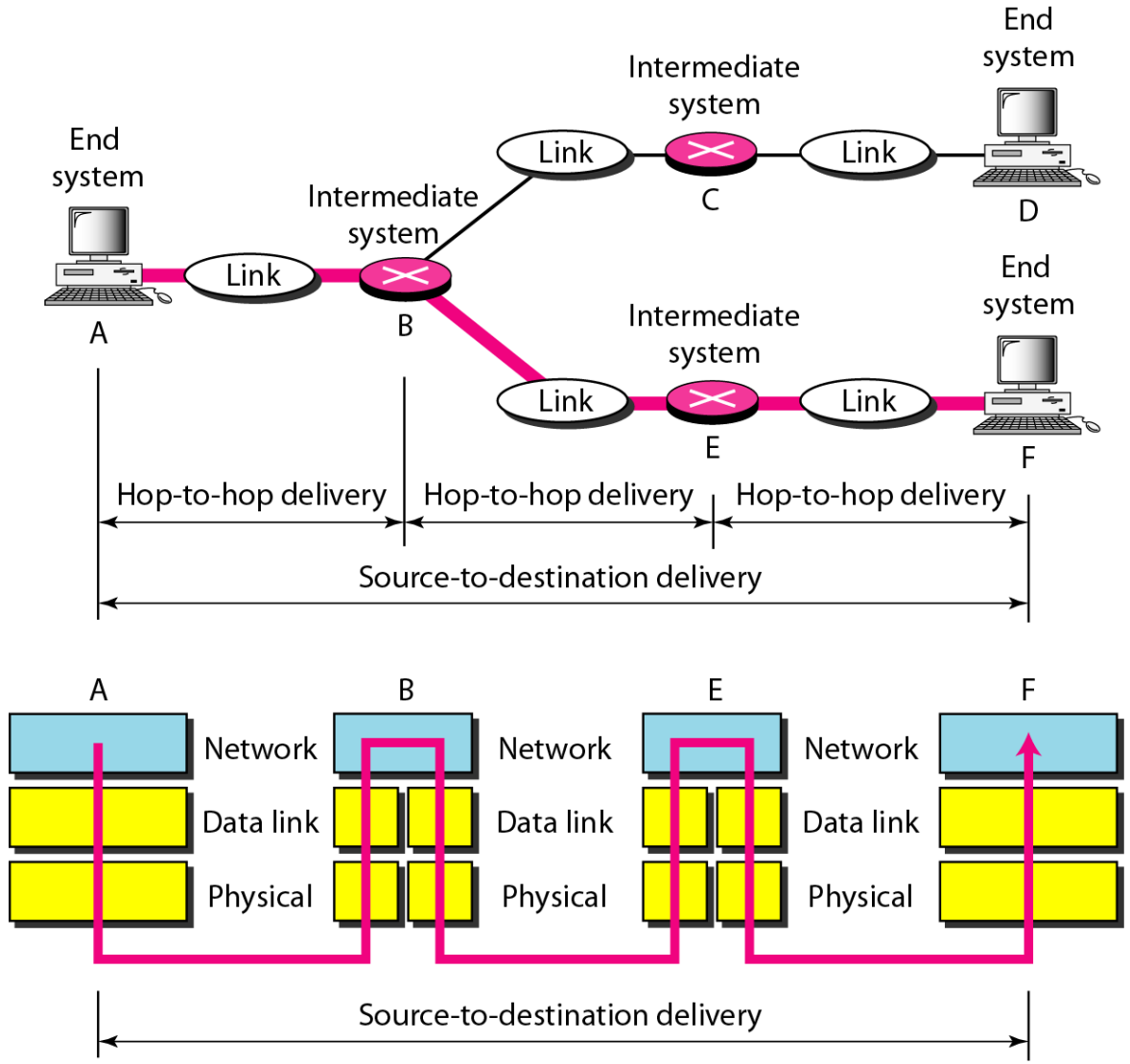


Figure Source-to-destination delivery

The OSI Model

- As the figure shows, now we need a source-to-destination delivery. The network layer at A sends the packet to the network layer at B.
- When the packet arrives at router B, the router makes a decision based on the final destination (F) of the packet.
- Router B uses its routing table to find that the next hop is router E. The network layer at B, therefore, sends the packet to the network layer at E. The network layer at E, in turn, sends the packet to the network layer at F.

The OSI Model

Transport Layer

- The transport layer is responsible for *process-to-process* delivery of the entire message.
- A process is an application program running on a host. Whereas the network layer oversees source-to-destination delivery of individual packets, it does not recognize any relationship between those packets.
- It treats each one independently, as though each piece belonged to a separate message, whether or not it does. The transport layer, on the other hand, ensures that the whole message arrives intact and in order, overseeing both error control and flow control at the source-to-destination level.
- Figure shows the relationship of the transport layer to the network and session layers.

The OSI Model

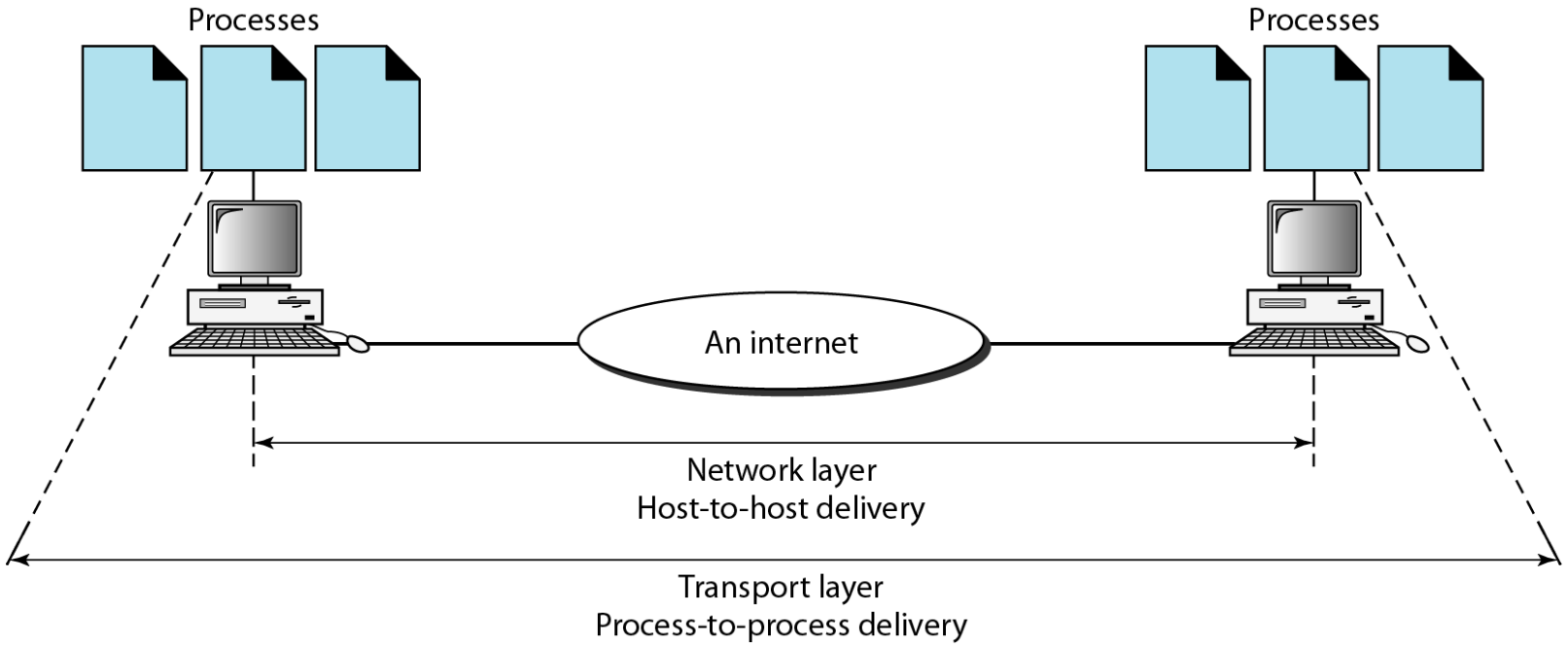


Figure Reliable process-to-process delivery of a message

The OSI Model

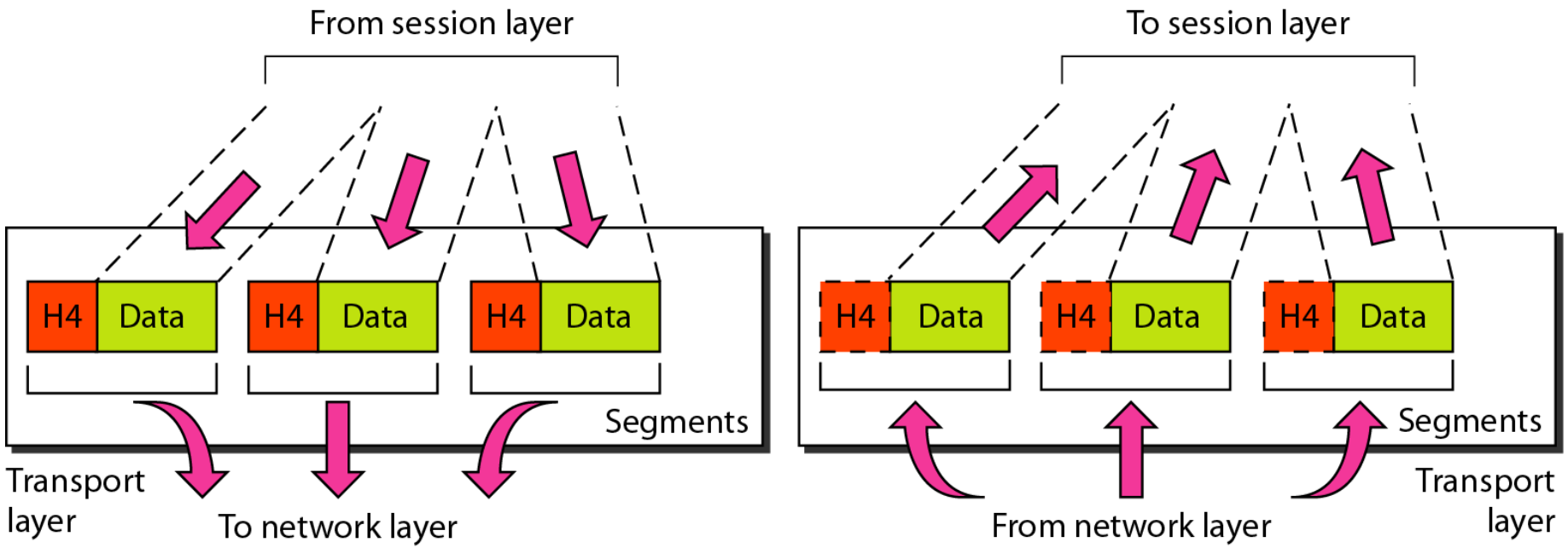


Figure Transport layer

The OSI Model

- Other responsibilities of the transport layer include the following:

- **Service-point addressing.** Computers often run several programs at the same time. For this reason, source-to-destination delivery means delivery not only from one computer to the next but also from a specific process (running program) on one computer to a specific process (running program) on the other. The transport layer header must therefore include a type of address called a service-point address (or port address). The network layer gets each packet to the correct computer; the transport layer gets the entire message to the correct process on that computer.

The OSI Model

- **Segmentation and reassembly.** A message is divided into transmittable segments, with each segment containing a sequence number. These numbers enable the transport layer to reassemble the message correctly upon arriving at the destination and to identify and replace packets that were lost in transmission.

- **Connection control.** The transport layer can be either connectionless or connection-oriented. A connectionless transport layer treats each segment as an independent packet and delivers it to the transport layer at the destination machine. A connection-oriented transport layer makes a connection with the transport layer at the destination machine first before delivering the packets. After all the data are transferred, the connection is terminated.

The OSI Model

- **Flow control.** Like the data link layer, the transport layer is responsible for flow control. However, flow control at this layer is performed end to end rather than across a single link.

- **Error control.** Like the data link layer, the transport layer is responsible for error control. However, error control at this layer is performed process-to-process rather than across a single link. The sending transport layer makes sure that the entire message arrives at the receiving transport layer without error (damage, loss, or duplication). Error correction is usually achieved through retransmission.

The OSI Model

Session Layer

- The services provided by the first three layers (physical, data link, and network) are not sufficient for some processes.
- The session layer is the network dialog controller. It establishes, maintains, and synchronizes the interaction among communicating systems.

The OSI Model

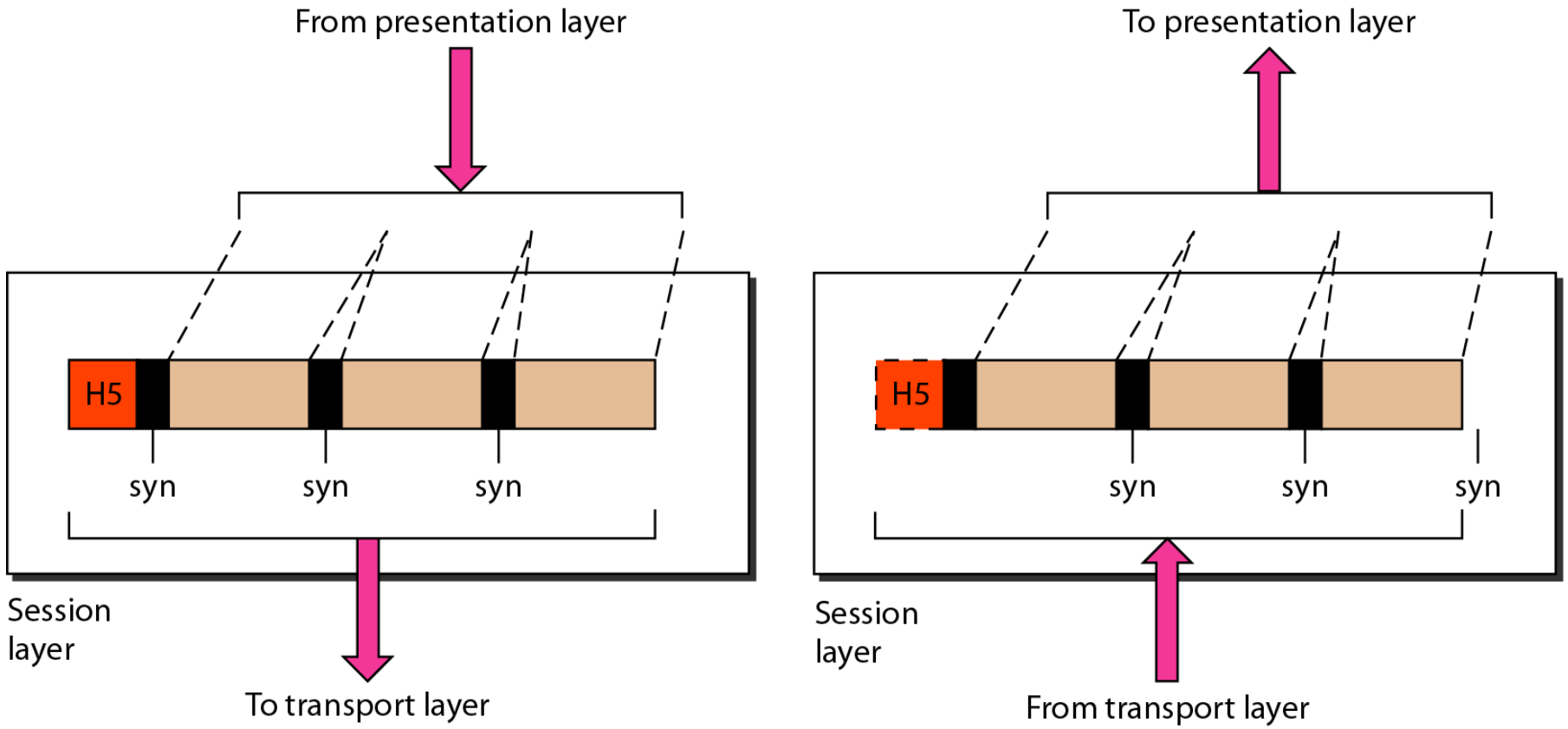


Figure Session layer

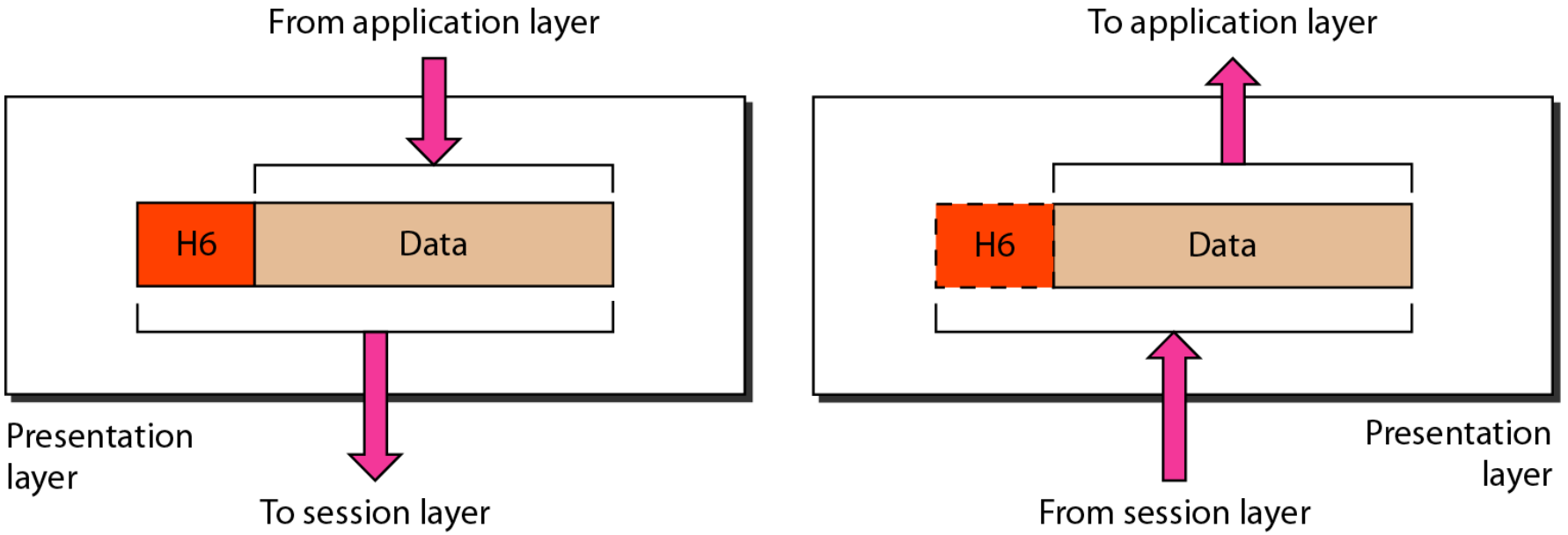
The OSI Model

- Specific responsibilities of the session layer include the following:
- **Dialog control.** The session layer allows two systems to enter into a dialog. It allows the communication between two processes to take place in either half-duplex (one way at a time) or full-duplex (two ways at a time) mode.
- **Synchronization.** The session layer allows a process to add checkpoints, or synchronization points, to a stream of data. For example, if a system is sending a file of 2000 pages, it is advisable to insert checkpoints after every 100 pages to ensure that each 100-page unit is received and acknowledged independently. In this case, if a crash happens during the transmission of page 523, the only pages that need to be resent after system recovery are pages 501 to 523. Pages previous to 501 need not be resent.

The OSI Model

Presentation Layer

- The presentation layer is concerned with the syntax and semantics of the information exchanged between two systems. Figure shows the relationship between the presentation layer and the application and session layer.



The OSI Model

- Specific responsibilities of the presentation layer include the following:
- **Translation.** The processes (running programs) in two systems are usually exchanging information in the form of character strings, numbers, and so on. The information must be changed to bit streams before being transmitted. Because different computers use different encoding systems, the presentation layer is responsible for interoperability between these different encoding methods. The presentation layer at the sender changes the information from its sender-dependent format into a common format. The presentation layer at the receiving machine changes the common format into its receiver-dependent format.

The OSI Model

- **Encryption.** To carry sensitive information, a system must be able to ensure privacy. Encryption means that the sender transforms the original information to another form and sends the resulting message out over the network. Decryption reverses the original process to transform the message back to its original form.

- **Compression.** Data compression reduces the number of bits contained in the information. Data compression becomes particularly important in the transmission of multimedia such as text, audio, and video.

The OSI Model

Application Layer

- The application layer enables the user, whether human or software, to access the network. It provides user interfaces and support for services such as electronic mail, remote file access and transfer, shared database management, and other types of distributed information services.
- Figure shows the relationship of the application layer to the user and the presentation layer. Of the many application services available, the figure shows only three: X.400 (message-handling services), X.500 (directory services), and file transfer, access, and management (FTAM). The user in this example employs X.400 to send an e-mail message

The OSI Model

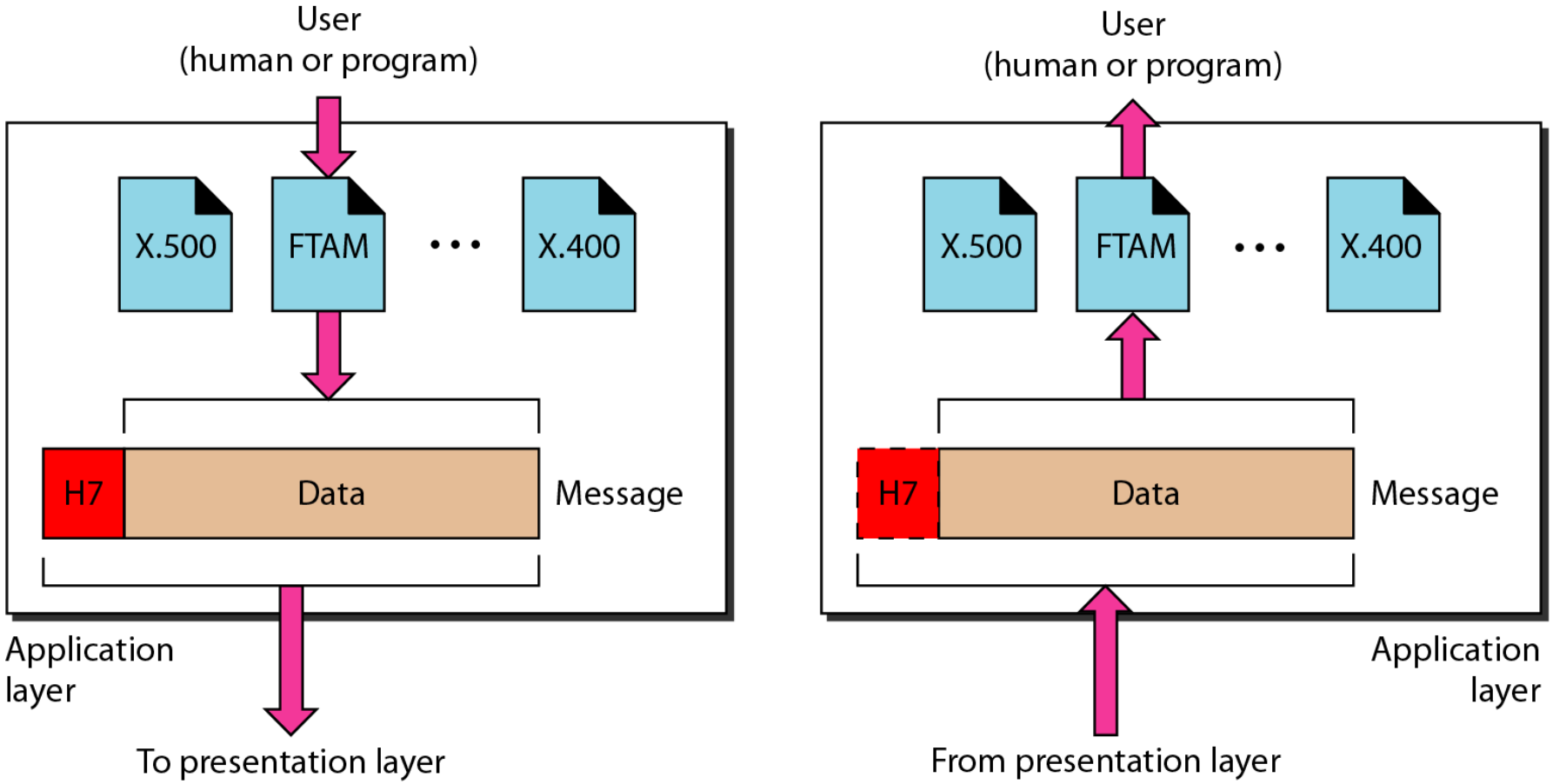


Figure Application layer

The OSI Model

- Specific services provided by the application layer include the following:
- **Network virtual terminal.** A network virtual terminal is a software version of a physical terminal, and it allows a user to log on to a remote host. To do so, the application creates a software emulation of a terminal at the remote host. The user's computer talks to the software terminal which, in turn, talks to the host, and vice versa. The remote host believes it is communicating with one of its own terminals and allows the user to log on.

The OSI Model

- **File transfer, access, and management.** This application allows a user to access files in a remote host (to make changes or read data), to retrieve files from a remote computer for use in the local computer, and to manage or control files in a remote computer locally.
- **Mail services.** This application provides the basis for e-mail forwarding and storage.
- **Directory services.** This application provides distributed database sources and access for global information about various objects and services.

