FUTURE VISION BIE

One Stop for All Study Materials

& Lab Programs

Future Vision

By K B Hemanth Raj

Scan the QR Code to Visit the Web Page

Or

Visit : https://hemanthrajhemu.github.io

Gain Access to All Study Materials according to VTU,
CSE – Computer Science Engineering,
ISE – Information Science Engineering,
ECE - Electronics and Communication Engineering
& MORE...

Join Telegram to get Instant Updates: https://bit.ly/VTU_TELEGRAM

Contact: MAIL: futurevisionbie@gmail.com

INSTAGRAM: www.instagram.com/hemanthraj_hemu/

INSTAGRAM: www.instagram.com/futurevisionbie/

WHATSAPP SHARE: https://bit.ly/FVBIESHARE

CHAPTER 13

Wired LANs: Ethernet

After discussing the general issues related to the data-link layer in Chapters 9 to 12, it is time in this chapter to discuss the wired LANs. Although over a few decades many wired LAN protocols existed, only the Ethernet technology survives today. This is the reason that we discuss only this technology and its evolution in this chapter.

This chapter is divided into five sections.

- The first section discusses the Ethernet protocol in general. It explains that IEEE Project 802 defines the LLC and MAC sublayers for all LANs including Ethernet. The section also lists the four generations of Ethernet.
- The second section discusses the Standard Ethernet. Although this generation is rarely seen in practice, most of the characteristics have been inherited by the following three generations. The section first describes some characteristics of the Standard Ethernet. It then discusses the addressing mechanism, which is the same in all Ethernet generations. The section next discusses the access method, CSMA/CD, which we discussed in Chapter 12. The section then reviews the efficiency of the Standard Ethernet