The OSI Model

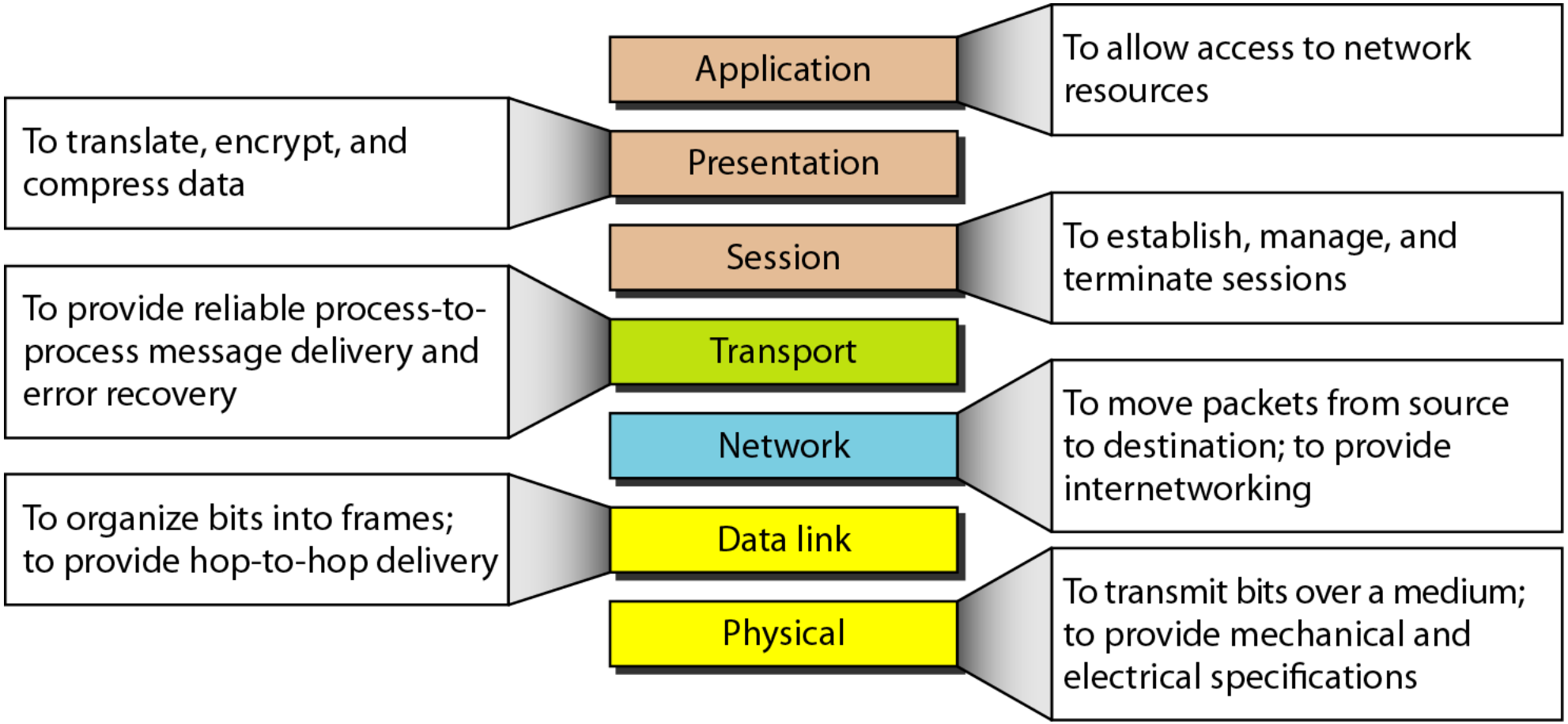


Figure Summary of layers

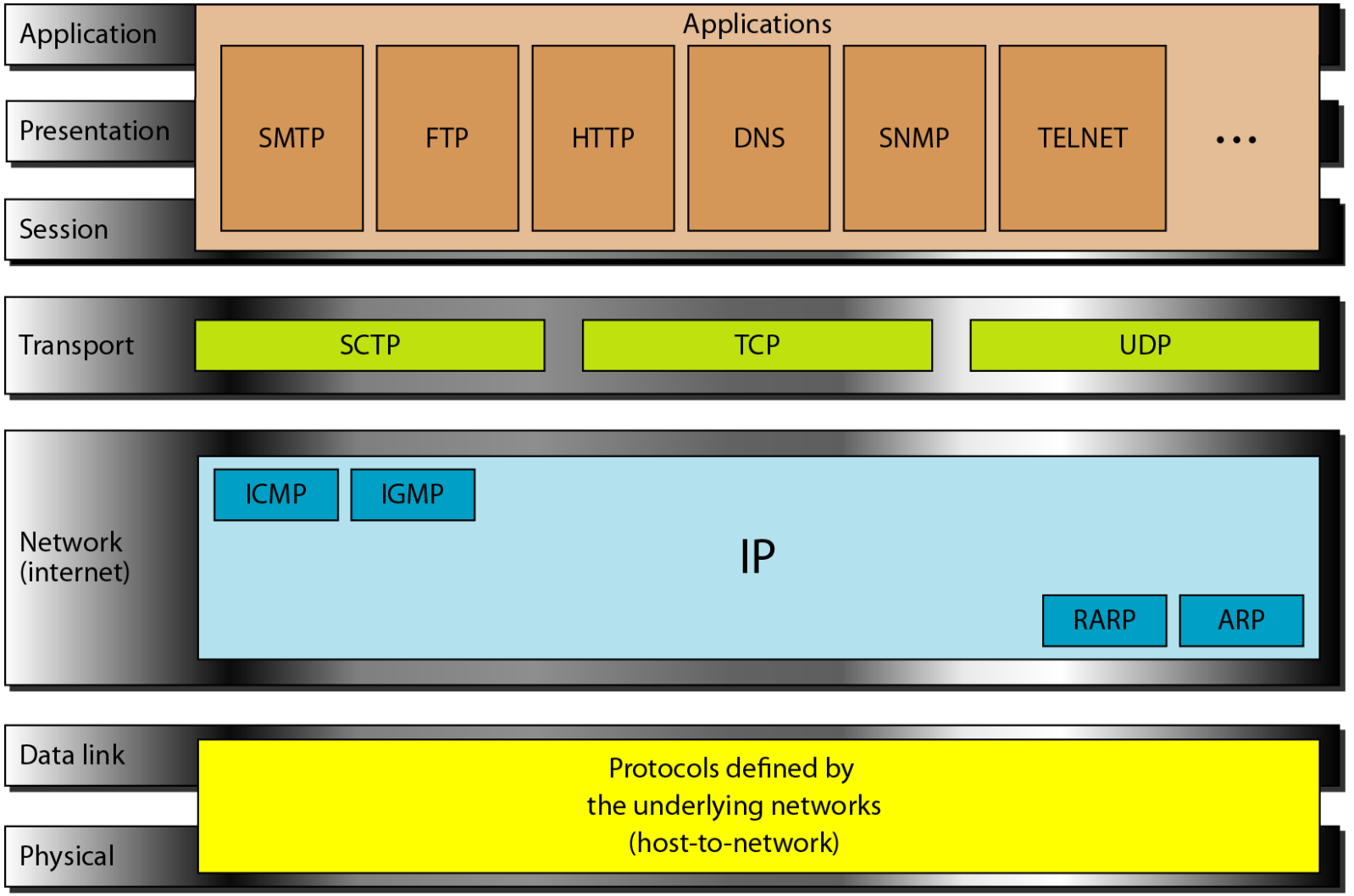
TCP/IP Protocol Suite

- The TCP/IP protocol suite was developed prior to the OSI model. Therefore, the layers in the TCP/IP protocol suite do not exactly match those in the OSI model.
- The original TCP/IP protocol suite was defined as having four layers: host-to-network, internet, transport, and application. However, when TCP/IP is compared to OSI, we can say that the host-to-network layer is equivalent to the combination of the physical and data link layers.
- The internet layer is equivalent to the network layer, and the application layer is roughly doing the job of the session, presentation, and application layers with the transport layer in TCP/IP taking care of part of the duties of the session layer.

TCP/IP Protocol Suite

- The TCP/IP protocol suite is made of five layers: physical, data link, network, transport, and application.
- The first four layers provide physical standards, network interfaces, internetworking, and transport functions that correspond to the first four layers of the OSI model.
- The three topmost layers in the OSI model, however, are represented in TCPIIP by a single layer called the application layer (see Figure).

TCP/IP and OSI model



TCP/IP Protocol Suite

- TCP/IP is a hierarchical protocol made up of interactive modules, each of which provides a specific functionality; however, the modules are not necessarily interdependent.
- Whereas the OSI model specifies which functions belong to each of its layers, the layers of the TCP/IP protocol suite contain relatively independent protocols that can be mixed and matched depending on the needs of the system. The term hierarchical means that each upper-level protocol is supported by one or more lower-level protocols.
- At the transport layer, TCP/IP defines three protocols: Transmission Control Protocol (TCP), User Datagram Protocol (UDP), and Stream Control Transmission Protocol (SCTP).
- At the network layer, the main protocol defined by TCP/IP is the Internetworking Protocol (IP); there are also some other protocols that support data movement in this layer.

TCP/IP Protocol Suite

Physical and Data Link Layers

- At the physical and data link layers, TCP/IP does not define any specific protocol. It supports all the standard and proprietary protocols.
- A network in a TCP/IP internetwork can be a local-area network or a wide-area network.

TCP/IP Protocol Suite

Network Layer

- At the network layer (or, more accurately, the internetwork layer), TCP/IP supports the Internetworking Protocol. IP, in turn, uses four supporting protocols: ARP, RARP, ICMP, and IGMP.

Internetworking Protocol (IP)

- The Internetworking Protocol (IP) is the transmission mechanism used by the TCP/IP protocols. It is an unreliable and connectionless protocol-a best-effort delivery service.
- The term best effort means that IP provides no error checking or tracking. IP assumes the unreliability of the underlying layers and does its best to get a transmission through to its destination, but with no guarantees.
- IP transports data in packets called datagrams, each of which is transported separately.
- Datagrams can travel along different routes and can arrive out of sequence or be duplicated. IP does not keep track of the routes and has no facility for reordering datagrams once they arrive at their destination.

TCP/IP Protocol Suite

Address Resolution Protocol

- The Address Resolution Protocol (ARP) is used to associate a logical address with a physical address. On a typical physical network, such as a LAN, each device on a link is identified by a physical or station address, usually imprinted on the network interface card (NIC). ARP is used to find the physical address of the node when its Internet address is known.

Reverse Address Resolution Protocol

- The Reverse Address Resolution Protocol (RARP) allows a host to discover its Internet address when it knows only its physical address. It is used when a computer is connected to a network for the first time or when a diskless computer is booted.

TCP/IP Protocol Suite

Internet Control Message Protocol

- The Internet Control Message Protocol (ICMP) is a mechanism used by hosts and gateways to send notification of datagram problems back to the sender. ICMP sends query and error reporting messages.

Internet Group Message Protocol

- The Internet Group Message Protocol (IGMP) is used to facilitate the simultaneous transmission of a message to a group of recipients.

TCP/IP Protocol Suite

Transport Layer

- Traditionally the transport layer was represented in TCP/IP by two protocols: TCP and UDP. IP is a host-to-host protocol, meaning that it can deliver a packet from one physical device to another.
- UDP and TCP are transport level protocols responsible for delivery of a message from a process (running program) to another process. A new transport layer protocol, SCTP, has been devised to meet the needs of some newer applications.

TCP/IP Protocol Suite

User Datagram Protocol

- The User Datagram Protocol (UDP) is the simpler of the two standard TCP/IP transport protocols. It is a process-to-process protocol that adds only port addresses, checksum, error control, and length information to the data from the upper layer.

Transmission Control Protocol

- The Transmission Control Protocol (TCP) provides full transport-layer services to applications. TCP is a reliable stream transport protocol. The term stream, in this context, means connection-oriented: A connection must be established between both ends of a transmission before either can transmit data.

TCP/IP Protocol Suite

- At the sending end of each transmission, TCP divides a stream of data into smaller units called *segments*. Each segment includes a sequence number for reordering after receipt, together with an acknowledgment number for the segments received. Segments are carried across the internet inside of IP datagrams. At the receiving end, TCP collects each datagram as it comes in and reorders the transmission based on sequence numbers.

Stream Control Transmission Protocol

- The Stream Control Transmission Protocol (SCTP) provides support for newer applications such as voice over the Internet. It is a transport layer protocol that combines the best features of UDP and TCP.

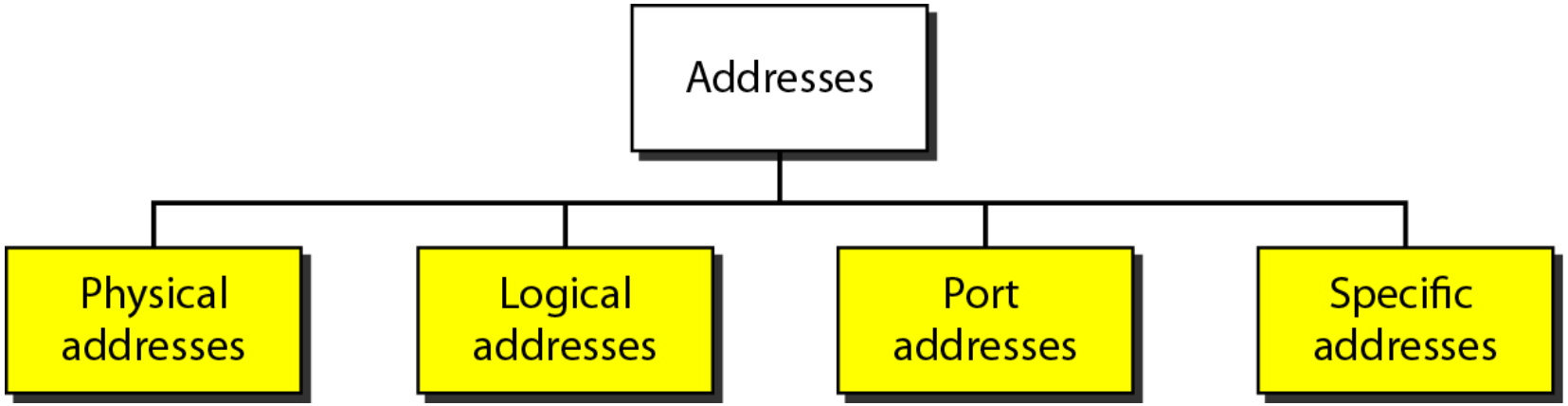
TCP/IP Protocol Suite

Application Layer

- The application layer in TCP/IP is equivalent to the combined session, presentation, and application layers in the OSI model. Many protocols are defined at this layer.

Addressing

- Four levels of addresses are used in an internet employing the TCP/IP protocols: physical (link) addresses, logical (IP) addresses, port addresses, and specific addresses



Addressing

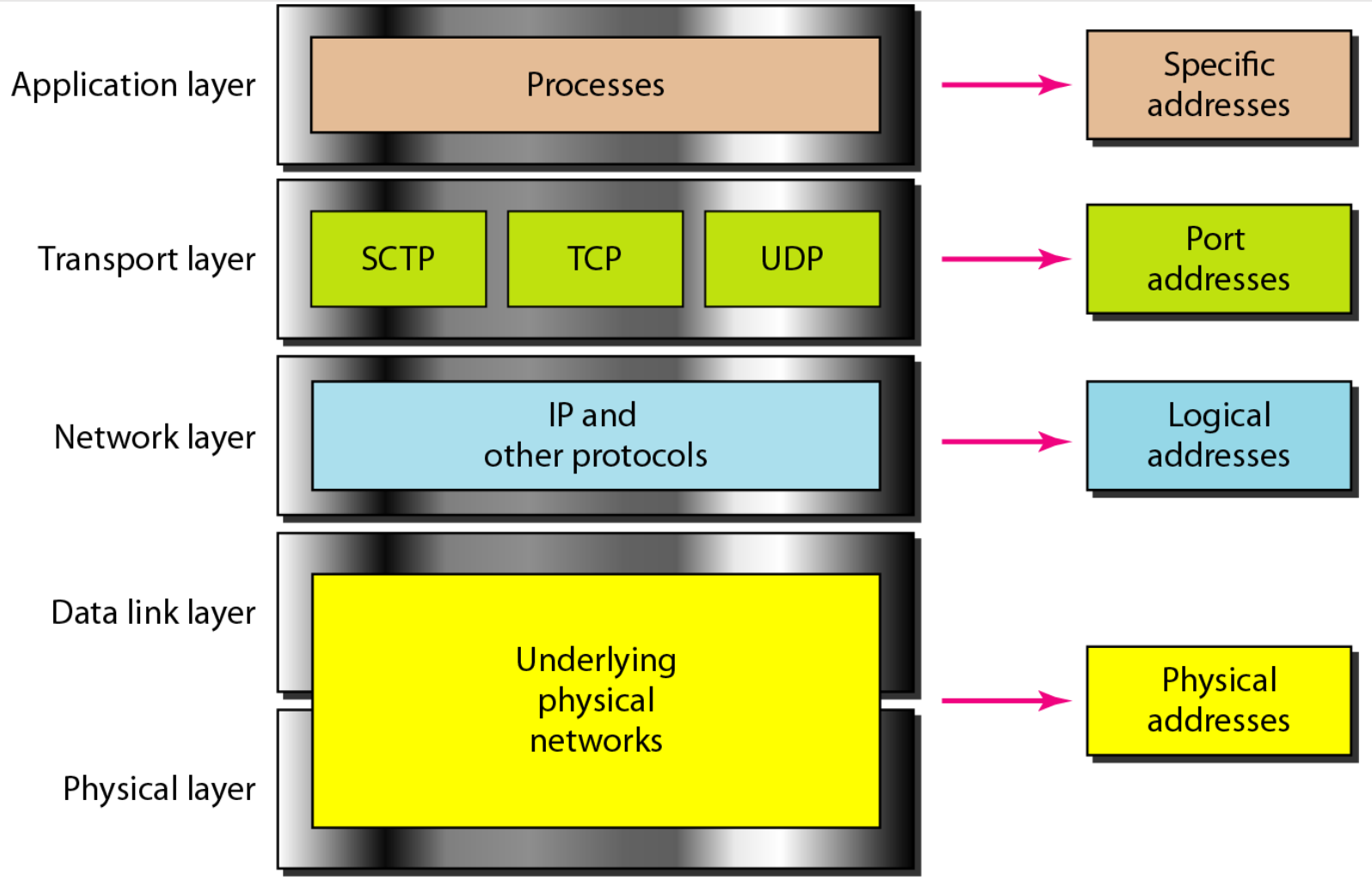


Figure Relationship of layers and addresses in TCP/IP

Addressing

Physical Addresses

- The physical address, also known as the link address, is the address of a node as defined by its LAN or WAN. It is included in the frame used by the data link layer. It is the lowest-level address.
- The physical addresses have authority over the network (LAN or WAN). The size and format of these addresses vary depending on the network.
- For example, Ethernet uses a **6-byte (48-bit)** physical address that is imprinted on the network interface card (NIC). LocalTalk (Apple), however, has a 1-byte dynamic address that changes each time the station comes up.

Addressing

Logical Addresses

- Logical addresses are necessary for universal communications that are independent of underlying physical networks. Physical addresses are not adequate in an internetwork environment where different networks can have different address formats. A universal addressing system is needed in which each host can be identified uniquely, regardless of the underlying physical network.
- The logical addresses are designed for this purpose. A logical address in the Internet is currently a **32-bit** address that can uniquely define a host connected to the Internet. No two publicly addressed and visible hosts on the Internet can have the same IP address.

Addressing

Port Addresses

- The IP address and the physical address are necessary for a quantity of data to travel from a source to the destination host.
- However, arrival at the destination host is not the final objective of data communications on the Internet. A system that sends nothing but data from one computer to another is not complete.
- Today, computers are devices that can run multiple processes at the same time.
- The end objective of Internet communication is a process communicating with another process.

Addressing

- For example, computer *A* can communicate with computer *C* by using TELNET. At the same time, computer *A* communicates with computer *B* by using the File Transfer Protocol (FTP).
- For these processes to receive data simultaneously, we need a method to label the different processes. In other words, they need addresses.
- In the TCP/IP architecture, the label assigned to a process is called a port address.
- A port address in TCP/IP is **16 bits** in length.

Addressing

Specific Addresses

- Some applications have user-friendly addresses that are designed for that specific address.
- Examples include the e-mail address (for example, forouzan@fhda.edu) and the Universal Resource Locator (URL) (for example, www.mhhe.com).
- The first defines the recipient of an e-mail; the second is used to find a document on the World Wide Web. These addresses, however, get changed to the corresponding port and logical addresses by the sending computer.

Guided and Unguided Transmission Media

- A transmission media define as anything that can carry information from a source to a destination.

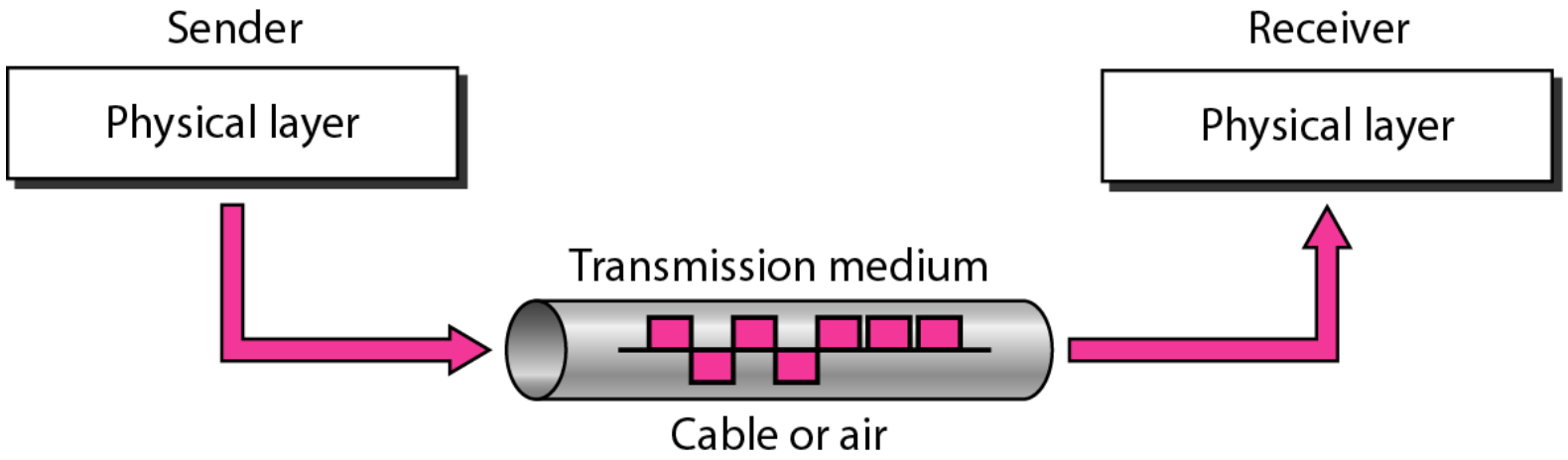


Figure Transmission medium and physical layer

Guided and Unguided Transmission Media

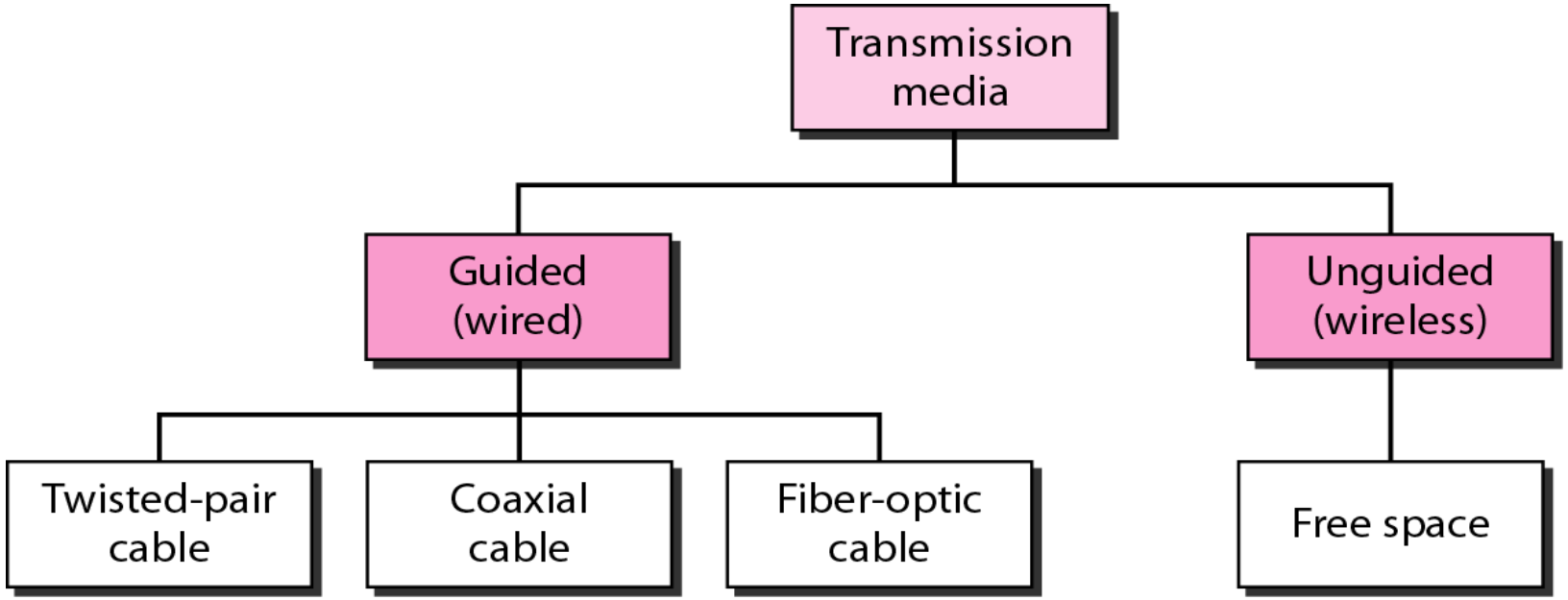


Figure Classes of transmission media

Guided and Unguided Transmission Media

Guided Media

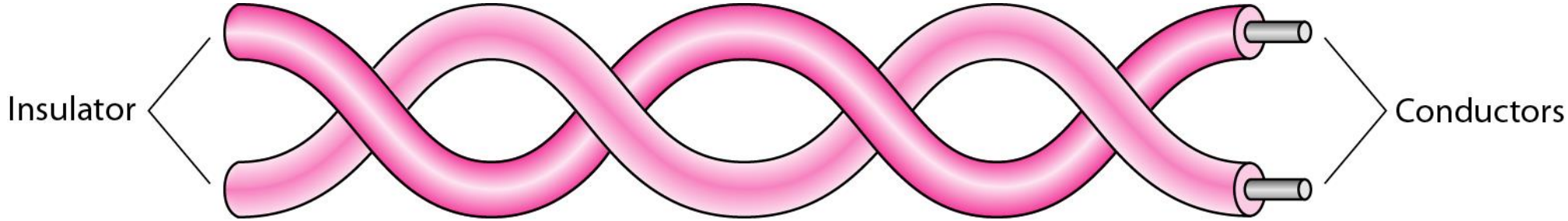
pipe

- Guided media, which are those that provide a conduit from one device to another, include twisted-pair cable, coaxial cable, and fiber-optic cable.
- A signal traveling along any of these media is directed and contained by the physical limits of the medium.
- Twisted-pair and coaxial cable use metallic (copper) conductors that accept and transport signals in the form of electric current. Optical fiber is a cable that accepts and transports signals in the form of light.

Guided and Unguided Transmission Media

Twisted-Pair Cable

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



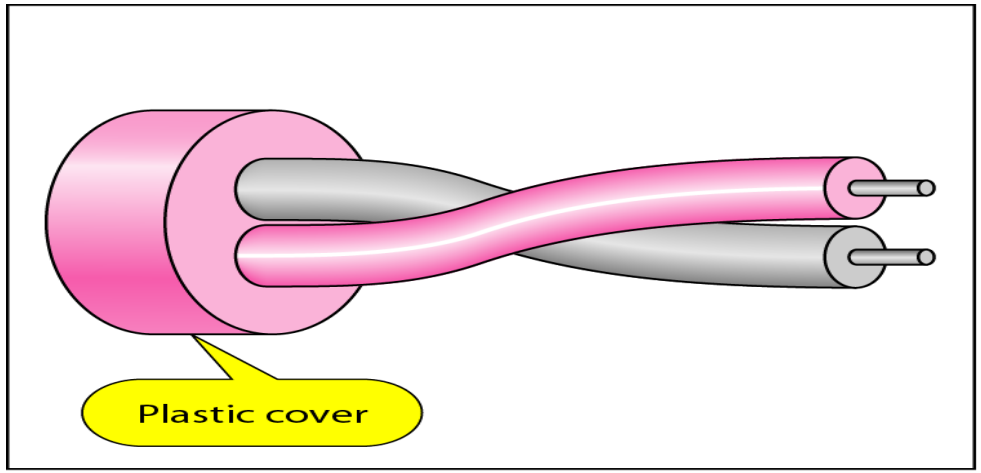
- One of the wires is used to carry signals to the receiver, and the other is used only as a ground reference. The receiver uses the difference between the two.
- In addition to the signal sent by the sender on one of the wires, interference (noise) and crosstalk may affect both wires and create unwanted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

Guided and Unguided Transmission Media

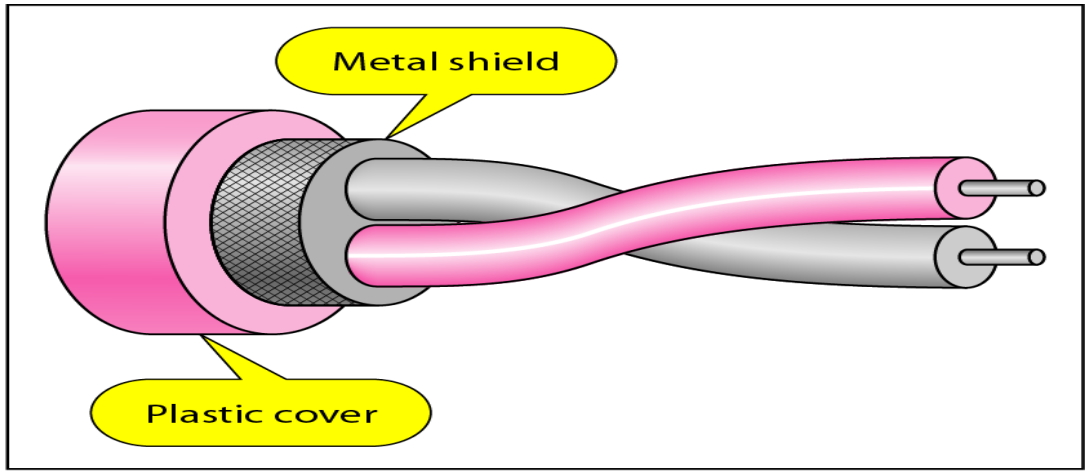
- If the two wires are parallel, the effect of these unwanted signals is not the same in both wires because they are at different locations relative to the noise or crosstalk sources. This results in a difference at the receiver. By twisting the pairs, a balance is maintained.
- For example, suppose in one twist, one wire is closer to the noise source and the other is farther; in the next twist, the reverse is true.

Guided and Unguided Transmission Media

- 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 shows the difference between UTP and STP.



a. UTP



b. STP

Category	*Specification*	*Data Rate (Mbps)*	*Use*
1	Unshielded twisted-pair used in telephone	< 0.1	Telephone
2	Unshielded twisted-pair originally used in T-lines	2	T-1 lines
3	Improved CAT 2 used in LANs	10	LANs
4	Improved CAT 3 used in Token Ring networks	20	LANs
5	Cable wire is normally 24 AWG with a jacket and outside sheath	100	LANs
5E	An extension to category 5 that includes extra features to minimize the crosstalk and electromagnetic interference	125	LANs
6	A new category with matched components coming from the same manufacturer. The cable must be tested at a 200-Mbps data rate.	200	LANs
7	Sometimes called SSTP (shielded screen twisted-pair). Each pair is individually wrapped in a helical metallic foil followed by a metallic foil shield in addition to the outside sheath. The shield decreases the effect of crosstalk and increases the data rate.	600	LANs

Guided and Unguided Transmission Media

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 :

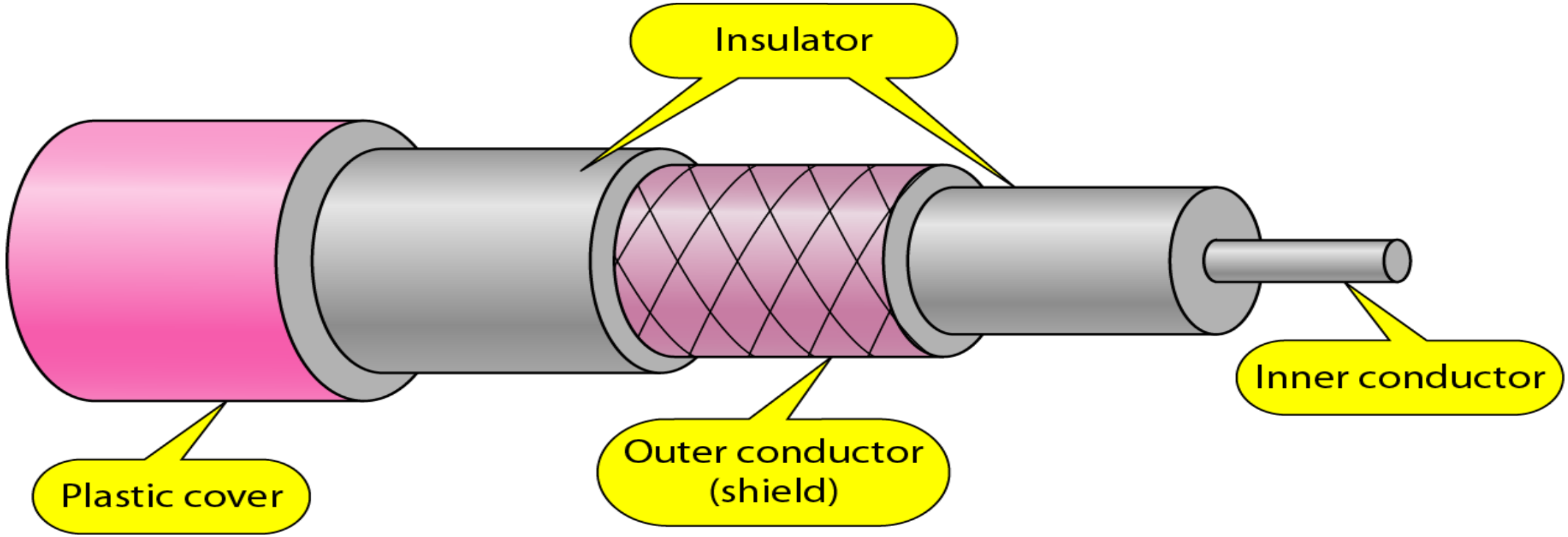
- Twisted-pair cables are used in telephone lines to provide voice and data channels. The local loop--the line that connects subscribers to the central telephone office---commonly consists of unshielded twisted-pair cables.
- The DSL lines that are used by the telephone companies to provide high-data-rate connections also use the high-bandwidth capability of unshielded twisted-pair cables.
- Local-area networks, such as 10Base-T and 100Base-T, also use twisted-pair cables.

Guided and Unguided Transmission Media

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).

Guided and Unguided Transmission Media



Guided and Unguided Transmission Media

- Coaxial cables are categorized by their radio government (RG) ratings. Table

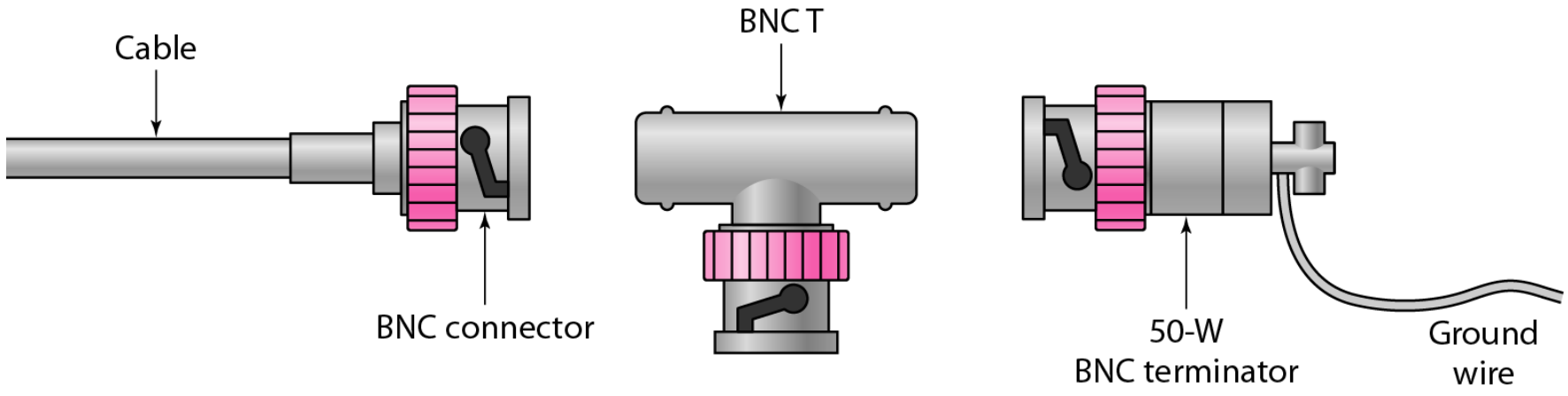
Category	*Impedance*	*Use*
RG-59	75 Ω	Cable TV
RG-58	50 Ω	Thin Ethernet
RG-11	50 Ω	Thick Ethernet

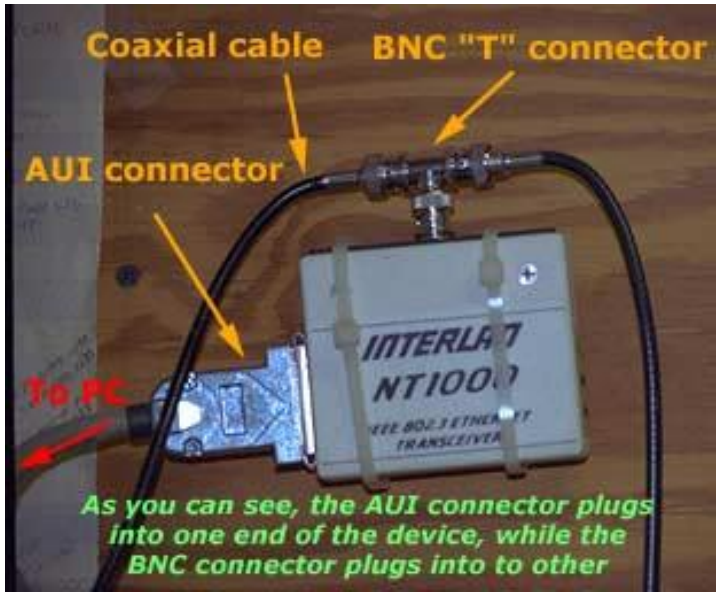
Two kinds:

- Thicknet(RG-11): it connect 100 devices with range 500 m (more expensive).
- Thinnet (RG-58): it connect 30 devices within 185 m (cheaper).

Guided and Unguided Transmission Media

- To connect coaxial cable to device, we need Bayonet Neill–Concelman (BNC).
- Carries signals of higher frequency ranges than twisted-pair cable.





Guided and Unguided Transmission Media

Fiber-Optic Cable

- A fiber-optic cable is made of glass or plastic and transmits signals in the form of light.

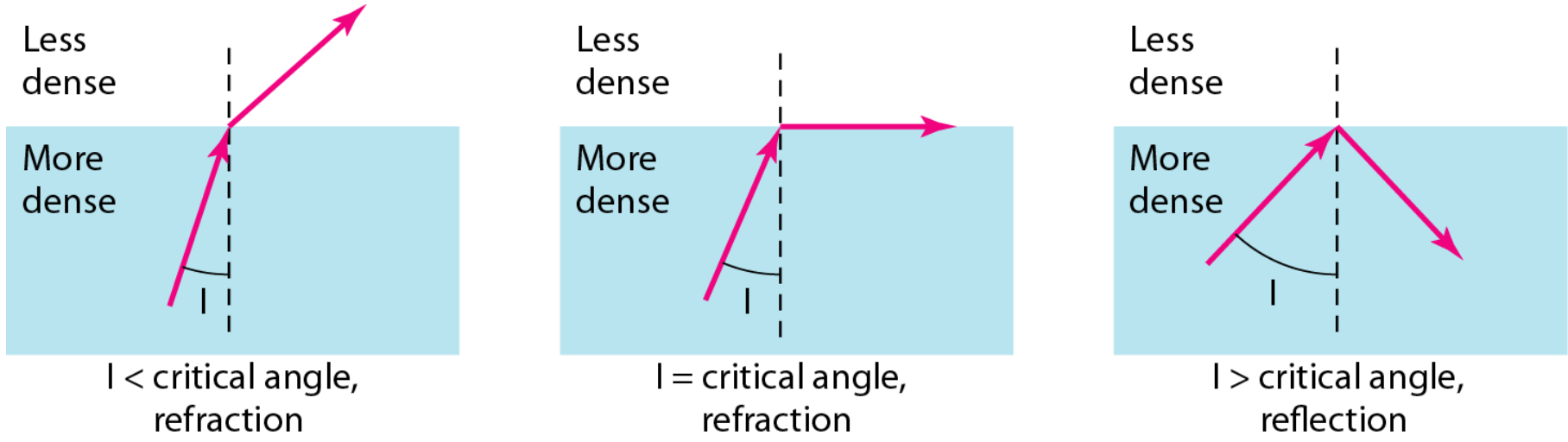


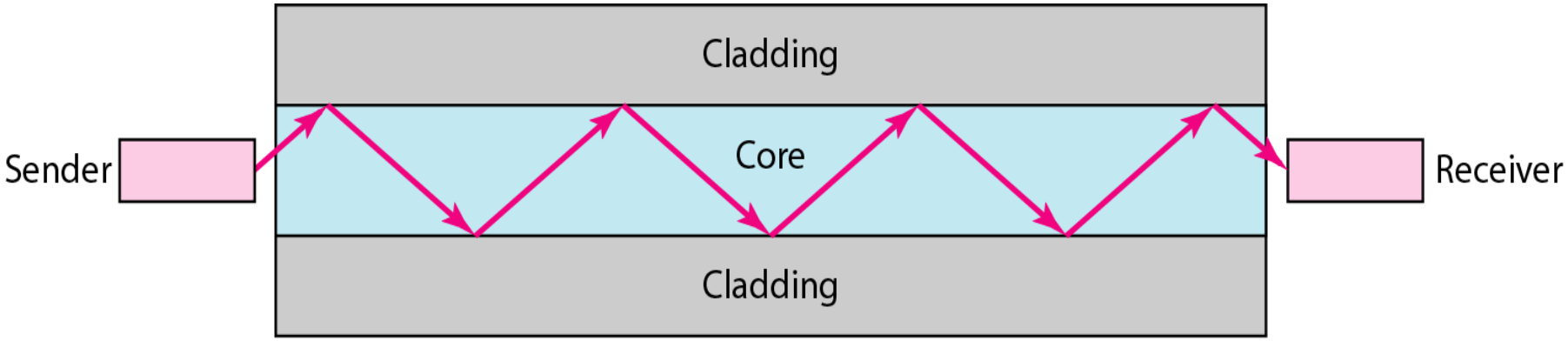
Figure Bending of light ray

Guided and Unguided Transmission Media

- If the angle of incidence I (the angle the ray makes with the line perpendicular to the interface between the two substances) is less than the critical angle, the ray refracts and moves closer to the surface. If the angle of incidence is equal to the critical angle, the light bends along the interface.
- If the angle is greater than the critical angle, the ray reflects (makes a turn) and travels again in the denser substance.
- Note that the critical angle is a property of the substance, and its value differs from one substance to another.
- 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.

Guided and Unguided Transmission Media

- 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.



Guided and Unguided Transmission Media

- **Propagation Modes**

1.Single-mode fiber

- Carries light pulses along single path.

2.Multimode fiber

- Many pulses of light travel at different angles

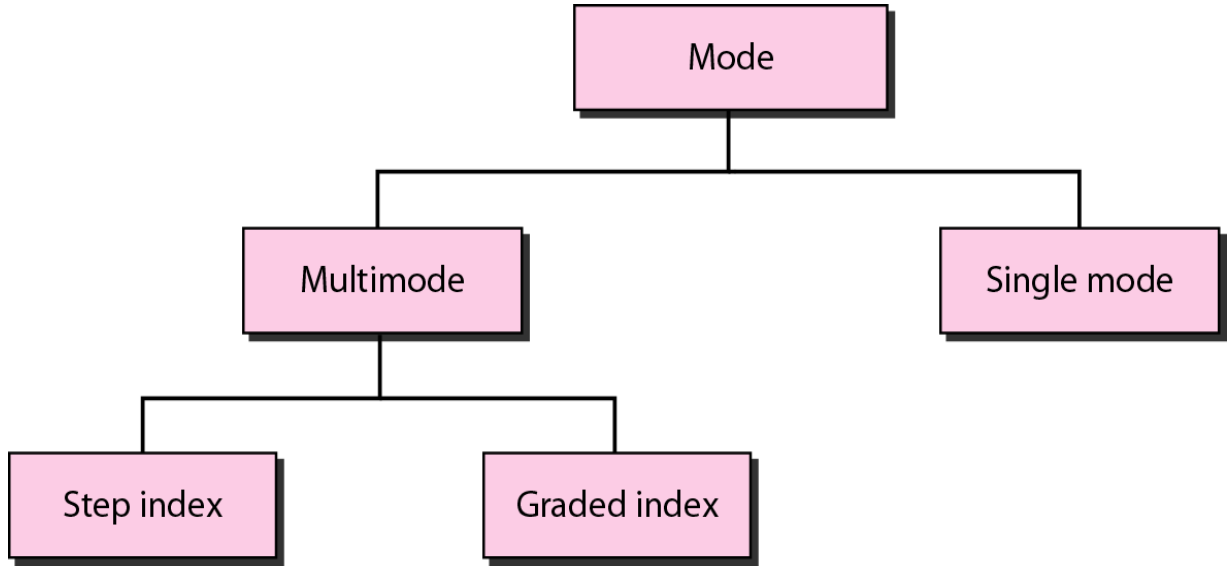


Figure Propagation modes

Guided and Unguided Transmission Media

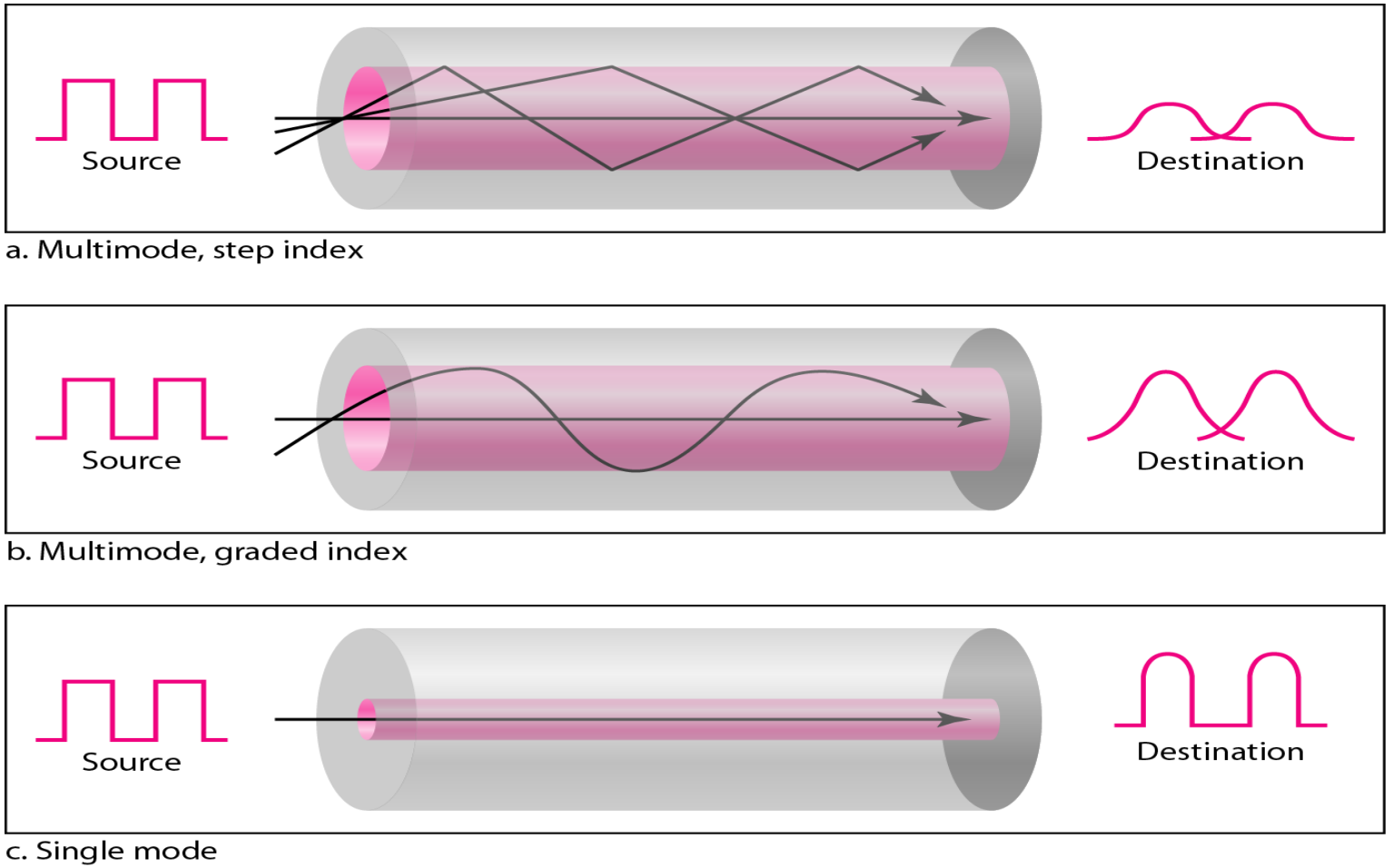


Figure **Modes**

Guided and Unguided Transmission Media

- **In multimode step-index fiber**, the density of the core remains constant from the center to the edges.
- A beam of light moves through this constant density in a straight line until it reaches the interface of the core and the cladding.
- At the interface, there is an abrupt change due to a lower density; this alters the angle of the beam's motion.
- The term step index refers to the suddenness of this change, which contributes to the distortion of the signal as it passes through the fiber.

Guided and Unguided Transmission Media

- **In multimode graded-index fiber**, decreases this distortion of the signal through the cable.
- The word index here refers to the index of refraction. As in figure, the index of refraction is related to density.
- A graded-index fiber, therefore, is one with varying densities. Density is highest at the center of the core and decreases gradually to its lowest at the edge.
- Figure shows the impact of this variable density on the propagation of light beams.

Guided and Unguided Transmission Media

Table Fiber types

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

Guided and Unguided Transmission Media

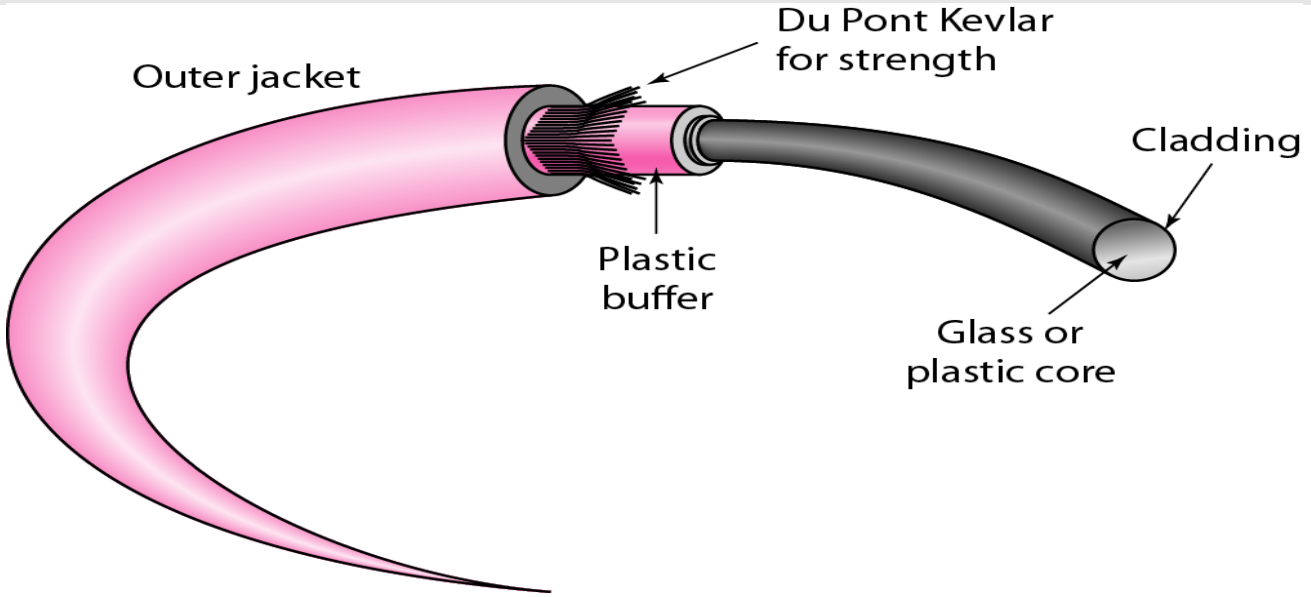
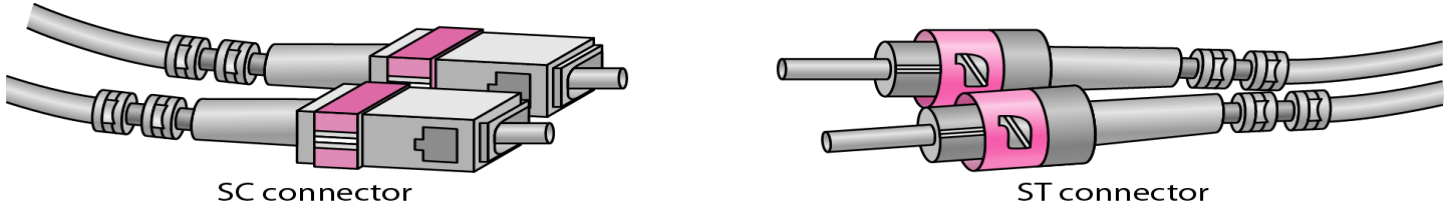


Figure Fiber construction



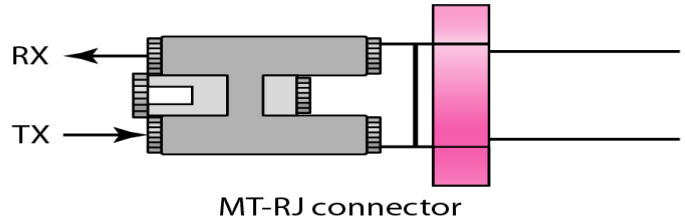


Figure Fiber-optic cable connectors

Guided and Unguided Transmission Media

Performance:

- Attenuation is flatter than in the case of twisted-pair cable and coaxial cable. The performance is such that we need fewer (actually 10 times less) repeaters when we use fiber-optic cable.

Applications:

- Fiber-optic cable is often found in backbone networks because its wide bandwidth is cost-effective. Today, with wavelength-division multiplexing (WDM), we can transfer data at a rate of 1600 Gbps.

Guided and Unguided Transmission Media

Advantages

- Fiber-optic cable has several advantages over metallic cable (twisted- pair or coaxial).
  1. Higher bandwidth.
  2. Less signal attenuation.
  3. Immunity to electromagnetic interference.
  4. Resistance to corrosive materials.
  5. Light weight.
  6. Greater immunity to tapping.

Guided and Unguided Transmission Media

Disadvantages

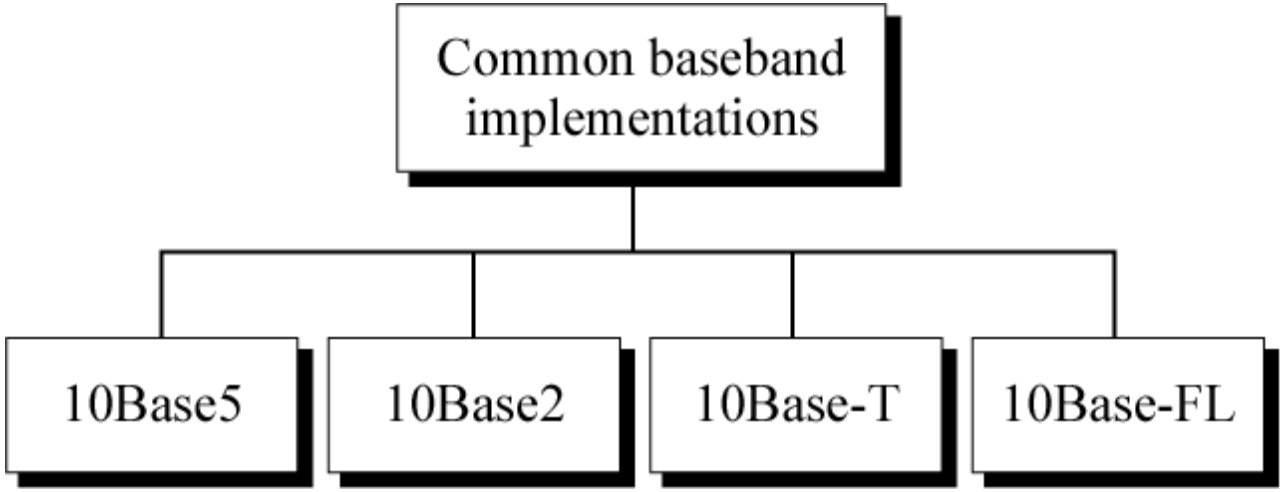
- There are some disadvantages in the use of optical fiber.

  1. *Installation and maintenance.* Fiber-optic cable is a relatively new technology. Its installation and maintenance require expertise that is not yet available everywhere.

  2. *Unidirectional light propagation.* Propagation of light is unidirectional. If we need bidirectional communication, two fibers are needed.

  3. *Cost.* The cable and the interfaces are relatively more expensive than those of other guided media. If the demand for bandwidth is not high, often the use of optical fiber cannot be justified.

Categories of 10-Mbps, Ethernet



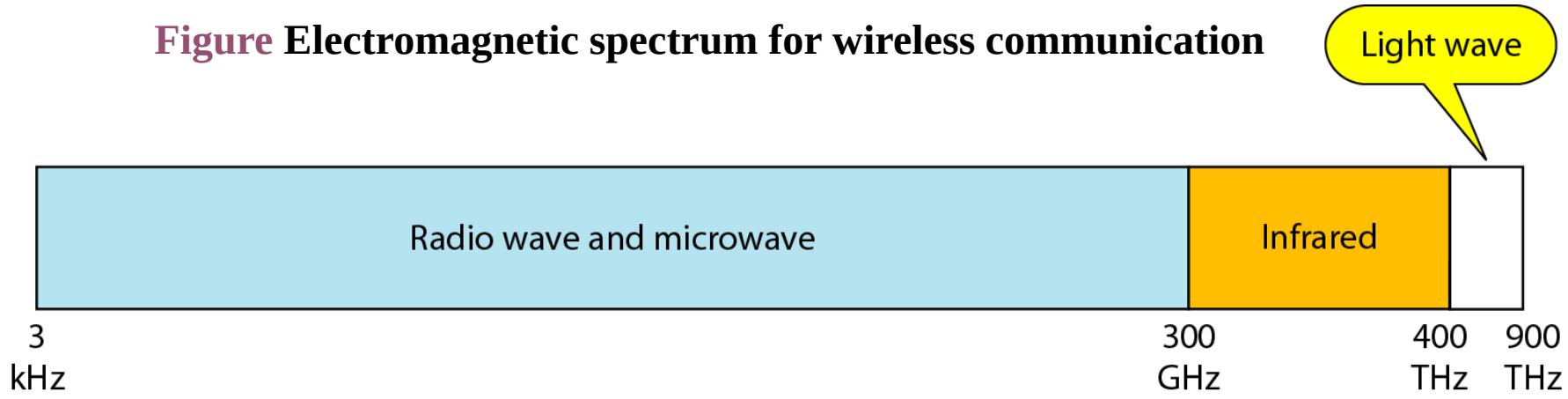
Name	Cable	Max. seg.	Nodes/seg.	Advantages
10Base5	Thick coax	500 m	100	Original cable; now obsolete
10Base2	Thin coax	185 m	30	No hub needed
10Base-T	Twisted pair	100 m	1024	Cheapest system
10Base-F	Fiber optics	2000 m	1024	Best between buildings

Guided and Unguided Transmission Media

Unguided Media

- Unguided media transport electromagnetic waves without using a physical conductor. This type of communication is often referred to as wireless communication.

Figure Electromagnetic spectrum for wireless communication



- Radio, satellite microwave,, Bluetooth, and infrared light are all different forms of electromagnetic waves that are used to transmit data.

Guided and Unguided Transmission Media

- Unguided signal can travel from the source to destination in several ways:

1. Ground Propagation:

- Radio waves travel through the lowest portion of the atmosphere, hugging the earth.
- The low frequency signal follow the curvature of the planet.
- Distance depends on the amount of the power.

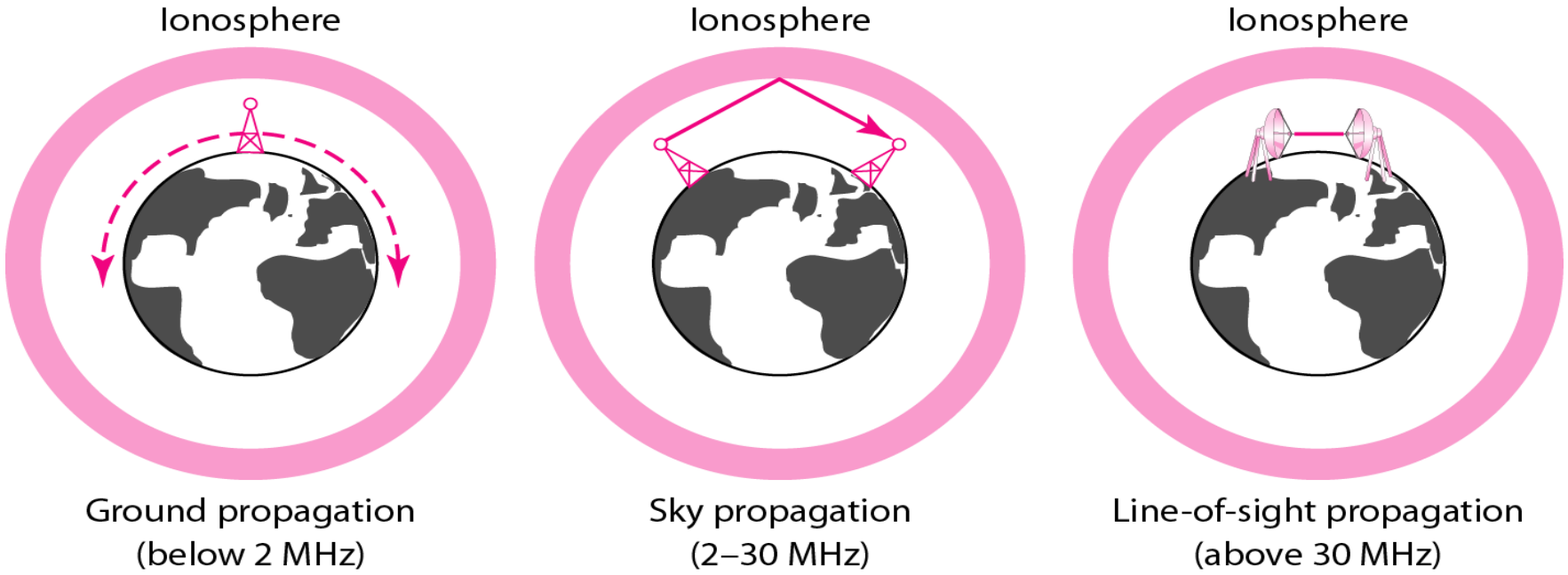
2. Sky Propagation:

- Higher frequency radio radiate upward into the ionosphere where they are reflected back to the earth.
- Sky propagation allow for greater distance with lower power output.

Guided and Unguided Transmission Media

3. line-of-sight Propagation:

- Very high frequency signals are transmitted in straight lines directly from antenna to antenna.



Guided and Unguided Transmission Media

- The section of the electromagnetic spectrum defined as radio waves and microwaves is divided into eight ranges, called bands, each regulated by government authorities.
- These bands are rated from very low frequency(VLF) to extremely high frequency (EHF).
- Table, lists these bands, their ranges, propagation methods, and some applications.

Table Bands

Band	*Range*	*Propagation*	*Application*
VLF (very low frequency)	3–30 kHz	Ground	Long-range radio navigation
LF (low frequency)	30–300 kHz	Ground	Radio beacons and navigational locators
MF (middle frequency)	300 kHz–3 MHz	Sky	AM radio
HF (high frequency)	3–30 MHz	Sky	Citizens band (CB), ship/aircraft communication
VHF (very high frequency)	30–300 MHz	Sky and line-of-sight	VHF TV, FM radio
UHF (ultrahigh frequency)	300 MHz–3 GHz	Line-of-sight	UHF TV, cellular phones, paging, satellite
SHF (superhigh frequency)	3–30 GHz	Line-of-sight	Satellite communication
EHF (extremely high frequency)	30–300 GHz	Line-of-sight	Radar, satellite

Guided and Unguided Transmission Media

Wireless Transmission Waves

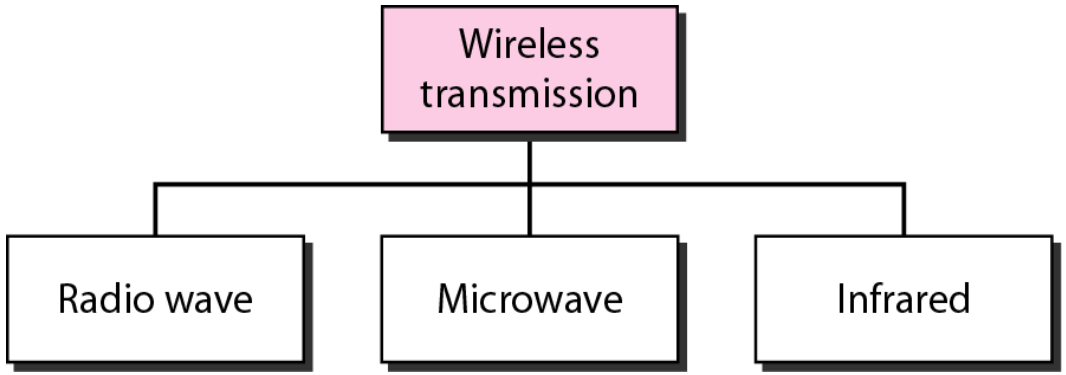


Figure Wireless transmission waves

Guided and Unguided Transmission Media

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. This means that the sending and receiving antennas do not have to be aligned. A sending antenna sends waves that can be received by any receiving antenna.

Guided and Unguided Transmission Media

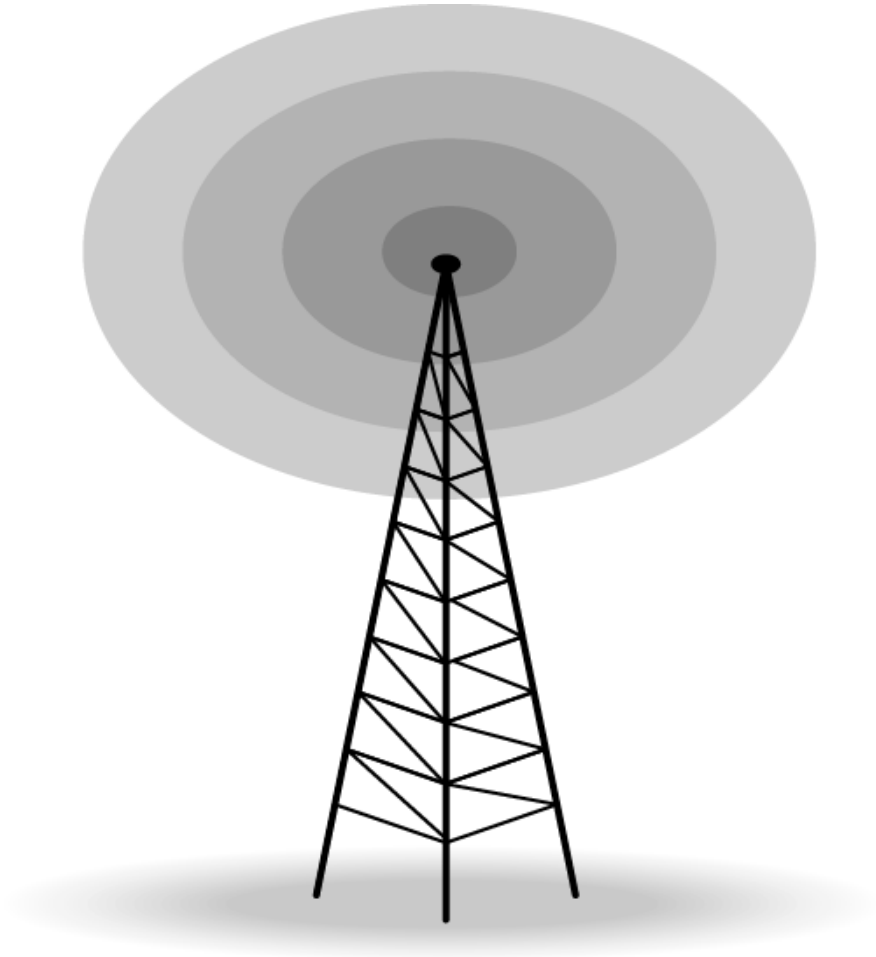


Figure Omnidirectional antenna

Guided and Unguided Transmission Media

- The omnidirectional property has a disadvantage, too. The radio waves transmitted by one antenna are susceptible to interference by another antenna that may send signals using the same frequency or band.
  - Between 3 KHz – 1 GHz.
  - Radio waves use omnidirectional antenna.
  - Radio waves used for multicast communication, such as radio and television.
  - Sky Propagation. This makes radio waves a good candidate for long-distance broadcasting such as AM radio.

Guided and Unguided Transmission Media

Microwave

- Microwave propagation is line-of-sight. Since the towers with the mounted antennas need to be in direct sight of each other, towers that are far apart need to be very tall. The curvature of the earth as well as other blocking obstacles do not allow two short towers to communicate by using microwaves. Repeaters are often needed for long distance communication.
- Very high-frequency microwaves cannot penetrate walls. This characteristic can be a disadvantage if receivers are inside buildings.
- The microwave band is relatively wide, almost 299 GHz. Therefore wider sub-bands can be assigned, and a high data rate is possible.
- Microwaves are used for unicast communication such as cellular telephones, satellite networks, and wireless LANs.

Guided and Unguided Transmission Media

- 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.
- *A parabolic dish antenna* is based on the geometry of a parabola: Every line parallel to the line of symmetry (line of sight) reflects off the curve at angles such that all the lines intersect in a common point called the focus.
- The parabolic dish works as a funnel, catching a wide range of waves and directing them to a common point. In this way, more of the signal is recovered than would be possible with a single-point receiver. Outgoing transmissions are broadcast through a horn aimed at the dish. The microwaves hit the dish and are deflected outward in a reversal of the receipt path.

Guided and Unguided Transmission Media

- *A horn antenna* looks like a gigantic scoop. Outgoing transmissions are broadcast up a stem (resembling a handle) and deflected outward in a series of narrow parallel beams by the curved head. Received transmissions are collected by the scooped shape of the horn, in a manner similar to the parabolic dish, and are deflected down into the stem.

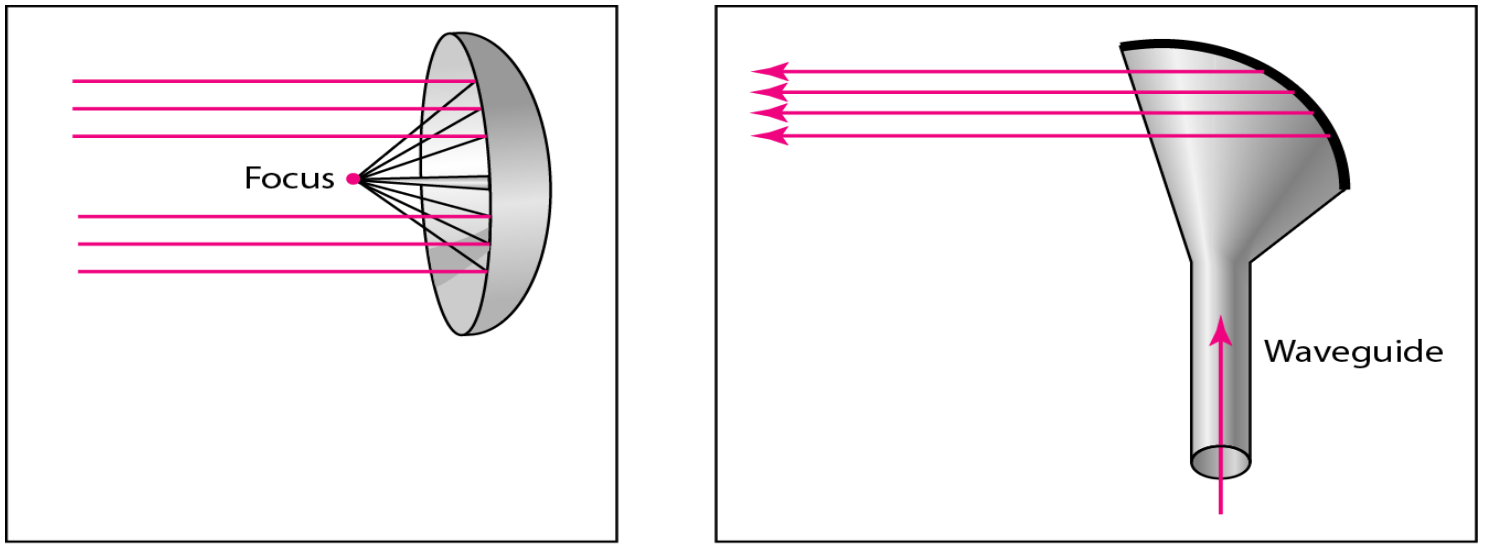


Figure Unidirectional antennas

Guided and Unguided Transmission Media

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. When we use our infrared remote control, we do not interfere with the use of the remote by our neighbors.
- However, this same characteristic makes infrared signals useless for long-range communication. In addition, we cannot use infrared waves outside a building because the sun's rays contain infrared waves that can interfere with the communication.

Guided and Unguided Transmission Media

- Between 300 GHz-400 THz
- Used for short-range communication.
- Very common with remote control devices, but can also be used for device-to-device transfers, such as PDA to computer.
- Line-of-sight propagation.

- Infrared signals can be used for short-range communication in a closed area using line-of-sight propagation.

Guided and Unguided Transmission Media

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.
- The Infrared Data Association (IrDA), an association for sponsoring the use of infrared waves, has established standards for using these signals for communication between devices such as keyboards, mice, PCs, and printers.

Network Devices

Hub

- Hub is a device that splits a network connection into multiple computers. It is like a distribution center. When a computer request information from a network or a specific computer, it sends the request to the hub through a cable. The hub will receive the request and transmit it to the entire network. Each computer in the network should then figure out whether the broadcast data is for them or not.

Network Devices

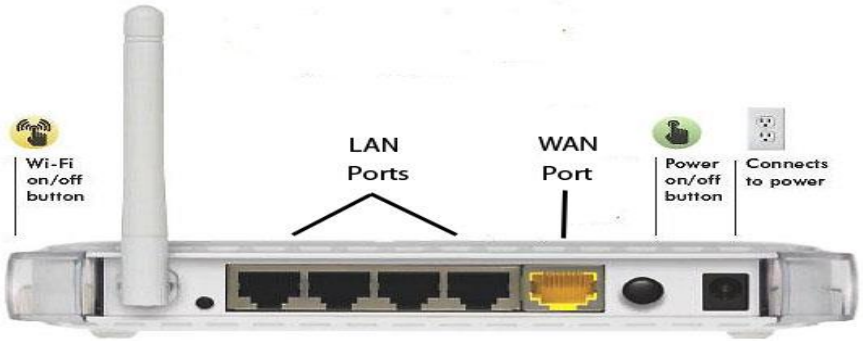
Switch

- Switch is a telecommunication device grouped as one of computer network components. Switch is like a Hub but built in with advanced features. It uses physical device addresses in each incoming messages so that it can deliver the message to the right destination or port.

Network Devices

Router

- When we talk about computer network components, the other device that used to connect a LAN with an internet connection is called Router. When you have two distinct networks (LANs) or want to share a single internet connection to multiple computers, we use a Router.
- In most cases, recent routers also include a switch which in other words can be used as a switch. You don't need to buy both switch and router, particularly if you are installing small business and home networks.



Network Devices

Modem

- A modem enables you to connect your computer to the available internet connection over the existing telephone line. Like NIC, Modem is not integrated with a computer motherboard. It comes as separate part which can be installed on the PCI slots found on motherboard.

Network Devices

Cables and Connector

- Cable is one way of transmission media which can transmit communication signals. The wired network topology uses special type of cable to connect computers on a network.

Hub	Switch	Router
Hub is a physical layer device i.e. layer 1.	Switch is a data link layer device i.e. layer 2.	Router is a network layer device i.e. layer 3.
A Hub works on the basis of broadcasting.	Switch works on the basis of MAC address.	A router works on the basis of IP address.
A Hub is a multiport repeater in which a signal introduced at the input of any port appears at the output of the all available ports.	A Switch is a tele-communication device which receives a message from any device connected to it and then transmits the message only to the device for which the message is intended.	A router reads the header of incoming packet and forward it to the port for which it is intended there by determines the route. It can also perform filtering and encapsulation.
Hub is not an intelligent device that may include amplifier on repeater.	A Switch is an intelligent device as it passes on the message to the selective device by inspecting the address.	A route is more sophisticated and intelligent device as it can read IP address and direct the packets to another network with specified IP address. Moreover routers can built address tables that helps in routing decisions.
At least single network is required to connect.	At least single network is required to connect.	Router needs at least two networks to connect.
Hub is cheaper as compared to switch and router.	Switch is an expensive device than hub.	Router is a relatively much more expensive device than hub and switch.
Speed of original hub 10Mbps and modern internet hub is 100Mbps.	maximum speed is 10Mbps to 100Mbps.	maximum speed for wireless is 1-10 Mbps and maximum speed for wired connections is 100 Mbps.
Hubs are used in LANs.	Switch is used in LANs.	Routers are used in LANs, MANs and WANs.

Fundamental of Circuit-Switched and Packet-Switched Network

- 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 shows a switched network.

Fundamental of Circuit-Switched and Packet-Switched Network

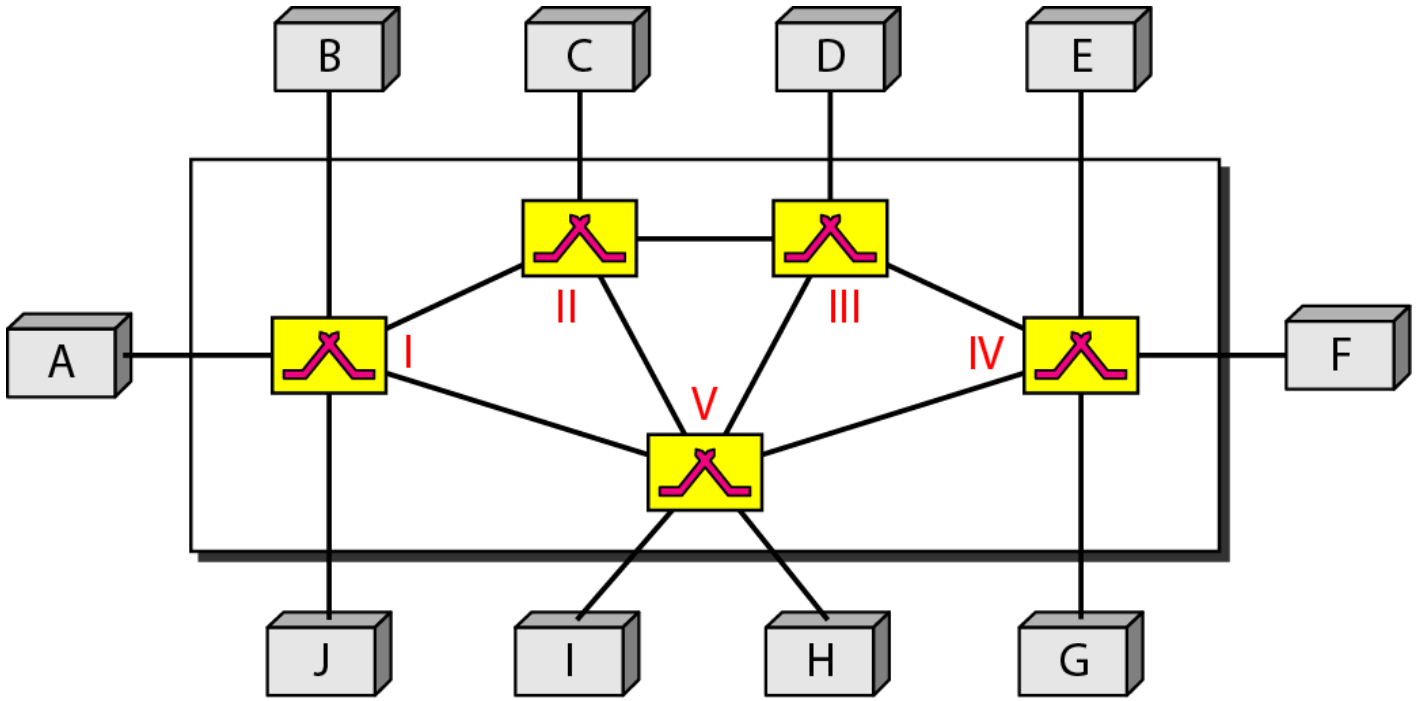


Figure **Switched network**

Methods of Switching

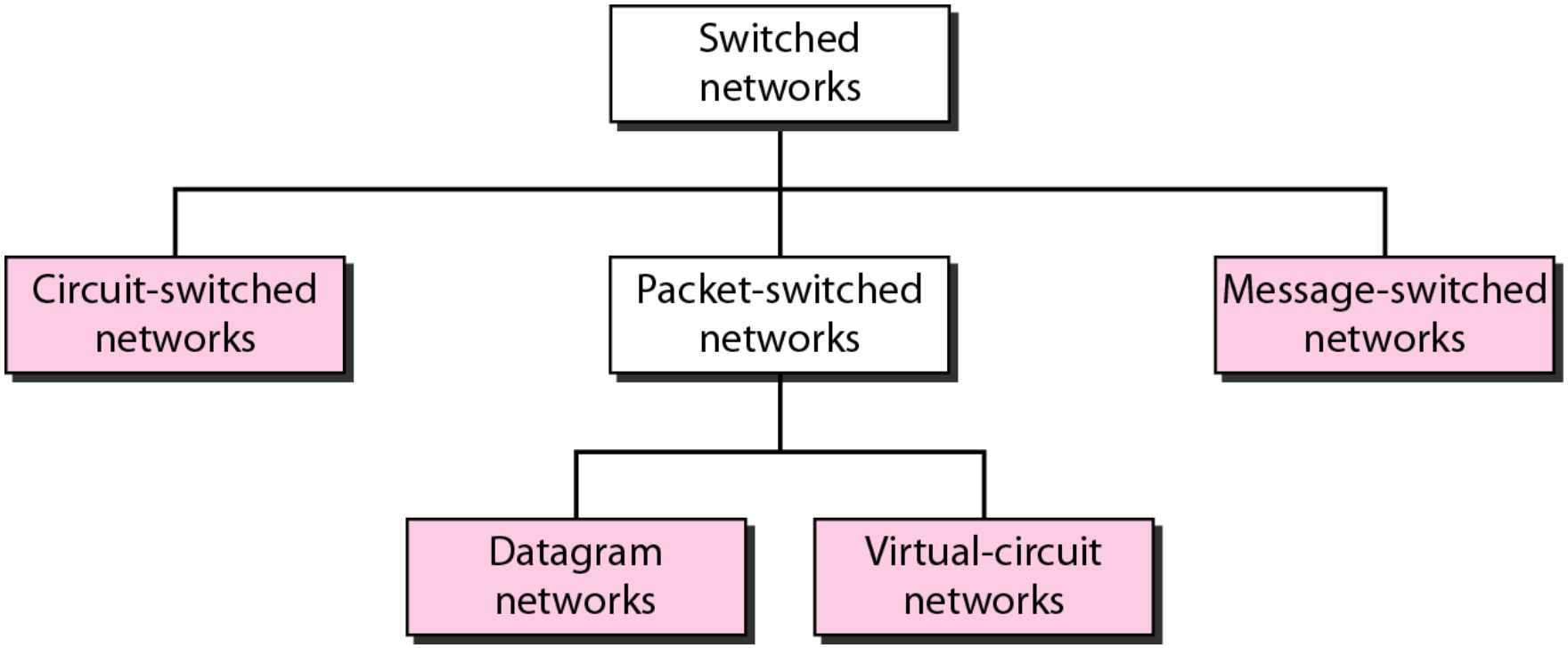


Figure **Taxonomy of switched networks**

Circuit-Switched Network

- 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.
- A circuit-switched network is made of a set of switches connected by physical links, in which each link is divided into *n* channels.

Circuit-Switched Network

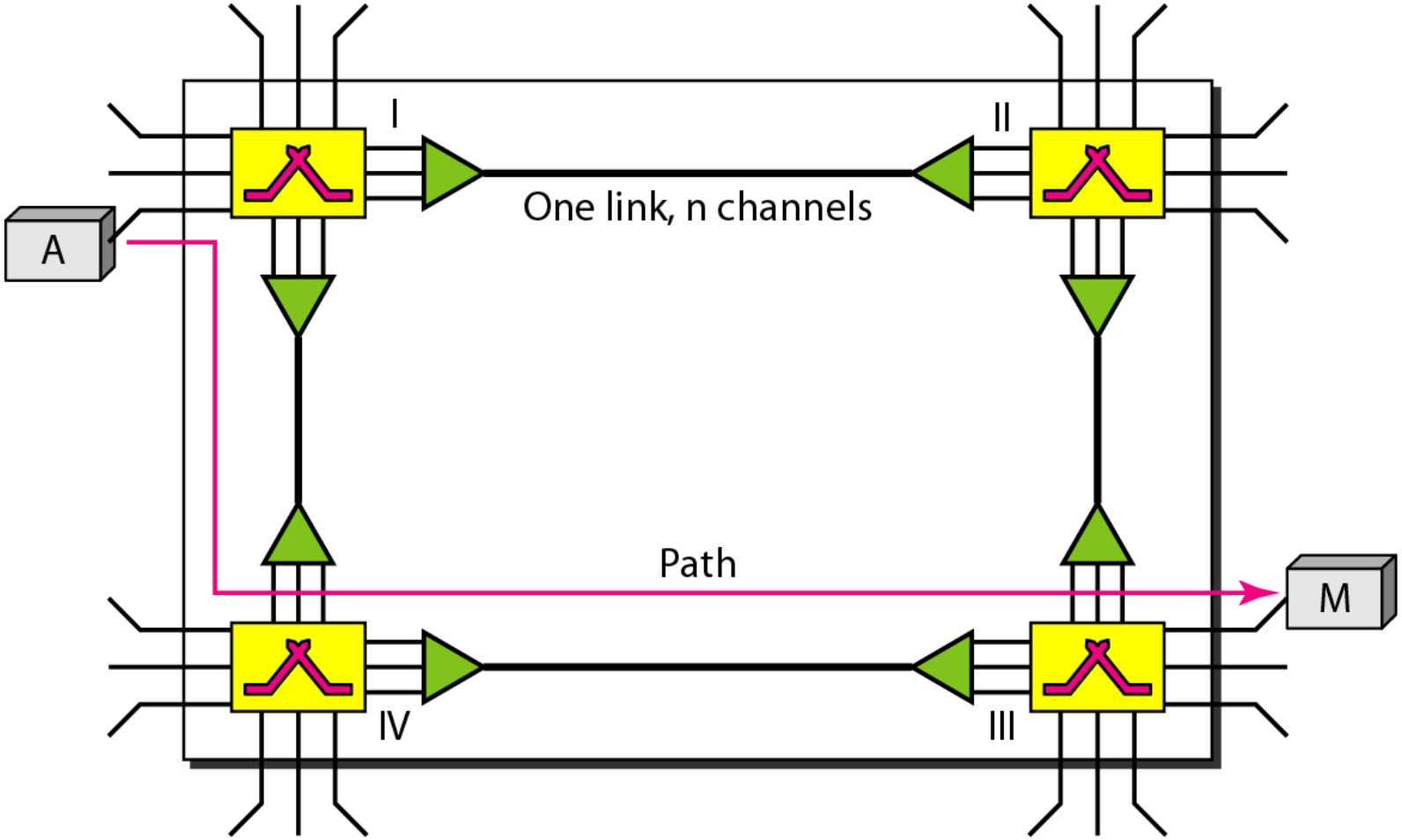


Figure **A trivial circuit-switched network**

Circuit-Switched Network

- Figure 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.
- 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, *data transfer* can take place.
- After all data have been transferred, the circuits are *teardown*.

Circuit-Switched Network

To emphasize several points here:

- Circuit switching takes place at the physical layer.

- Before starting communication, the stations must make a reservation for the resources to be used during the communication. These resources, such as channels (bandwidth in FDM and time slots in TDM), switch buffers, switch processing time, and switch input/output ports, must remain dedicated during the entire duration of data transfer until the teardown phase.

Circuit-Switched Network

- Data transferred between the two stations are not packetized (physical layer transfer of the signal). The data are a continuous flow sent by the source station and received by the destination station, although there may be periods of silence.

- There is no addressing involved during data transfer. The switches route the data based on their occupied band (FDM) or time slot (TDM). Of course, there is end-to-end addressing used during the setup phase.

Circuit-Switched Network

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 (combination of channels in links) 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.
- Note that end-to-end addressing is required for creating a connection between the two end systems. These can be, the addresses of the computers assigned by the administrator in a TDM network, or telephone numbers in an FDM network.

Circuit-Switched Network

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.

Circuit-Switched Network

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.

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.
- Figure shows the idea of delay in a circuit-switched network when only two switches are involved.

Circuit-Switched Network

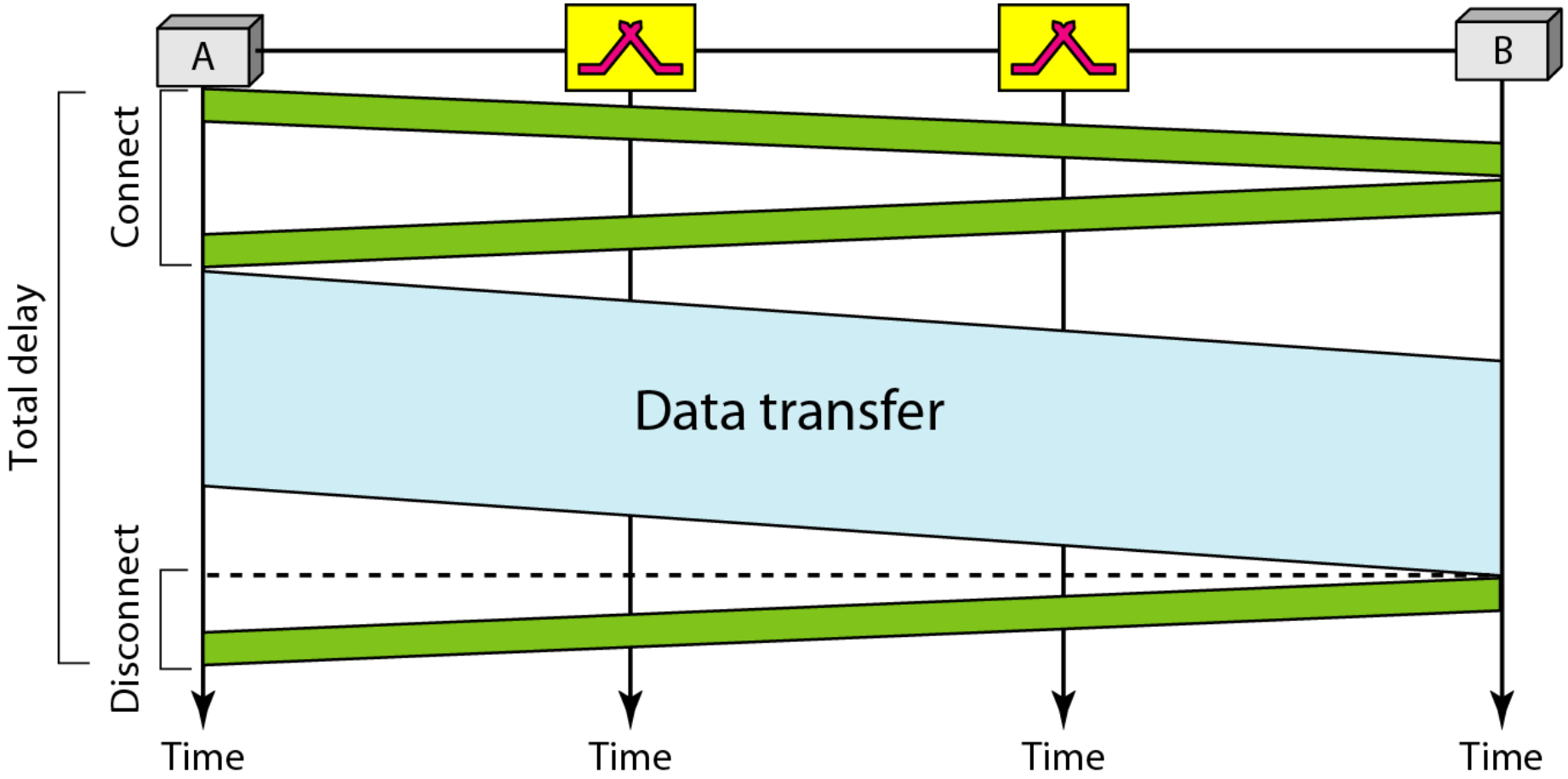


Figure **Delay in a circuit-switched network**

Packet-Switched Network

- 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.

Packet-Switched Network

- When a switch receives a packet, no matter what is the source or destination, the packet must wait if there are other packets being processed.
- In a datagram network, each packet is treated independently of all others. Even if a packet is part of a multi-packet 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 shows how the datagram approach is used to deliver four packets from station *A* to station *X*.

Packet-Switched Network

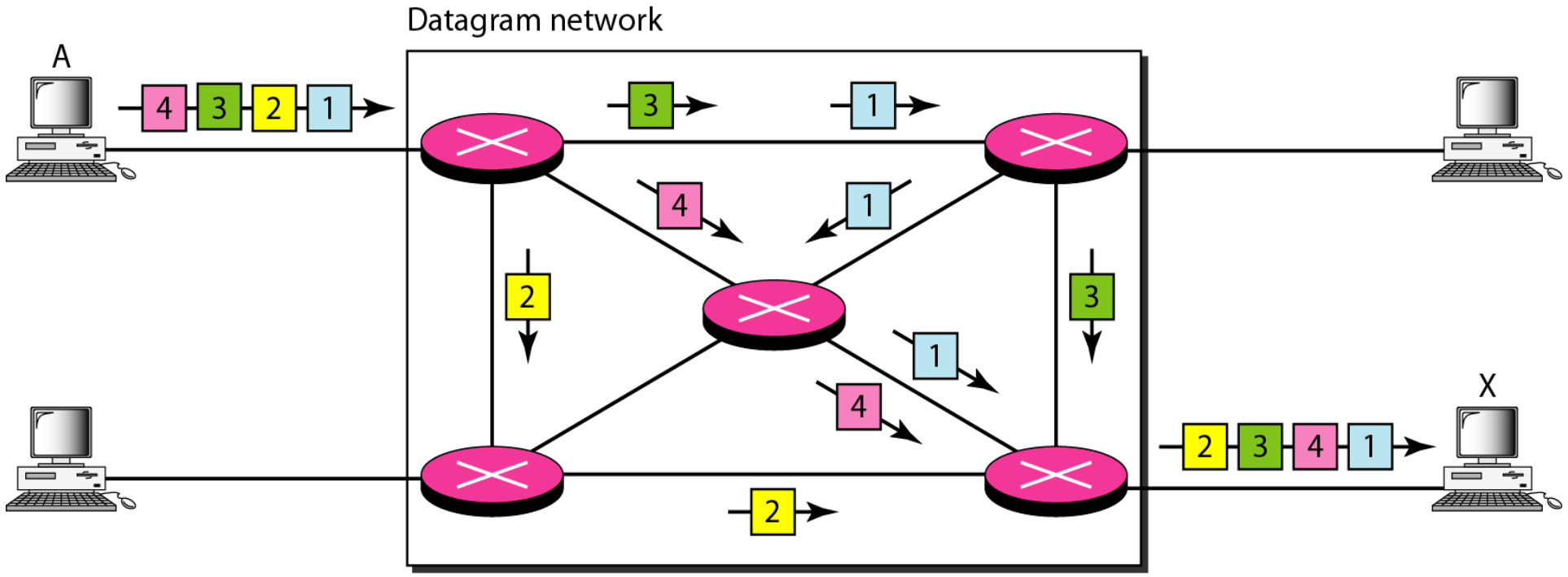


Figure **A datagram network with four switches (routers)**

Packet-Switched Network

- In this example, all four packets (or datagrams) belong to the same message, but may travel different paths to reach their destination. This is so because the links may be involved in carrying packets from other sources and do not have the necessary bandwidth available to carry all the packets from *A* to *X*.
- 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.

Packet-Switched Network

Routing Table

- If there are no setup or teardown phases, how are the packets routed to their destinations in a datagram network?
- In this type of network, each switch (or packet switch) has a routing table which is based on the destination address. The routing tables are dynamic and are updated periodically.

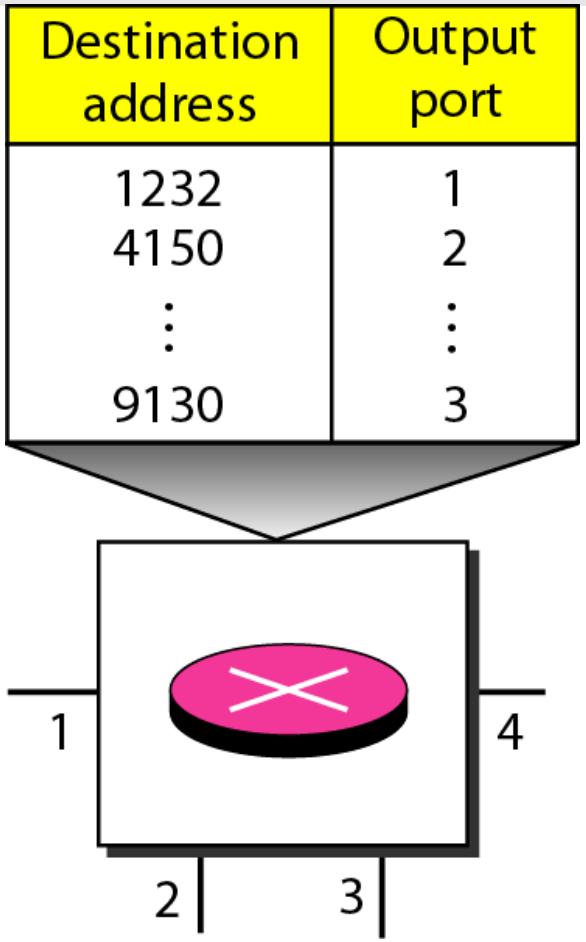


Figure **Routing table in a datagram network**

Packet-Switched Network

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. Since not all packets in a message necessarily travel through the same switches, the delay is not uniform for the packets of a message.

Packet-Switched Network

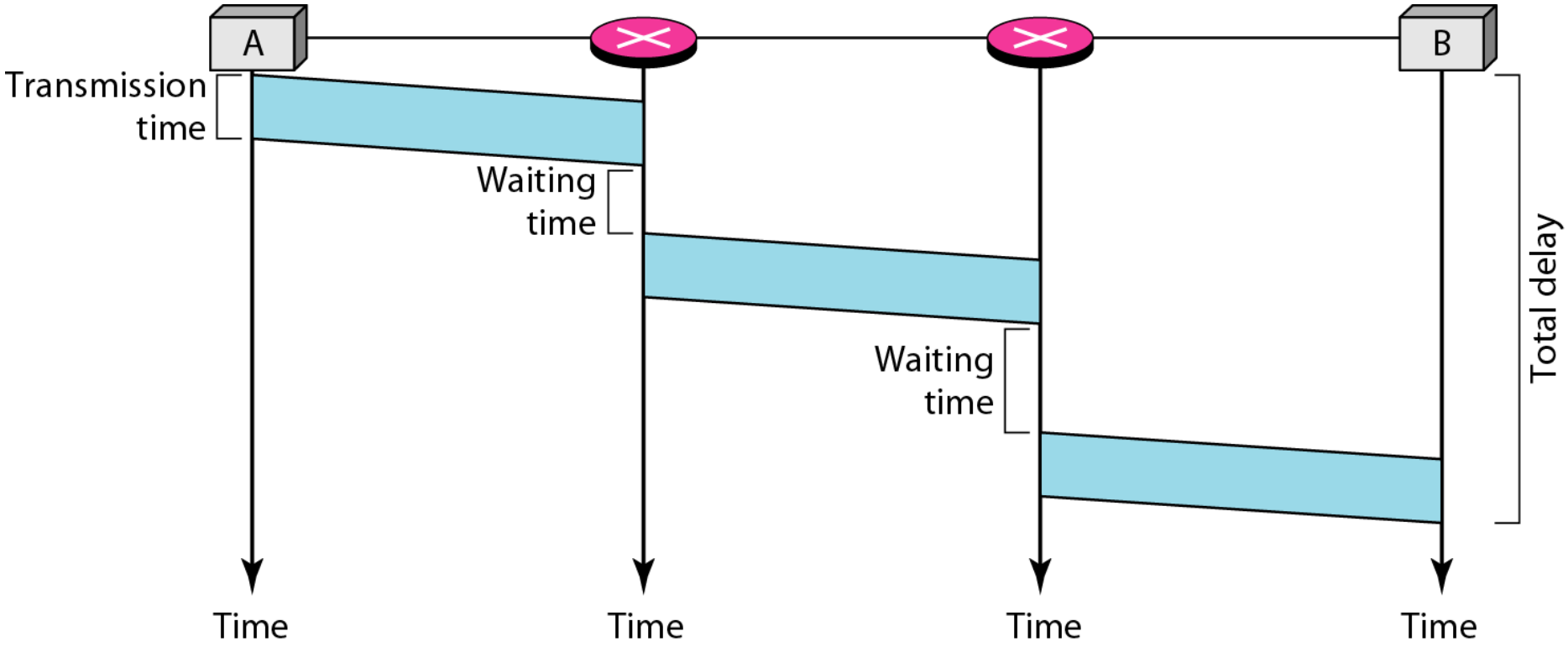


Figure **Delay in a datagram network**

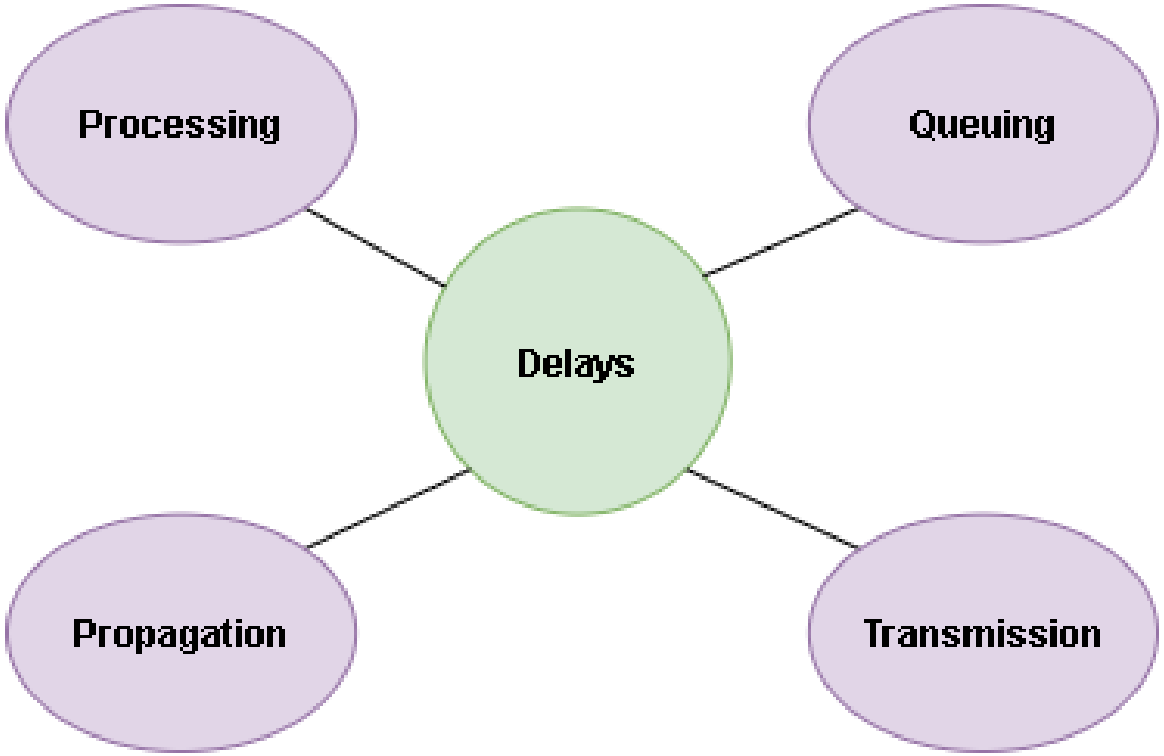
Delay in Packet-Switching Network

- A *delay* in computer networking is the time taken for a packet to go from its sender to its receiver.

- The packet starts from a source, passes through some routers, and then reaches its destination. When it passes from one node (Host or Router) to another node (Host or Router), it experiences different delays on its path at each node.

Delay in Packet-Switching Network

- **Types of Delay**



Delay in Packet-Switching Network

1. Transmission Delay:

- The time taken to transmit a packet from the host to the transmission medium is called *Transmission delay*.

- For example, if bandwidth is 1 bps (every second 1 bit can be transmitted onto the transmission medium) and data size is 20 bits then what is the transmission delay? If in one second, 1 bit can be transmitted. To transmit 20 bits, 20 seconds would be required.

Delay in Packet-Switching Network

- Let B bps is the bandwidth and L bit is the size of the data then transmission delay is,

$$Ttrans = L/B$$

- This delay depends upon the following factors:
  - If there are multiple active sessions, the delay will become significant.
  - Increasing bandwidth decreases transmission delay.
  - MAC protocol largely influences the delay if the link is shared among multiple devices.
  - Sending and receiving a packet involves a context switch in the operating system, which takes a finite time.

Delay in Packet-Switching Network

2. Propagation delay:

- After the packet is transmitted to the transmission medium, it has to go through the medium to reach the destination.
- Hence the time taken by the last bit of the packet to reach the destination is called *propagation delay*.

Delay in Packet-Switching Network

- Factors affecting propagation delay:
  - **Distance** – It takes more time to reach the destination if the distance of the medium is longer.
  - **Velocity** – If the velocity(speed) of the signal is higher, the packet will be received faster.

$$Tprop = Distance/Velocity$$

Delay in Packet-Switching Network

3. Queueing delay:

- Let the packet is received by the destination, the packet will not be processed by the destination immediately. It has to wait in a queue in something called a buffer.
- So the amount of time it waits in queue before being processed is called *queueing delay*.
- This delay depends upon the following factors:
  - If the size of the queue is large, the queuing delay will be huge. If the queue is empty there will be less or no delay.
  - If more packets are arriving in a short or no time interval, queuing delay will be large.
  - The less the number of servers/links, the greater is the queuing delay.

Delay in Packet-Switching Network

4. Processing delay:

- Now the packet will be taken for the processing which is called *processing delay*.
- Time is taken to process the data packet by the processor that is the time required by intermediate routers to decide where to forward the packet, update TTL, perform header checksum calculations.

- The sum of all delays (Transmission, Propagation, Queuing, and Processing) is known as the **Nodal Delay** and is given as:

$$\mathbf{T}_{\text{nodal}} = \mathbf{T}_{\text{trans}} + \mathbf{T}_{\text{prop}} + \mathbf{T}_{\text{queue}} + \mathbf{T}_{\text{proc}}$$

Packet Loss in Packet-Switching Network

- In the **queuing delay**, it was mentioned that the delay time will approach infinity; this is of course not realistic as the queues usually has a finite number of packets which can be queued.
- Basically this means that when sending a lot of packets in to a queue at a high rate (or at the same time), packet loss will be experienced as the queue will be m*axed out* and the router will drop packets.
- This will start a "chain-reaction" of increasing the rate of incoming packets, as the dropped packets will need to be retransmitted to the router.
- A lost packet can be retransmitted on an end-to-end basis in order to ensure that all data are eventually transferred from source to destination.

Throughput in Packet-Switching Network

- In data transmission, network **throughput** is the amount of data moved successfully from one place to another in a given time period, and typically measured in bits per second (bps), as in megabits per second (Mbps) or gigabits per second (Gbps).

- The **instantaneous throughput** at any instant of time is the rate (in bits/sec) at which *Host B* is receiving the file.

- The **average throughput** of the file is F/T bits/sec, where the file consists of F bits and the transfer time is T (in seconds).

Connection-oriented & Connectionless Services

No.	Connection-oriented Service	Connection-less Service
1.	Connection-oriented service is related to the telephone system.	Connection-less service is related to the postal system.
2.	Connection-oriented service is preferred by long and steady communication.	Connection-less Service is preferred by bursty communication.
3.	Connection-oriented Service is necessary.	Connection-less Service is not compulsory.
4.	Connection-oriented Service is feasible.	Connection-less Service is not feasible.
5.	In connection-oriented Service, Congestion is not possible.	In connection-less Service, Congestion is possible.
6.	Connection-oriented Service gives the guarantee of reliability.	Connection-less Service does not give a guarantee of reliability.
7.	In connection-oriented Service, Packets follow the same route.	In connection-less Service, Packets do not follow the same route.
8.	Connection-oriented services require a bandwidth of a high range.	Connection-less Service requires a bandwidth of low range.
9.	Ex: TCP (Transmission Control Protocol)	Ex: UDP (User Datagram Protocol)

Summary

- Understanding of Network and Internet,
- Network Topologies,
- The OSI Model,
- TCP/IP Protocol Suite,
- Guided and Unguided Transmission Media,
- Network Devices,
- Fundamental of Circuit-Switched and Packet-Switched Networks,
- Performance Metrics,
- Understanding of Delay, Loss and Throughput in the packet-switching network

Thank You

Unit 2

Data Link Layer: Logical Link Control Sublayer

Prof. Keyur J Patel

Outline

- Introduction and Design Issues,
- Flow and Error Control,
- Techniques for Error Detection and Correction,
- Elementary Data Link Layer Protocols: Simplex, Stop and Wait, Sliding Window Protocol

Introduction of Data Link Layer

- Data link layer performs the most reliable **node to node** delivery of data. It forms frames from the packets that are received from network layer and gives it to physical layer.
- It also synchronizes the information which is to be transmitted over the data. Error controlling is easily done. The encoded data are then passed to physical.
- Error detection bits are used by the data link layer. It also corrects the errors. Outgoing messages are assembled into frames. Then the system waits for the acknowledgements to be received after the transmission. It is reliable to send message.

Introduction of Data Link Layer

- The main task of the **data link layer** is to transform a raw transmission facility into a line that appears free of undetected transmission errors to the network layer.
- It accomplishes this task by having the sender break up the input data into **data frames**(typically a few hundred or few thousand bytes) and transmit the frames sequentially. If the service is reliable, the receiver confirms correct receipt of each frame by send back an **acknowledgement frame**.

Introduction of Data Link Layer

Functions of Data Link Layer

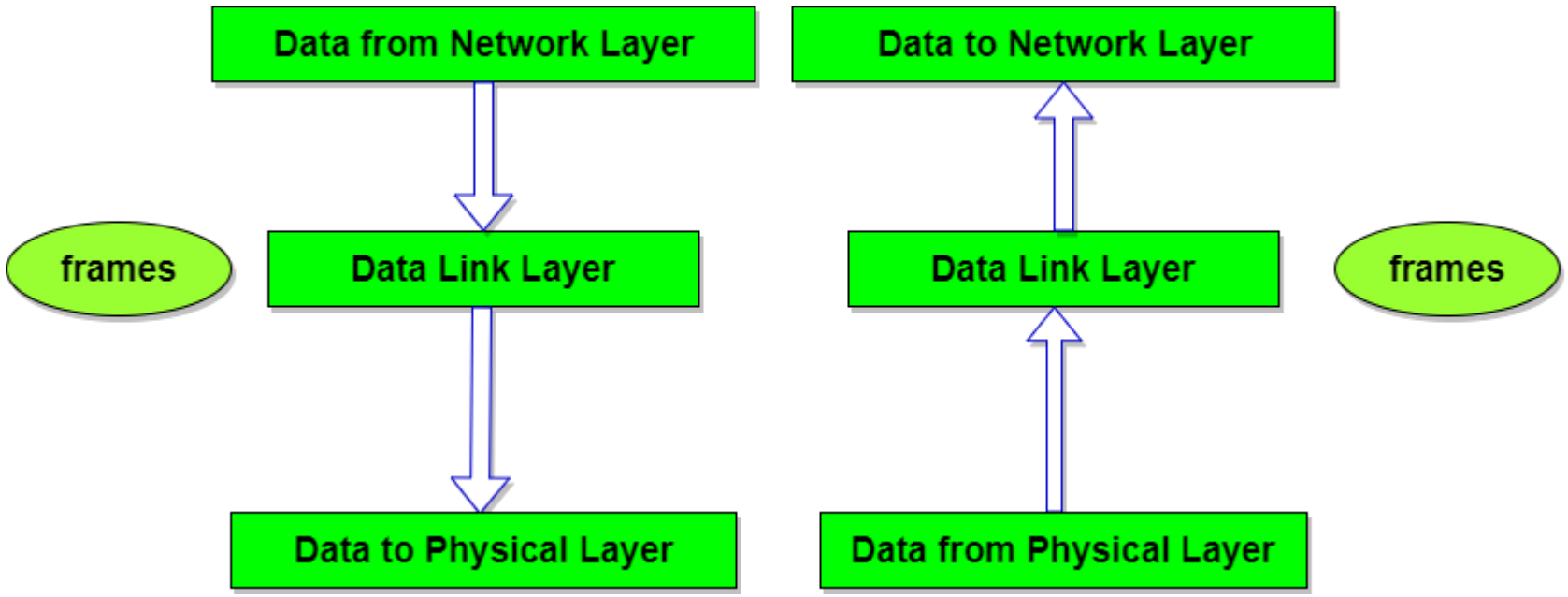
- **Framing:** Frames are the streams of bits received from the network layer into manageable data units. This division of stream of bits is done by Data Link Layer.
- **Physical Addressing:** The Data Link layer adds a header to the frame in order to define physical address of the sender or receiver of the frame, if the frames are to be distributed to different systems on the network.
- **Flow Control:** A flow control mechanism to avoid a fast transmitter from running a slow receiver by buffering the extra bit is provided by flow control. This prevents traffic jam at the receiver side.

Introduction of Data Link Layer

- **Error Control:** Error control is achieved by adding a trailer at the end of the frame. Duplication of frames are also prevented by using this mechanism. Data Link Layers adds mechanism to prevent duplication of frames.

- **Access Control:** Protocols of this layer determine which of the devices has control over the link at any given time, when two or more devices are connected to the same link.

Introduction of Data Link Layer



Design Issues with Data Link Layer

- The issue that arises in the data link layer(and most of the higher layers as well) is how to keep a fast transmitter from drowning a slow receiver in data. Some traffic regulation mechanism is often needed to let the transmitter know how much buffer space the receiver has at the moment. Frequently, the flow regulation and the error handling are integrated.
- Broadcast networks have an additional issue in the data link layer: How to control access to the shared channel. A special sublayer of the data link layer, the Medium Access Control(MAC) sublayer, deals with this problem.

Flow and Error Control

Flow Control

- Flow control coordinates the amount of data that can be sent before receiving an acknowledgment and is one of the most important duties of the data link layer.
- In most protocols, flow control is a set of procedures that tells the sender how much data it can transmit before it must wait for an acknowledgment from the receiver.
- The flow of data must not be allowed to overwhelm the receiver. Any receiving device has a limited speed at which it can process incoming data and a limited amount of memory in which to store incoming data.

Flow and Error Control

- The receiving device must be able to inform the sending device before those limits are reached and to request that the transmitting device send fewer frames or stop temporarily.
- Incoming data must be checked and processed before they can be used. The rate of such processing is often slower than the rate of transmission.
- For this reason, each receiving device has a block of memory, called a buffer, reserved for storing incoming data until they are processed. If the buffer begins to fill up, the receiver must be able to tell the sender to halt transmission until it is once again able to receive.

Flow and Error Control

Error Control

- Error control is both error detection and error correction. It allows the receiver to inform the sender of any frames lost or damaged in transmission and coordinates the retransmission of those frames by the sender.
- In the data link layer, the term error control refers primarily to methods of error detection and retransmission.
- Error control in the data link layer is often implemented simply: Any time an error is detected in an exchange, specified frames are retransmitted. This process is called automatic repeat request (ARQ).

Introduction of Error

- Data can be corrupted during transmission.

- Some applications require that errors be detected and corrected.

Introduction of Error

Types of Errors

- Whenever bits flow from one point to another, they are subject to unpredictable changes because of interference.
- This interference can change the shape of the signal.
- In a *single-bit error*, a 0 is changed to a 1 or a 1 to a 0.
- In a *burst error*, multiple bits are changed.

Introduction of Error

Single-Bit Error

- The term single-bit error means that only 1 bit of a given data unit (such as a byte, character, or packet) is changed from 1 to 0 or from 0 to 1.
- Figure shows the effect of a single-bit error on a data unit.

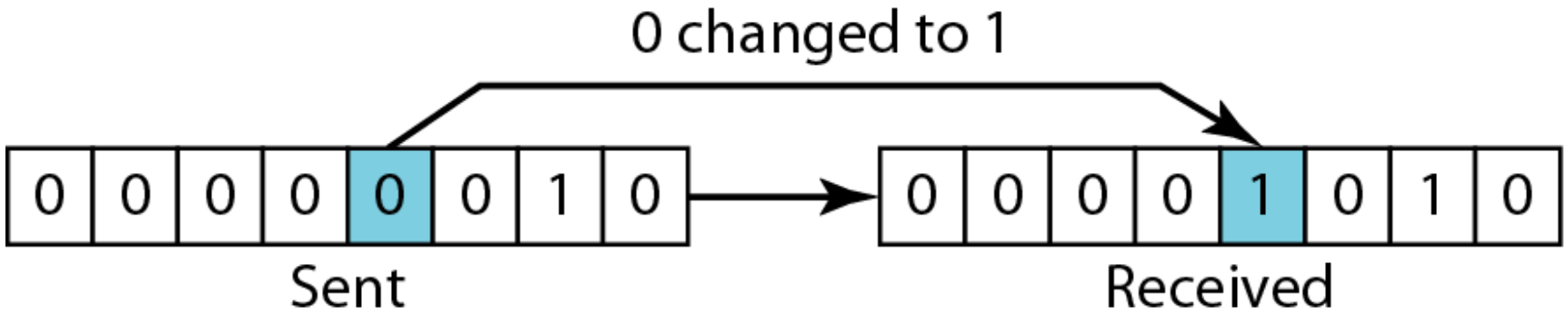


Figure Single-bit error

Introduction of Error

Burst Error

- The term burst error means that 2 or more bits in the data unit have changed from 1 to 0 or from 0 to 1.
- *Note that* a burst error does not necessarily mean that the errors occur in consecutive bits.
- The length of the burst is measured from the first corrupted bit to the last corrupted bit. Some bits in between may not have been corrupted.
- Figure shows the effect of a burst error on a data unit.

Introduction of Error

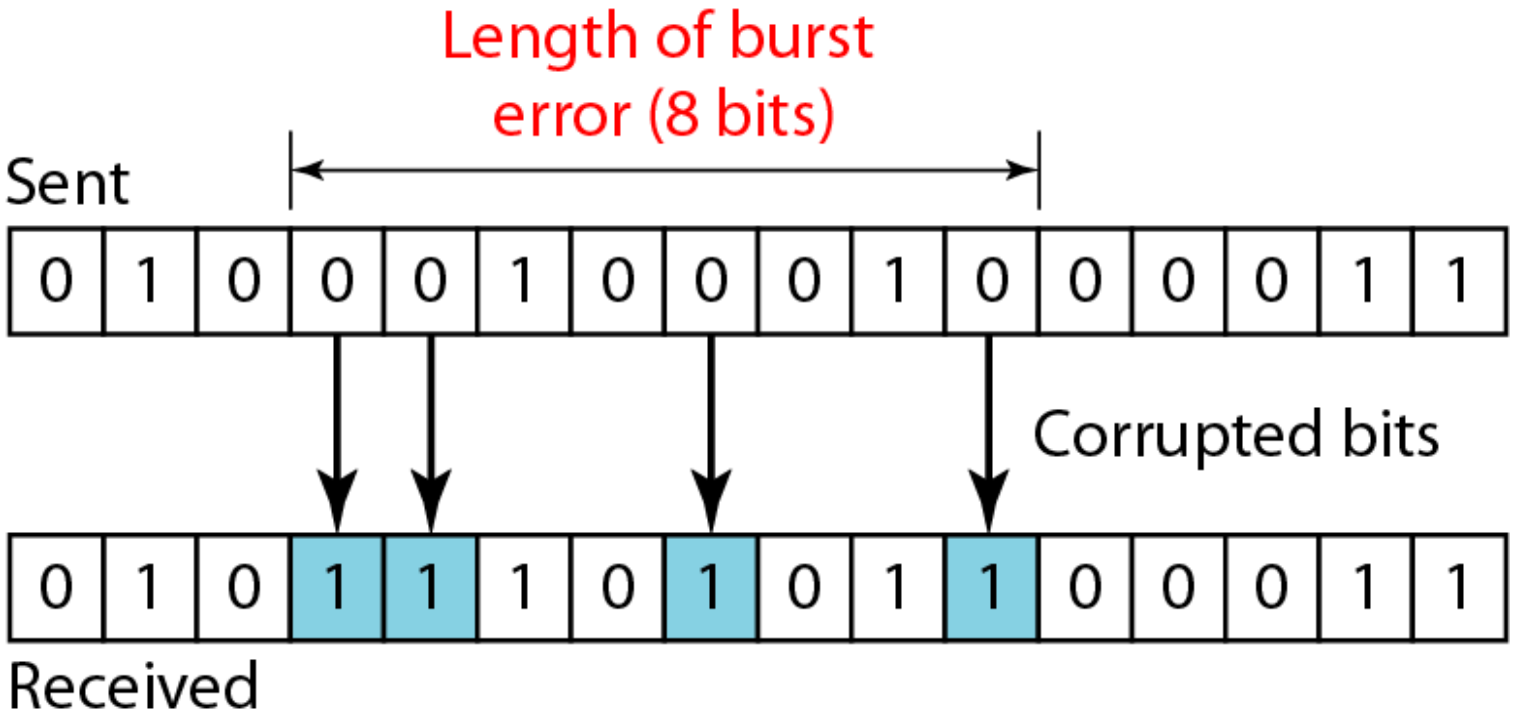


Figure Burst error of length 8

Introduction of Error

Redundancy

- The central concept in detecting or correcting errors is redundancy.
- To be able to detect or correct errors, we need to send some extra bits with our data.
- These redundant bits are added by the sender and removed by the receiver.
- Their presence allows the receiver to detect or correct corrupted bits.

- *To detect or correct errors, we need to send **extra (redundant)** bits with data.*

Techniques for Error Detection & Correction

Detection Versus Correction

- The correction of errors is more difficult than the detection.
- In error detection, we are looking only to see if any error has occurred. The answer is a simple yes or no. We are not even interested in the number of errors. A single-bit error is the same for us as a burst error.
- In error correction, we need to know the exact number of bits that are corrupted and more importantly, their location in the message. The number of the errors and the size of the message are important factors.

Techniques for Error Detection & Correction

Coding

- Redundancy is achieved through various coding schemes.
- The sender adds redundant bits through a process that creates a relationship between the redundant bits and the actual data bits.
- The receiver checks the relationships between the two sets of bits to detect or correct the errors.
- The ratio of redundant bits to the data bits and the robustness of the process are important factors in any coding scheme. Figure shows the general idea of coding.
- It can divide coding schemes into two broad categories: *block coding* and *convolution coding*.

Techniques for Error Detection & Correction

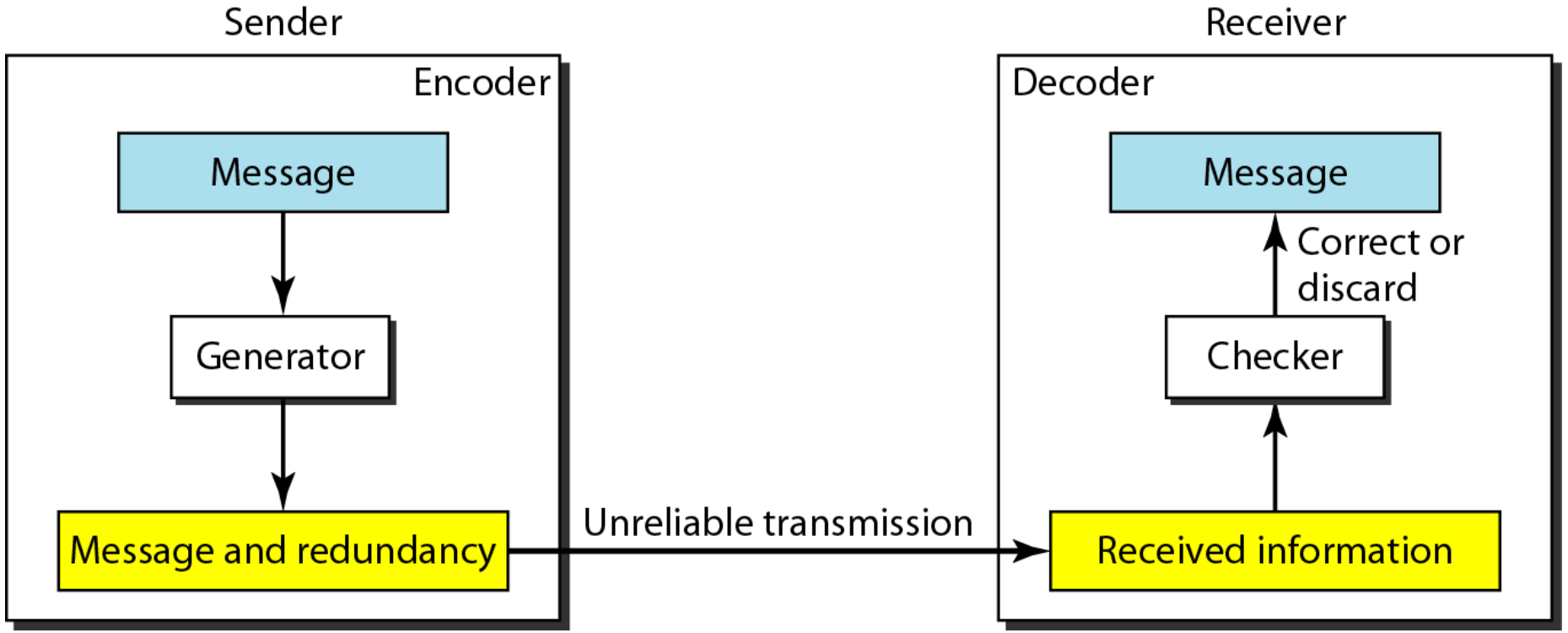


Figure The structure of encoder and decoder

Techniques for Error Detection & Correction

Block Coding

- In block coding, we divide our message into blocks, each of k bits, called datawords. We add r redundant bits to each block to make the length $n = k + r$. The resulting n-bit blocks are called codewords.
- It is important to know that we have a set of datawords, each of size k, and a set of codewords, each of size of n.
- With k bits, we can create a combination of 2^k datawords; with n bits, we can create a combination of 2^n codewords.
- Since $n > k$, the number of possible codewords is larger than the number of possible datawords.

Techniques for Error Detection & Correction

- The block coding process is one-to-one; the same dataword is always encoded as the same codeword. This means that we have $2^n - 2^k$ codewords that are not used. We call these codewords invalid or illegal.

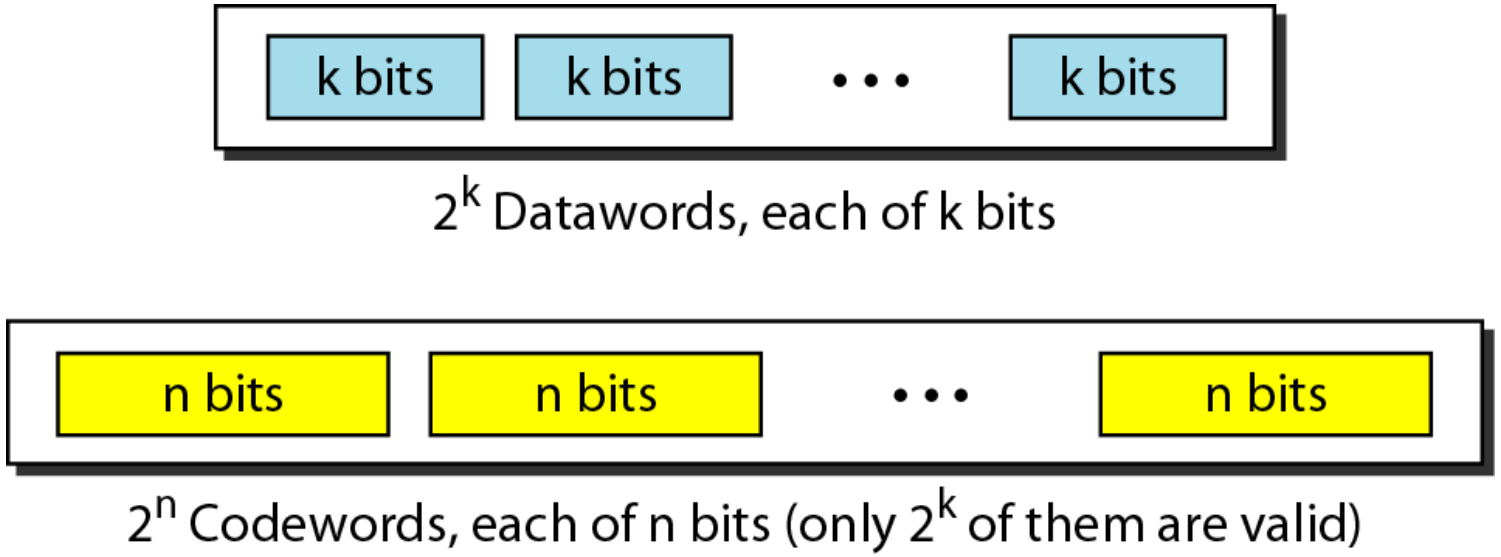


Figure Datawords and codewords in block coding

Techniques for Error Detection & Correction

Example

- The 4B/5B block coding example of this type of coding. In this coding scheme, $k = 4$ and $n = 5$.

Answer

- As we saw, we have $2^k = 16$ datawords and $2^n = 32$ codewords. We saw that 16 out of 32 codewords are used for message transfer and the rest are either used for other purposes or unused.

Techniques for Error Detection & Correction

Error Detection

- How can errors be detected by using block coding?
- If the following two conditions are met, the receiver can detect a change in the original codeword.
  1. The receiver has (or can find) a list of valid codewords.
  2. The original codeword has changed to an invalid one.

- Figure shows the role of block coding in error detection.

Techniques for Error Detection & Correction

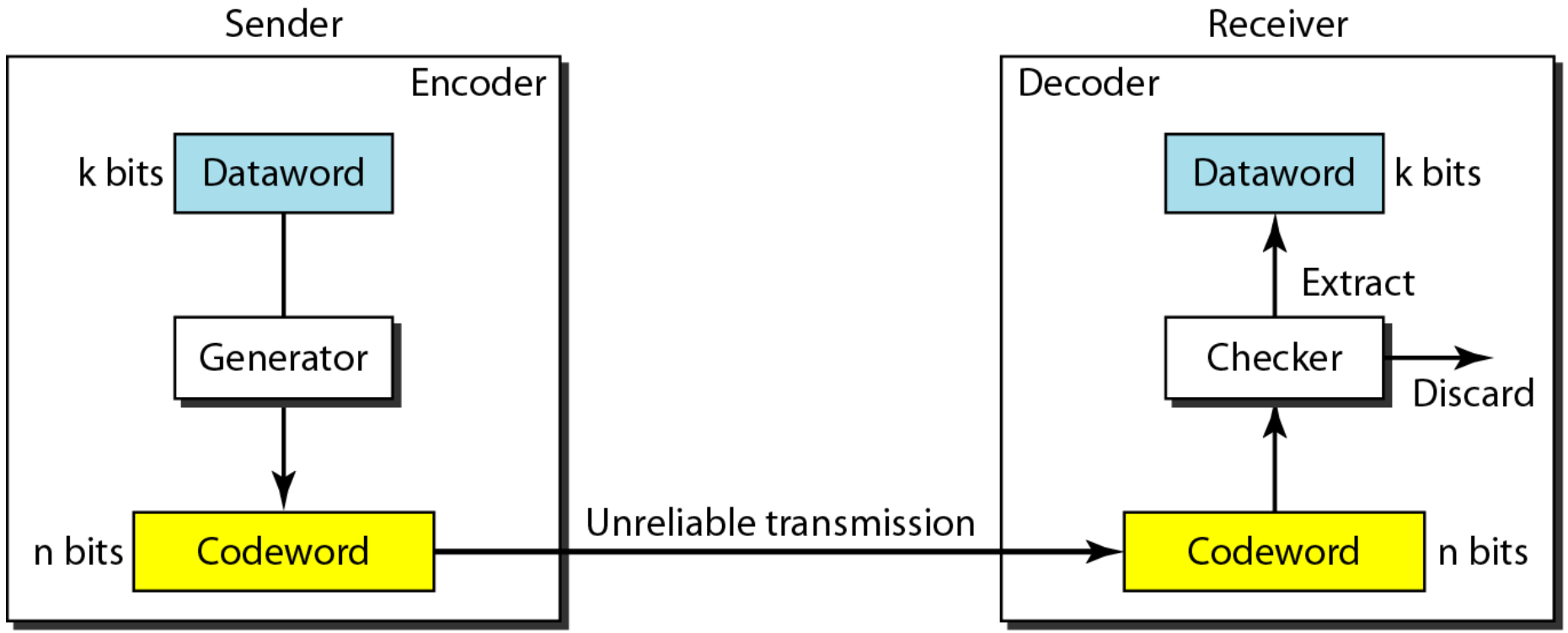


Figure Process of error detection in block coding

Techniques for Error Detection & Correction

- The sender creates codewords out of datawords by using a generator that applies the rules and procedures of encoding.
- Each codeword sent to the receiver may change during transmission. If the received codeword is the same as one of the valid codewords, the word is accepted; the corresponding dataword is extracted for use.
- If the received codeword is not valid, it is discarded. However, if the codeword is corrupted during transmission but the received word still matches a valid codeword, the error remains undetected.
- This type of coding can detect only single errors. Two or more errors may remain undetected.

Techniques for Error Detection & Correction

Example 1

- Let us assume that $k = 2$ and $n = 3$. Table 1 shows the list of datawords and codewords. Now, we will see how to derive a codeword from a dataword.

Table 1 A code for error detection

Datawords	*Codewords*
00	000
01	011
10	101
11	110

Techniques for Error Detection & Correction

- Assume the sender encodes the dataword 01 as 011 and sends it to the receiver. Consider the following cases:

1. The receiver receives 011. It is a valid codeword. The receiver extracts the dataword 01 from it.
2. The codeword is corrupted during transmission, and 111 is received. This is not a valid codeword and is discarded.
3. The codeword is corrupted during transmission, and 000 is received. This is a valid codeword. The receiver incorrectly extracts the dataword 00. Two corrupted bits have made the error undetectable.

Techniques for Error Detection & Correction

Error Correction

- As in, error correction is much more difficult than error detection.
- In error detection, the receiver needs to know only that the received codeword is invalid; in error correction the receiver needs to find (or guess) the original codeword sent.
- We can say that we need more redundant bits for error correction than for error detection.
- Figure shows the role of block coding in error correction.

Techniques for Error Detection & Correction

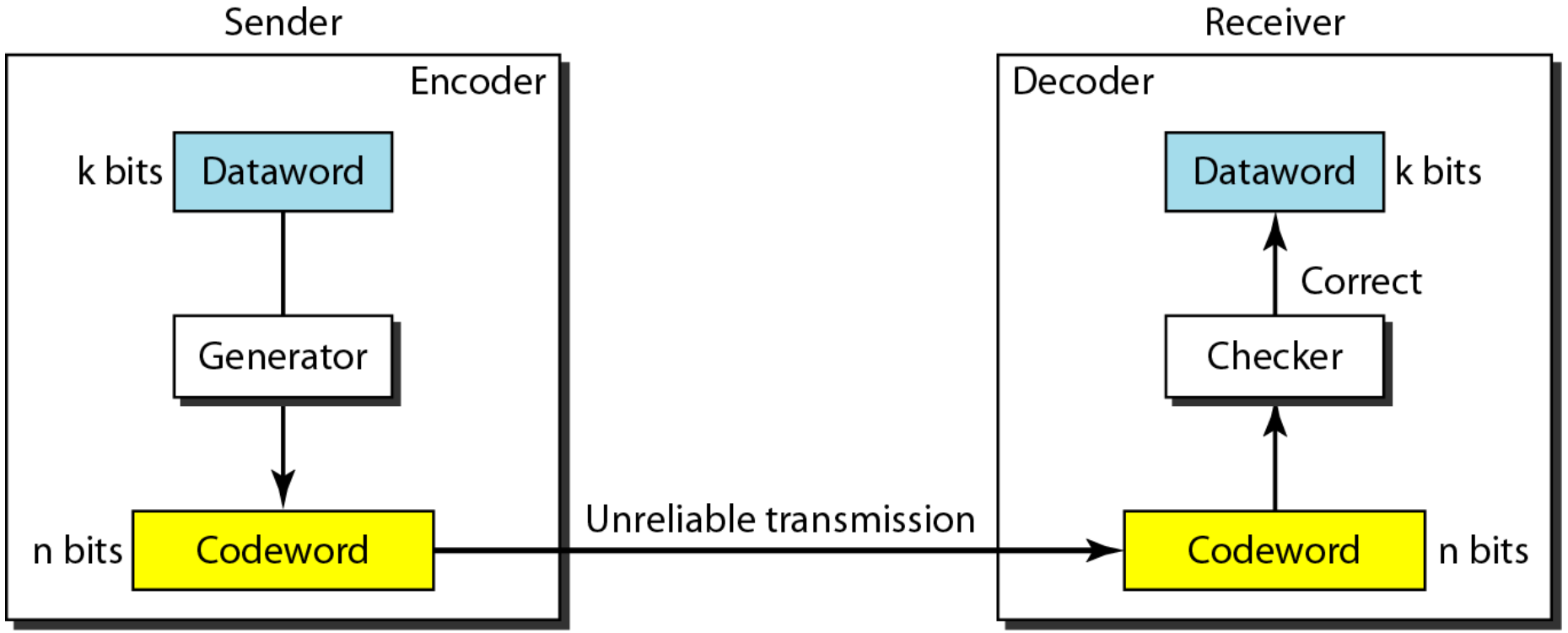


Figure Structure of encoder and decoder in error correction

Techniques for Error Detection & Correction

Example 2

- Let us add more redundant bits to Example 1 to see if the receiver can correct an error without knowing what was actually sent. We add 3 redundant bits to the 2-bit dataword to make 5-bit codewords. Table 2 shows the datawords and codewords.

Table 2 **A code for error correction**

Dataword	*Codeword*
00	00000
01	01011
10	10101
11	11110

Techniques for Error Detection & Correction

- Assume the dataword is 01. The sender creates the codeword 01011. The codeword is corrupted during transmission, and 01001 is received.
- First, the receiver finds that the received codeword is not in the table. This means an error has occurred.
- The receiver, assuming that there is only 1 bit corrupted, uses the following strategy to guess the correct dataword.
    1. Comparing the received codeword with the first codeword in the table (01001 versus 00000), the receiver decides that the first codeword is not the one that was sent because there are two different bits.

Techniques for Error Detection & Correction

2. By the same reasoning, the original codeword cannot be the third or fourth one in the table.

3. The original codeword must be the second one in the table because this is the only one that differs from the received codeword by 1 bit. The receiver replaces 01001 with 01011 and consults the table to find the dataword 01.

Techniques for Error Detection & Correction

Hamming Distance

- The Hamming distance between two words (of the same size) is the number of differences between the corresponding bits.
- The Hamming distance between two words *x* and *y* as *d(x, y)*.
- The Hamming distance can easily be found if apply the XOR operation (⊕) on the two words and count the number of 1s in the result.
- *Note that* the Hamming distance is a value greater than zero.

Techniques for Error Detection & Correction

- Let us find the Hamming distance between two pairs of words.

1. The Hamming distance d(000, 011) is 2 because

$$000 \oplus 011 \text{ is } 011 \text{ (two 1s)}$$

2. The Hamming distance d(10101, 11110) is 3 because

$$10101 \oplus 11110 \text{ is } 01011 \text{ (three 1s)}$$

Techniques for Error Detection & Correction

Minimum Hamming Distance

- The minimum Hamming distance is the smallest Hamming distance between all possible pairs.
- We use d_{min} to define the minimum Hamming distance in a coding scheme.
- To find this value, we find the Hamming distances between all words and select the smallest one.

Techniques for Error Detection & Correction

Example 3

- Find the minimum Hamming distance of the coding scheme in Table 1.

Datawords	*Codewords*
00	000
01	011
10	101
11	110

Solution

- We first find all Hamming distances.

$d(000, 011) = 2$ $d(000, 101) = 2$ $d(000, 110) = 2$ $d(011, 101) = 2$
$d(011, 110) = 2$ $d(101, 110) = 2$

- The d_{min} in this case is 2.

Techniques for Error Detection & Correction

Example 4

- Find the minimum Hamming distance of the coding scheme in Table 2.

Dataword	*Codeword*
00	00000
01	01011
10	10101
11	11110

Solution

- We first find all the Hamming distances.

$d(00000, 01011) = 3$ $d(00000, 10101) = 3$ $d(00000, 11110) = 4$
$d(01011, 10101) = 4$ $d(01011, 11110) = 3$ $d(10101, 11110) = 3$

- The d_{min} in this case is 3.

Techniques for Error Detection & Correction

Linear Block Codes

- The formal definition of linear block codes requires the knowledge of abstract algebra.
- A linear block code is a code in which the *exclusive* OR (addition modulo-2) of two valid codewords creates another valid codeword.

Techniques for Error Detection & Correction

Simple Parity-Check Code

- The most familiar error-detecting code is the simple parity-check code.
- In this code, a *k*-bit dataword is changed to an *n*-bit codeword where $n = k + 1$. The extra bit, called the **parity bit**, is selected to make the total number of 1s in the codeword even.
- The minimum Hamming distance for this category is $d_{min} = 2$, which means that the code is a single-bit error-detecting code; it cannot correct any error.

Techniques for Error Detection & Correction

- Table 3 is also a parity-check code with k =4 and n =5.

Table 3 Simple parity-check code C(5, 4)

Datawords	*Codewords*	*Datawords*	*Codewords*
0000	00000	1000	10001
0001	00011	1001	10010
0010	00101	1010	10100
0011	00110	1011	10111
0100	01001	1100	11000
0101	01010	1101	11011
0110	01100	1110	11101
0111	01111	1111	11110

Techniques for Error Detection & Correction

- Figure shows a possible structure of an encoder (at the sender) and a decoder (at the receiver).

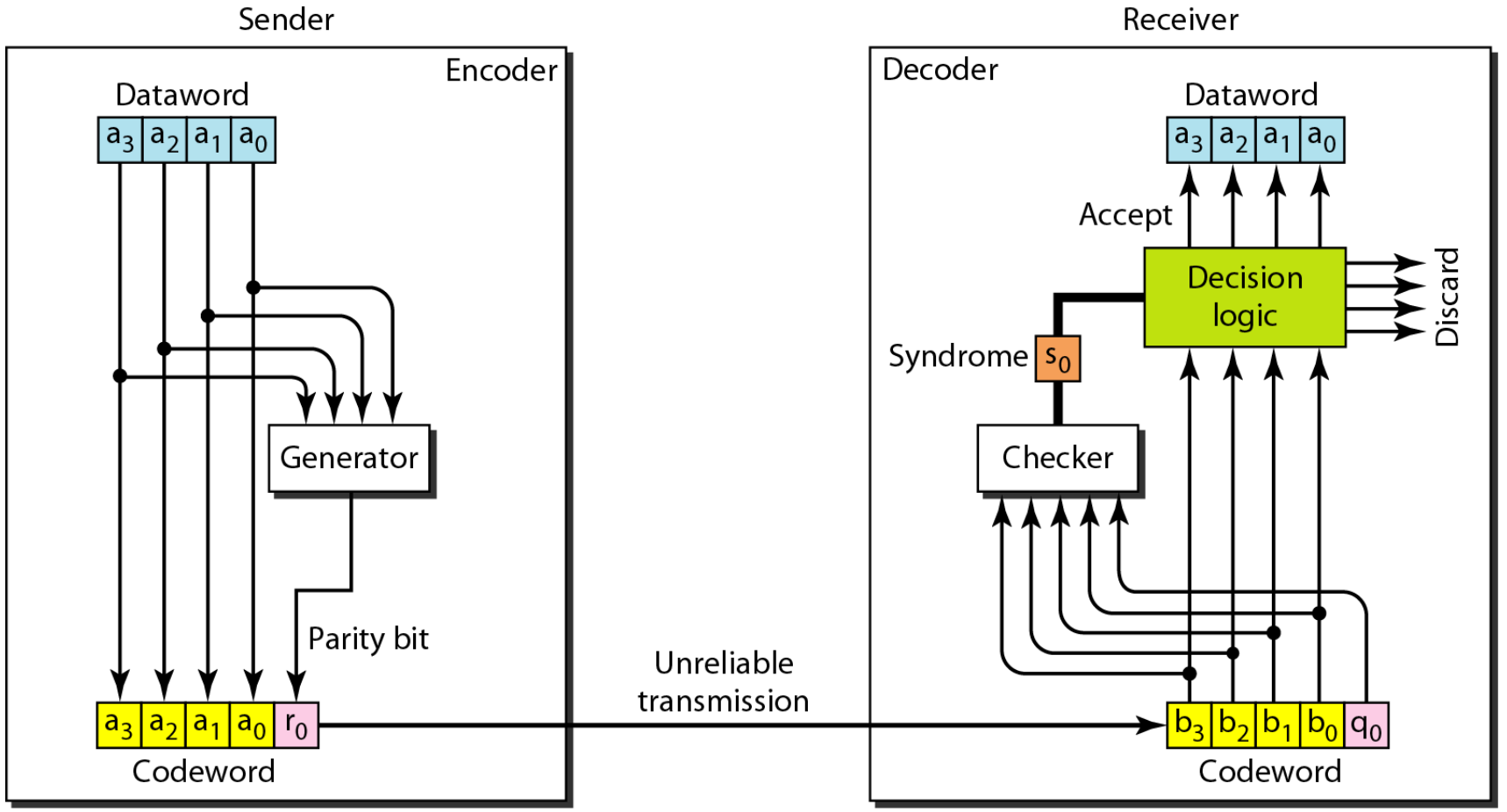


Figure Encoder and decoder for simple parity-check code

Techniques for Error Detection & Correction

- This is normally done by adding the 4 bits of the dataword (modulo-2); the result is the parity bit.
- In other words,

$$r_0 = a_0 + a_1 + a_2 + a_3$$

- If the number of 1s is even, the result is 0; if the number of 1s is odd, the result is 1. In both cases, the total number of 1s in the codeword is even.
- The sender sends the codeword which may be corrupted during transmission. The receiver receives a 5-bit word.

Techniques for Error Detection & Correction

- The checker at the receiver does the same thing as the generator in the sender with one exception: The addition is done over all 5 bits. The result, which is called the syndrome, is just 1 bit. The syndrome is 0 when the number of 1s in the received codeword is even; otherwise, it is 1.

$$s_0 = b_3 + b_2 + b_1 + b_0 + q_0$$

- The syndrome is passed to the decision logic analyzer.
- If the syndrome is 0, there is no error in the received codeword; the data portion of the received codeword is accepted as the dataword; if the syndrome is 1, the data portion of the received codeword is discarded. The dataword is not created.

Techniques for Error Detection & Correction

Example 5

- Let us look at some transmission scenarios. Assume the sender sends the dataword 1011. The codeword created from this dataword is 10111, which is sent to the receiver.
- We examine five cases:
  1. No error occurs; the received codeword is 10111. The syndrome is 0. The dataword 1011 is created.
  2. One single-bit error changes a_1 . The received codeword is 10011. The syndrome is 1. No dataword is created.
  3. One single-bit error changes r_0 . The received codeword is 10110. The syndrome is 1. No dataword is created.

Techniques for Error Detection & Correction

4. An error changes r_0 and a second error changes a_3. The received codeword is 00110. The syndrome is 0. The dataword 0011 is created at the receiver. Note that here the dataword is wrongly created due to the syndrome value.
5. Three bits - a_3, a_2, and a_1 are changed by errors. The received codeword is 01011. The syndrome is 1. The dataword is not created. This shows that the simple parity check, guaranteed to detect one single error, can also find any odd number of errors.

Techniques for Error Detection & Correction

Cyclic Codes

- Cyclic codes are special linear block codes with one extra property. In a cyclic code, if a codeword is cyclically shifted (rotated), the result is another codeword.
- For example, if 1011000 is a codeword and we cyclically left-shift, then 0110001 is also a codeword.
- In this case, if we call the bits in the first word a_0 to a_6 and the bits in the second word b_0 to b_6, we can shift the bits by using the following:

$$b_1 = a_0 \quad b_2 = a_1 \quad b_3 = a_2 \quad b_4 = a_3 \quad b_5 = a_4 \quad b_6 = a_5 \quad b_0 = a_6$$

- In the rightmost equation, the last bit of the first word is wrapped around and becomes the first bit of the second word.

Techniques for Error Detection & Correction

Cyclic Redundancy Check

- Cyclic codes to correct errors.
- A category of cyclic codes called the cyclic redundancy check (CRC) that is used in networks such as LANs and WANs.
- Table 4 shows an example of a CRC code. We can see both the linear and cyclic properties of this code.

Techniques for Error Detection & Correction

Table 4 A CRC code with C(7, 4)

Dataword	*Codeword*	*Dataword*	*Codeword*
0000	0000000	1000	1000101
0001	0001011	1001	1001110
0010	0010110	1010	1010011
0011	0011101	1011	1011000
0100	0100111	1100	1100010
0101	0101100	1101	1101001
0110	0110001	1110	1110100
0111	0111010	1111	1111111

Techniques for Error Detection & Correction

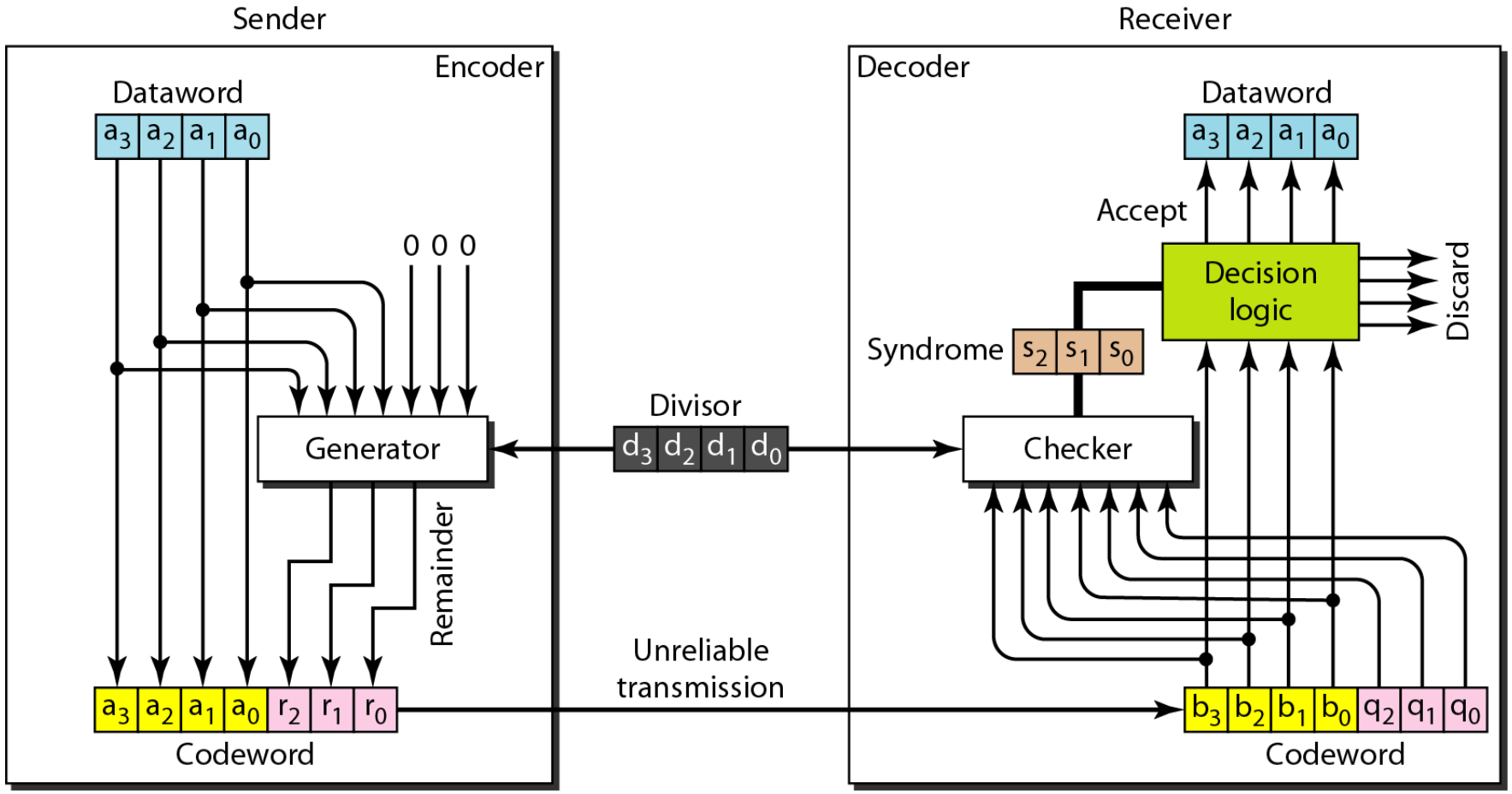


Figure CRC encoder and decoder

Techniques for Error Detection & Correction

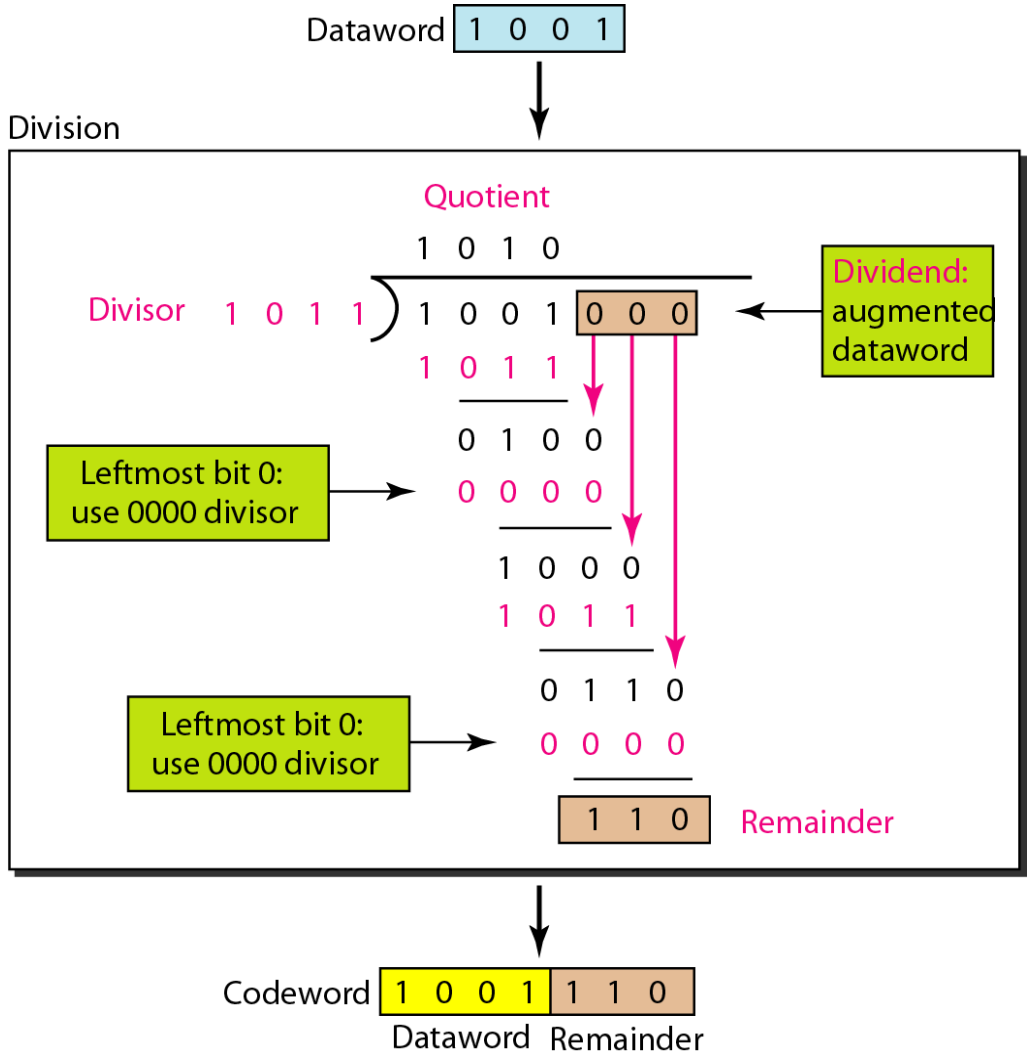


Figure Division in CRC encoder

Techniques for Error Detection & Correction

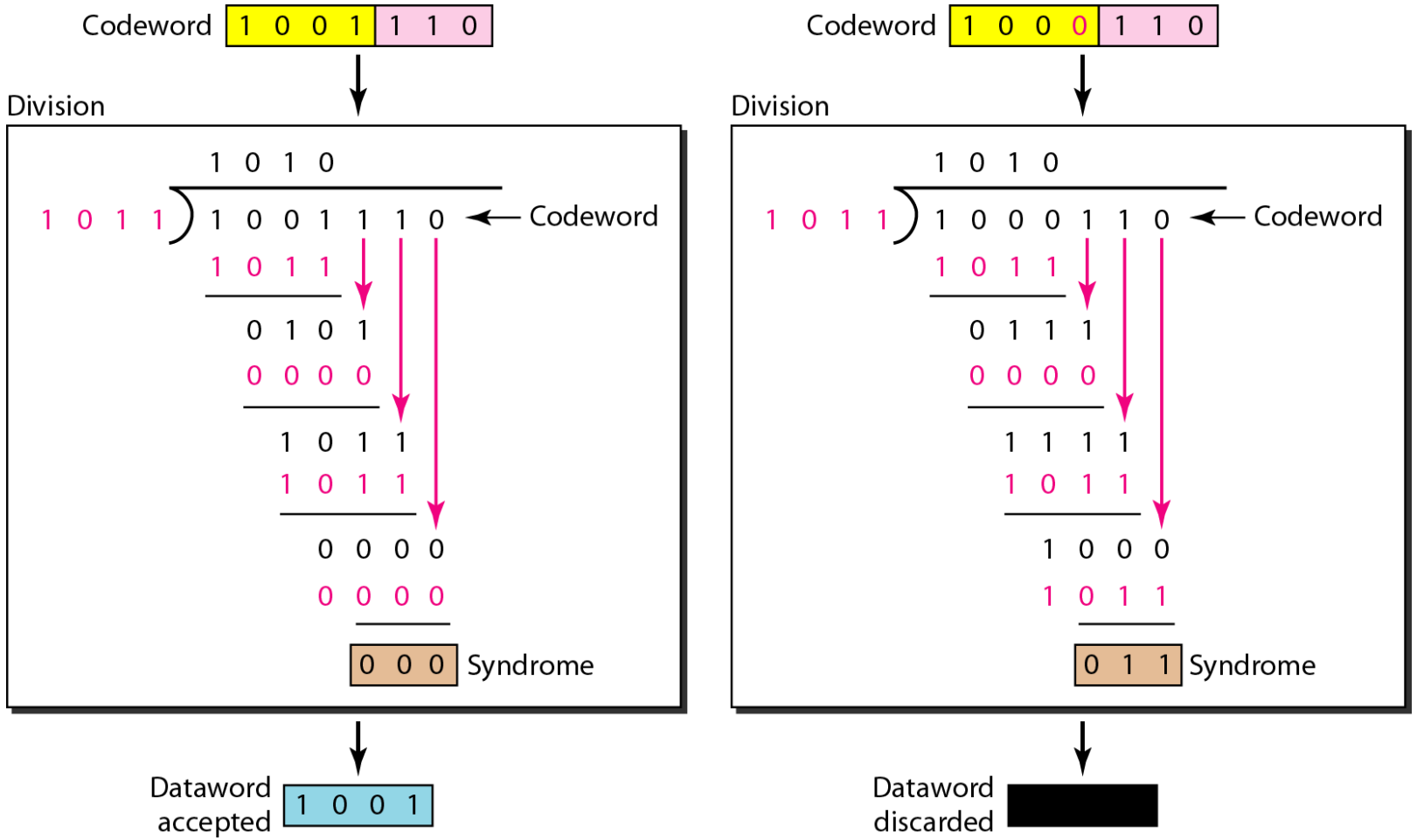


Figure **Division in the CRC decoder for two cases**

Techniques for Error Detection & Correction

Checksum

- This is the last error detection method, which is called the checksum.
- The checksum is used in the Internet by several protocols although not at the data link layer.
- Like linear and cyclic codes, the checksum is based on the concept of redundancy. Several protocols still use the checksum for error detection.

Techniques for Error Detection & Correction

Idea

Example 6

- Suppose our data is a list of five 4-bit numbers that we want to send to a destination. In addition to sending these numbers, we send the sum of the numbers.
- For example, if the set of numbers is (7, 11, 12, 0, 6), we send (7, 11, 12, 0, 6, 36), where 36 is the sum of the original numbers. The receiver adds the five numbers and compares the result with the sum.
- If the two are the same, the receiver assumes no error, accepts the five numbers, and discards the sum. Otherwise, there is an error somewhere and the data are not accepted.

Techniques for Error Detection & Correction

Example 7

- We can make the job of the receiver easier if we send the negative (complement) of the sum, called the checksum.
- In this case, we send (7, 11, 12, 0, 6, −36). The receiver can add all the numbers received (including the checksum). If the result is 0, it assumes no error; otherwise, there is an error.

Techniques for Error Detection & Correction

One's Complement

- One solution is to use one's complement arithmetic. In this arithmetic, we can represent unsigned numbers between 0 and $2^n - 1$ using only *n* bits.
- If the number has more than *n* bits, the extra leftmost bits need to be added to the *n* rightmost bits (wrapping).
- In one's complement arithmetic, a negative number can be represented by inverting all bits (changing a 0 to a 1 and a 1 to a 0). This is the same as subtracting the number from $2^n - 1$.

Techniques for Error Detection & Correction

Example 8

- How can we represent the number 21 in one's complement arithmetic using only four bits?

Solution

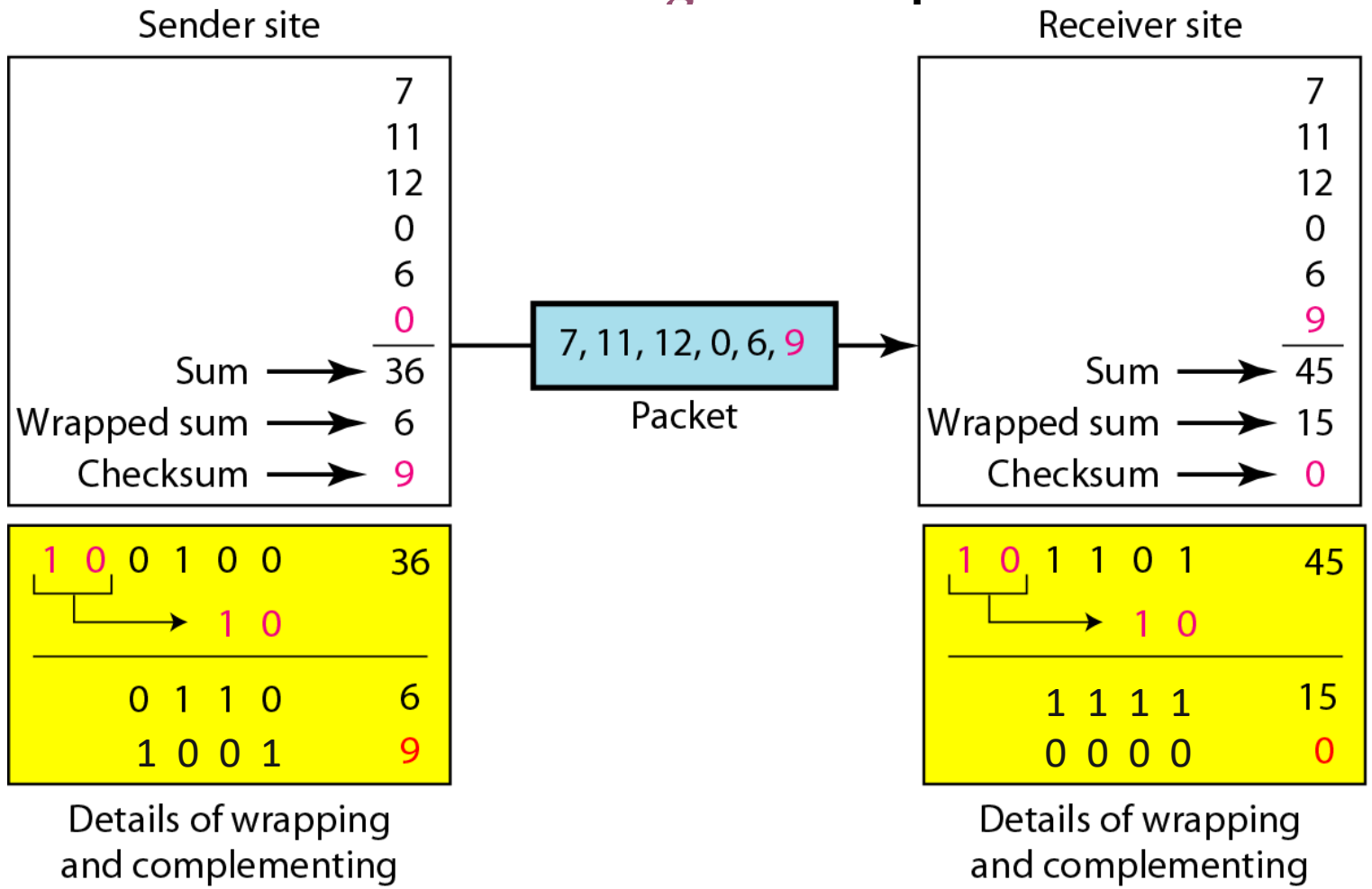
- The number 21 in binary is 10101 (it needs five bits). We can wrap the leftmost bit and add it to the four rightmost bits. We have (0101 + 1) = 0110 or 6.

Techniques for Error Detection & Correction

Example 9

- Figure shows the process at the sender and at the receiver.

Figure Example 9



Techniques for Error Detection & Correction

- The sender initializes the checksum to 0 and adds all data items and the checksum (the checksum is considered as one data item and is shown in color).
- The result is 36. However, 36 cannot be expressed in 4 bits. The extra two bits are wrapped and added with the sum to create the wrapped sum value 6.
- In the figure, we have shown the details in binary. The sum is then complemented, resulting in the checksum value 9 (15 − 6 = 9). The sender now sends six data items to the receiver including the checksum 9.

Techniques for Error Detection & Correction

- The receiver follows the same procedure as the sender. It adds all data items (including the checksum); the result is 45. The sum is wrapped and becomes 15. The wrapped sum is complemented and becomes 0.
- Since the value of the checksum is 0, this means that the data is not corrupted.
- The receiver drops the checksum and keeps the other data items. If the checksum is not zero, the entire packet is dropped.

Techniques for Error Detection & Correction

Internet Checksum

- Traditionally, the Internet has been using a 16-bit checksum. The sender calculates the checksum by following these steps.
- Sender site:
  1. The message is divided into 16-bit words.
  2. The value of the checksum word is set to 0.
  3. All words including the checksum are added using one's complement addition.
  4. The sum is complemented and becomes the checksum.
  5. The checksum is sent with the data

Techniques for Error Detection & Correction

- Receiver site:
  1. The message (including checksum) is divided into 16-bit words.
  2. All words are added using one's complement addition.
  3. The sum is complemented and becomes the new checksum.
  4. If the value of checksum is 0, the message is accepted; otherwise, it is rejected

Elementary Data Link Layer Protocols

- Data Link Layer Protocols:
  - Simplex
  - Stop and Wait
  - Sliding Window Protocol

Elementary Data Link Layer Protocols

Simplex(Simplest) Protocol

- First protocol, which we call the Simplest Protocol for lack of any other name, is one that has no flow or error control.
- It is a unidirectional protocol in which data frames are traveling in only one direction-from the sender to receiver.
- We assume that the receiver can immediately handle any frame it receives with a processing time that is small enough to be negligible.
- The data link layer of the receiver immediately removes the header from the frame and hands the data packet to its network layer, which can also accept the packet immediately.
- In other words, the receiver can never be overwhelmed with incoming frames.

Elementary Data Link Layer Protocols

Design

- There is no need for flow control in this scheme. The data link layer at the sender site gets data from its network layer, makes a frame out of the data, and sends it.
- The data link layer at the receiver site receives a frame from its physical layer, extracts data from the frame, and delivers the data to its network layer.
- The data link layers of the sender and receiver provide transmission services for their network layers.
- The data link layers use the services provided by their physical layers (such as signaling, multiplexing, and so on) for the physical transmission of bits.
- Figure shows a design.

Elementary Data Link Layer Protocols

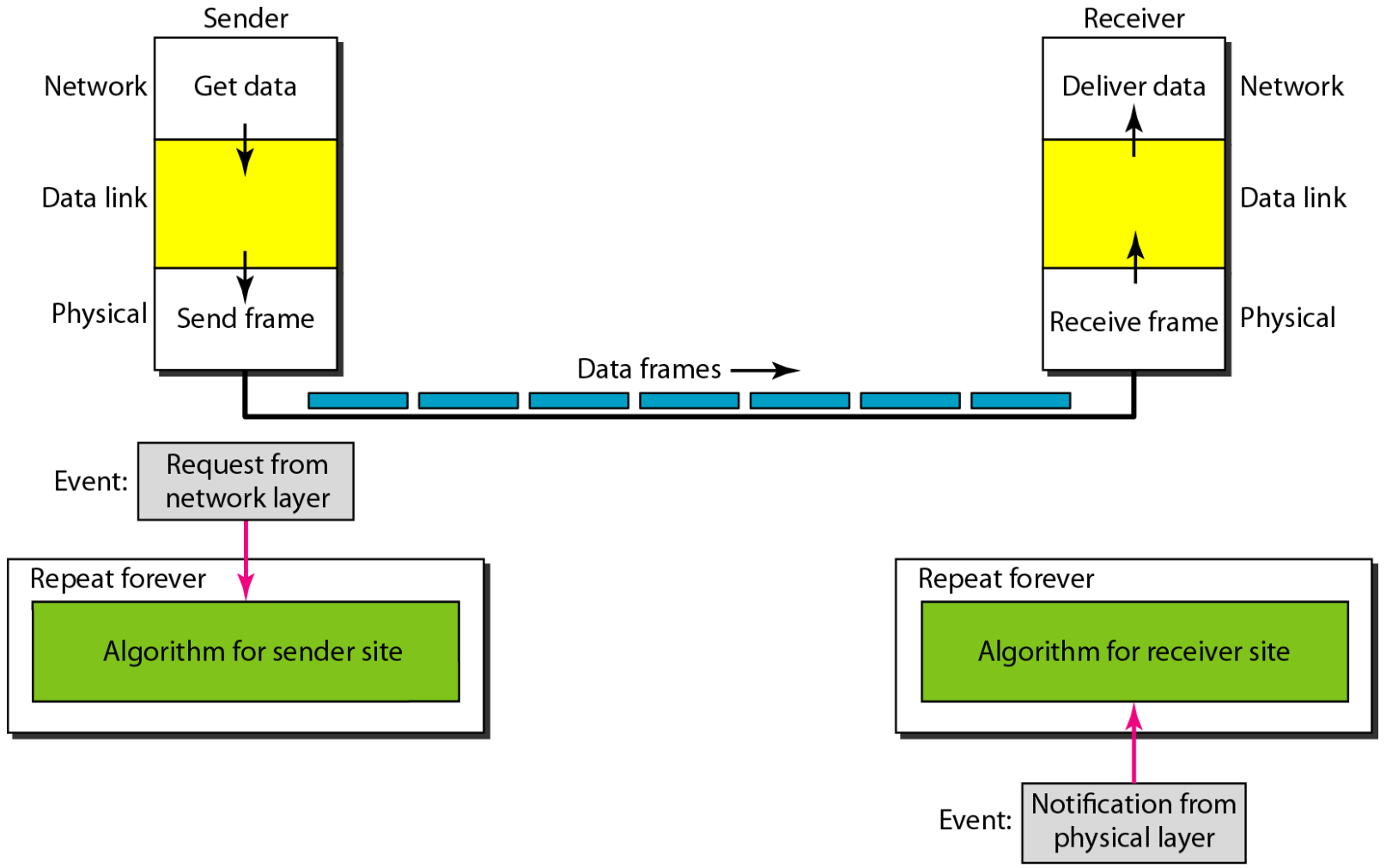


Figure The design of the simplest protocol with no flow or error control

Elementary Data Link Layer Protocols

Stop-and-wait Protocol

- If data frames arrive at the receiver site faster than they can be processed, the frames must be stored until their use.
- Normally, the receiver does not have enough storage space, especially if it is receiving data from many sources. This may result in either the discarding of frames or denial of service.
- To prevent the receiver from becoming overwhelmed with frames, we somehow need to tell the sender to slow down. There must be feedback from the receiver to the sender.

Elementary Data Link Layer Protocols

- The protocol we discuss now is called the Stop-and-Wait Protocol because the sender sends one frame, stops until it receives confirmation from the receiver (okay to go ahead), and then sends the next frame.
- We still have unidirectional communication for data frames, but auxiliary ACK frames (simple tokens of acknowledgment) travel from the other direction. We add flow control to our previous protocol.

Elementary Data Link Layer Protocols

Design

- Figure illustrates the mechanism. Comparing this figure with Simplest protocol, we can see the traffic on the forward channel (from sender to receiver) and the reverse channel.
- At any time, there is either one data frame on the forward channel or one ACK frame on the reverse channel. We therefore need a half-duplex link.

Elementary Data Link Layer Protocols

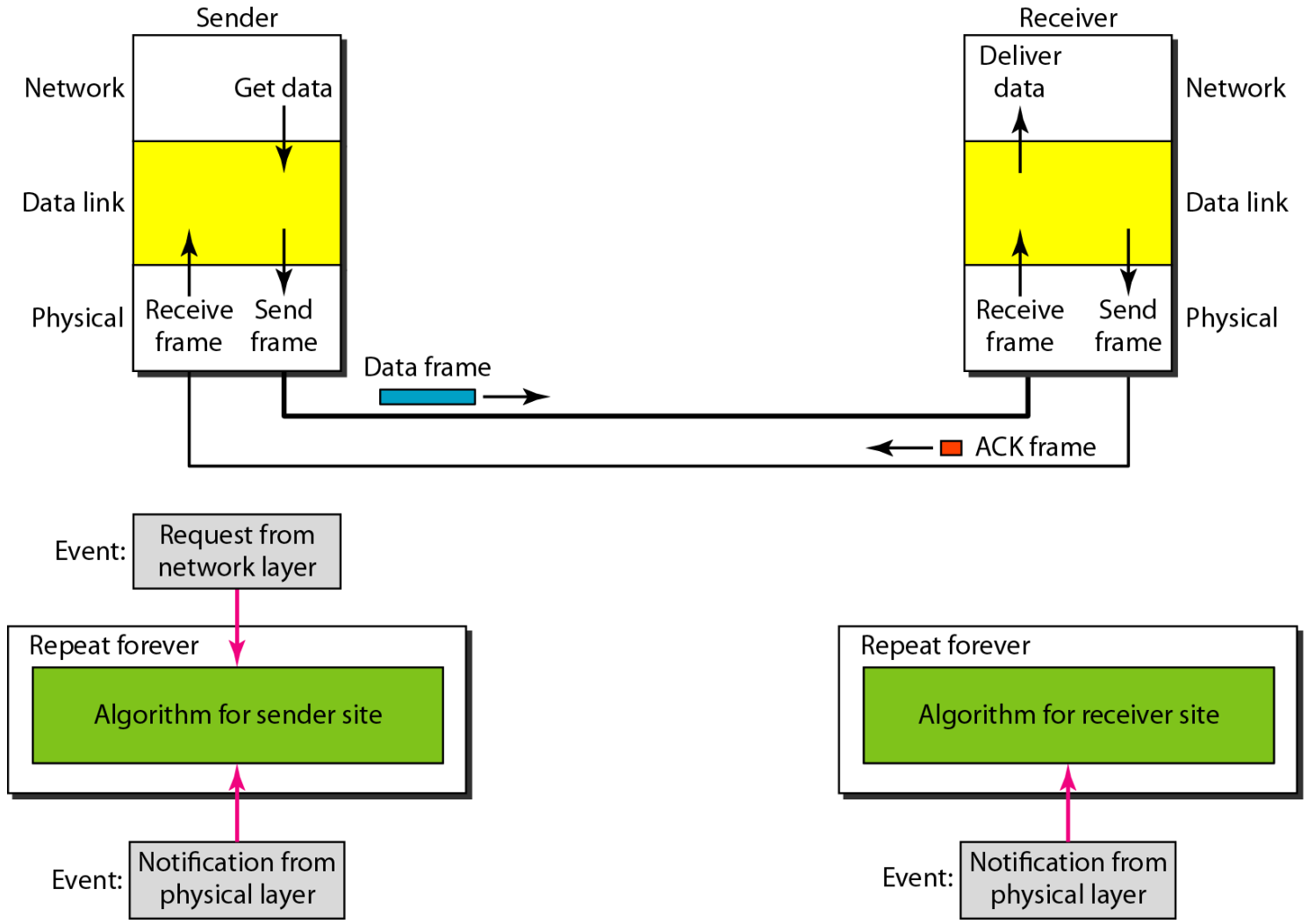


Figure Design of Stop-and-Wait Protocol

Elementary Data Link Layer Protocols

Sliding Window Protocol

- Sliding Window protocols are those protocols that are used as a method of flow control in networks for the transfer of data.
- With the help of the sliding window technique, multiple frames can be sent at a time by the sender before receiving any acknowledgment from the receiver.
- Sliding Window protocols make the use of **TCP**(transmission control protocol).
- The receiver can send the acknowledgment of **multiple frames** transmitted by the sender using a single **ACK frame**.
- In the Sliding Window protocols, the term **sliding window** mainly refers to the imaginary box that can hold the frames of both the sender side as well as receiver side.

Elementary Data Link Layer Protocols

- The receiver maintains a buffer to manage the flow of data packets. The receive buffer holds the packets that have been sent by the sender but have not yet been processed.
- During data transmission, the receiver notifies the sender of the amount of free space available in the receive buffer. This space is referred to as the receive window, which is the buffer size less the amount of unprocessed data.
- The sender cannot send more data packets than the amount of space available in the receive window.

Elementary Data Link Layer Protocols

- Data packets are numbered sequentially so they can be tracked when data is being transmitted from the sender to the receiver.
- During the transmission process, the data packets pass through one of four stages:
  1. Sent and acknowledged by the receiver.
  2. Sent but not acknowledged by the receiver.
  3. Not sent but the receiver is ready accept them.
  4. Not sent and the receiver is not ready to accept them.

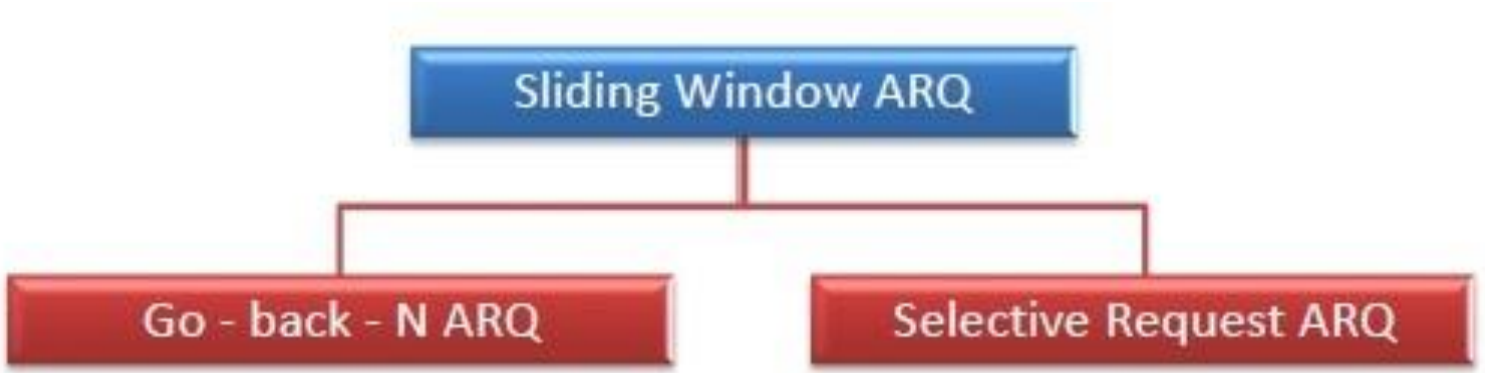
Elementary Data Link Layer Protocols

Elementary Data Link Layer Protocols

- The Sliding Window mainly provides the upper limit on the number of frames that can be transmitted before the requirement of an acknowledgment.
- The frames get acknowledged by the receiver at any point even when the window is not completely full on the receiver side.
- Also, the Frames may be transmitted by the source side even when at the time the window is not yet full on the sender side.
- There is the specific size of the window, where the frames are numbered *modulo-n*, which simply means frames are numbered from **0 to *n*-1**. For e.g. if ***n* = 10**, the frames are numbered 0, 1,2,3,4,5,6, 7,8,9, 0, 1,2,3,4,5,6, 7, 8,9,0, 1, ….

Elementary Data Link Layer Protocols

- Sliding window protocol has two types:
  - Go-Back-N ARQ
  - Selective Repeat ARQ



Elementary Data Link Layer Protocols

Go-Back-N ARQ

- Go-Back-N ARQ protocol is also known as Go-Back-N Automatic Repeat Request. It is a data link layer protocol that uses a sliding window method. In this, if any frame is corrupted or lost, all subsequent frames have to be sent again.
- The size of the sender window is N in this protocol. For example, Go-Back-8, the size of the sender window, will be 8. The receiver window size is always 1.
- If the receiver receives a corrupted frame, it cancels it. The receiver does not accept a corrupted frame. When the timer expires, the sender sends the correct frame again. The design of the Go-Back-N ARQ protocol is shown below.

Elementary Data Link Layer Protocols

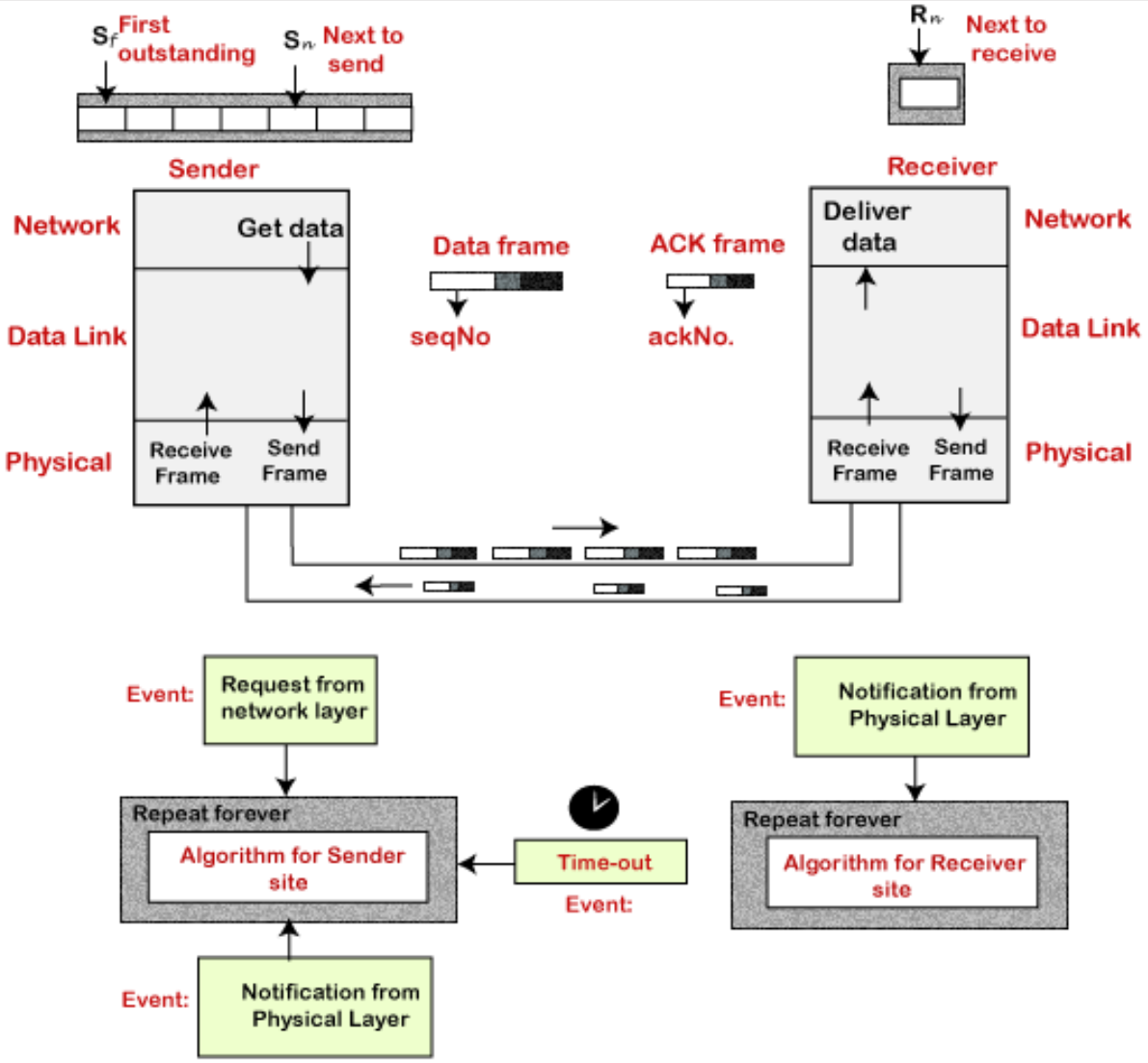


Figure Go-Back-N ARQ

Elementary Data Link Layer Protocols

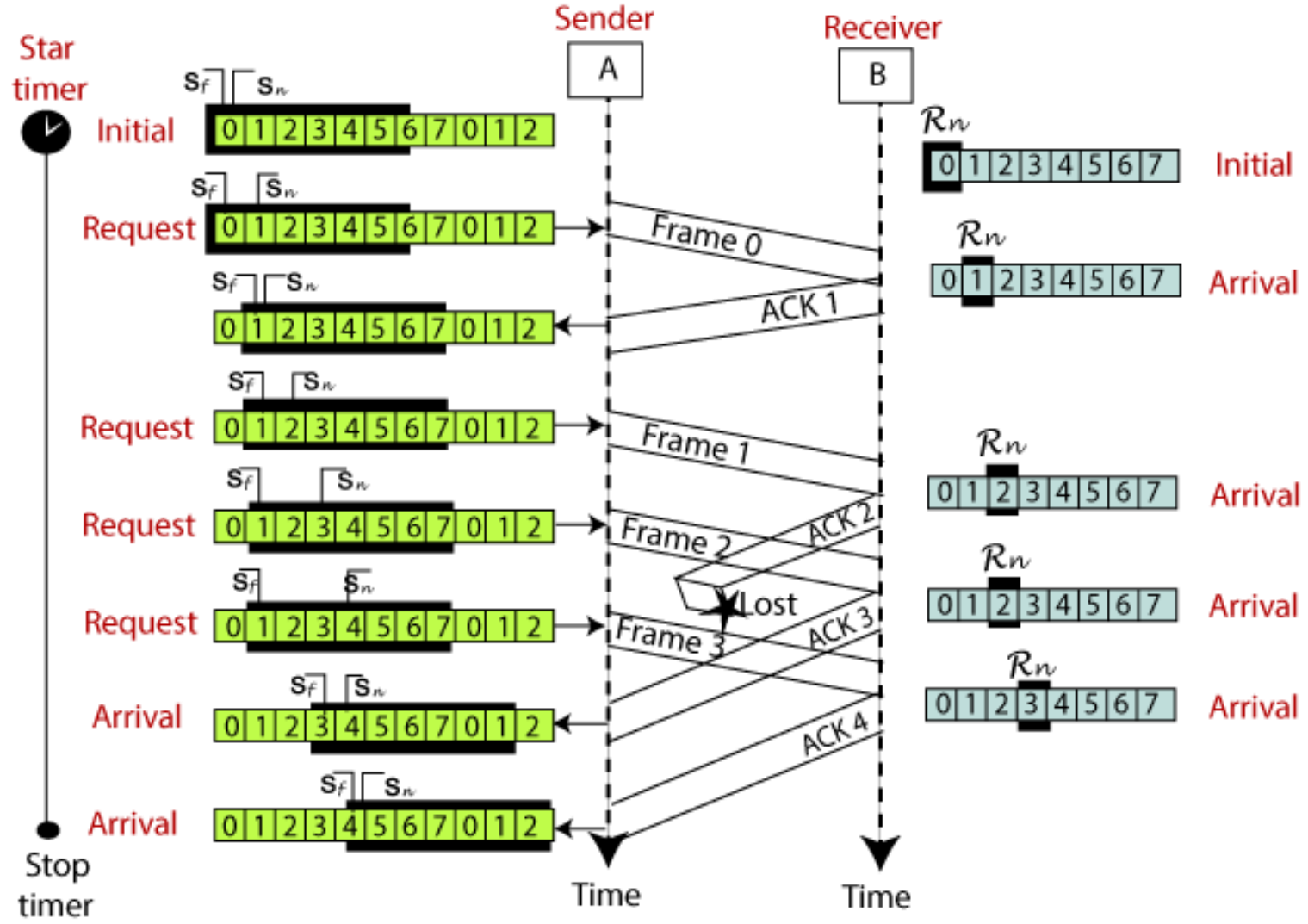


Figure Example of Go-Back-N ARQ

Elementary Data Link Layer Protocols

Selective Repeat ARQ

- Selective Repeat ARQ is also known as the Selective Repeat Automatic Repeat Request. It is a data link layer protocol that uses a sliding window method.
- The Go-back-N ARQ protocol works well if it has fewer errors. But if there is a lot of error in the frame, lots of bandwidth loss in sending the frames again. So, we use the Selective Repeat ARQ protocol.
- In this protocol, the size of the sender window is always equal to the size of the receiver window. The size of the sliding window is always greater than 1.
- If the receiver receives a corrupt frame, it does not directly discard it. It sends a negative acknowledgment to the sender. The sender sends that frame again as soon as on the receiving negative acknowledgment. There is no waiting for any time-out to send that frame. The design of the Selective Repeat ARQ protocol is shown below.

Elementary Data Link Layer Protocols

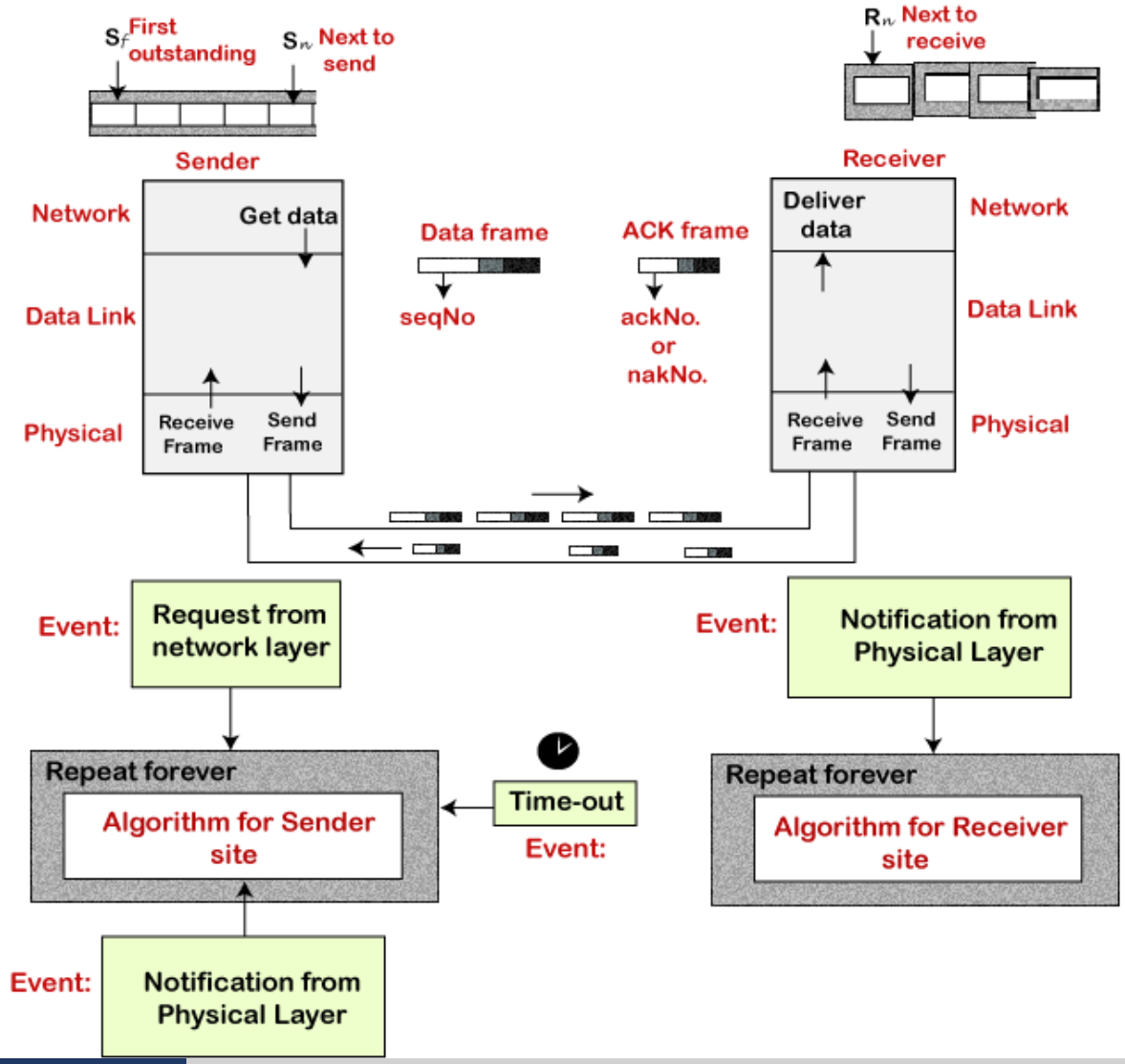
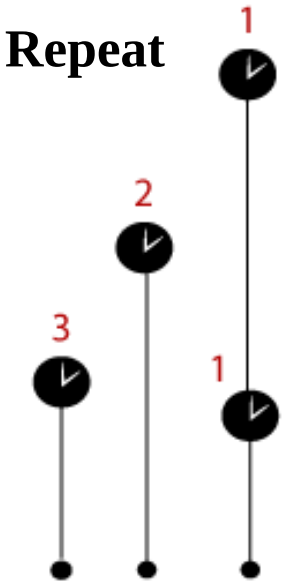


Figure Selective Repeat ARQ

Elementary Data Link Layer Protocols



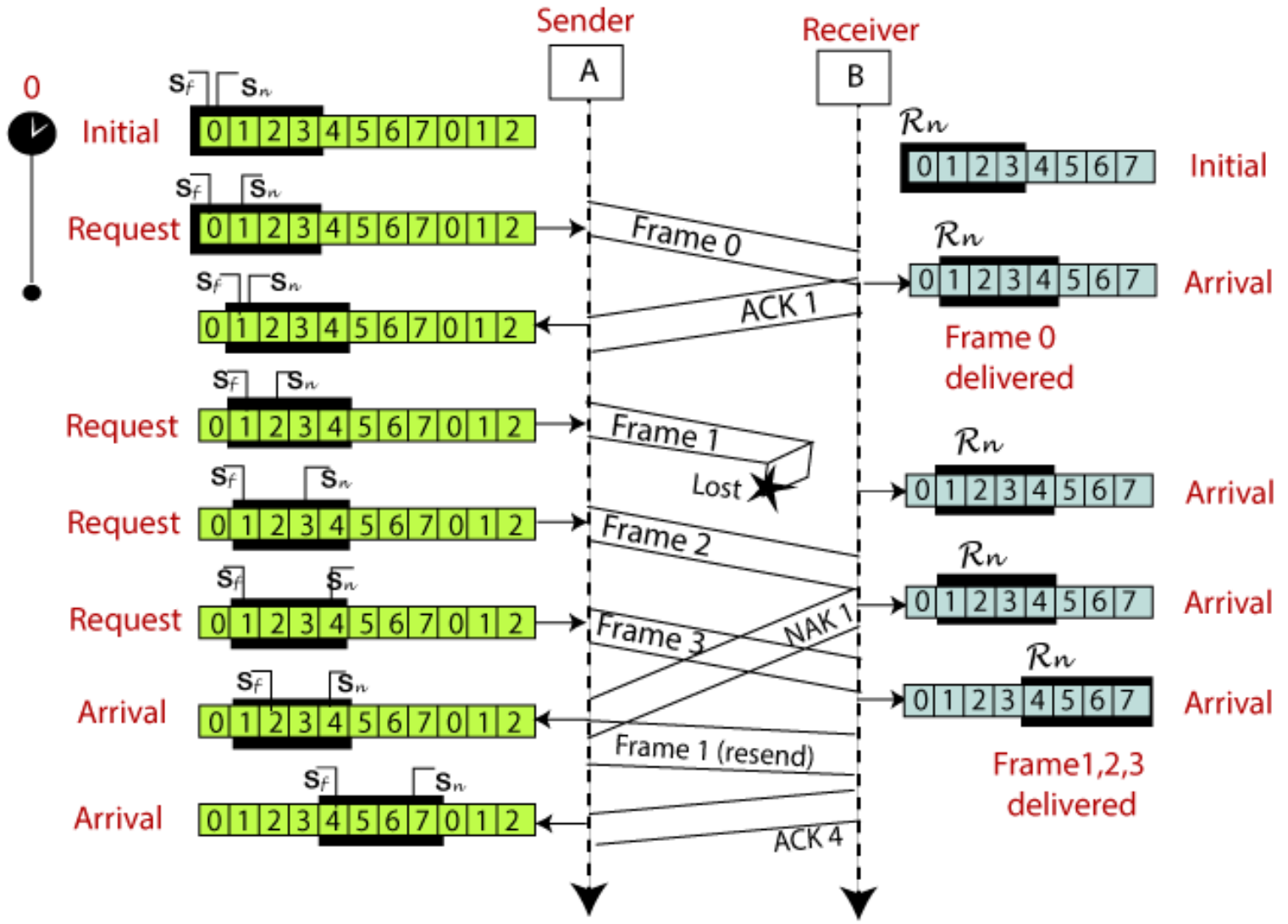


Figure Example of Selective Repeat ARQ

Difference between the Go-Back-N ARQ and Selective Repeat ARQ

Go-Back-N ARQ	Selective Repeat ARQ
If a frame is corrupted or lost in it,all subsequent frames have to be sent again.	In this, only the frame is sent again, which is corrupted or lost.
If it has a high error rate,it wastes a lot of bandwidth.	There is a loss of low bandwidth.
It is less complex.	It is more complex because it has to do sorting and searching as well. And it also requires more storage.
It does not require sorting.	In this, sorting is done to get the frames in the correct order.
It does not require searching.	The search operation is performed in it.
It is used more.	It is used less because it is more complex.

Summary

- Introduction and Design Issues,
- Flow and Error Control,
- Techniques for Error Detection and Correction,
- Elementary Data Link Layer Protocols: Simplex, Stop and Wait, Sliding Window Protocol

Thank You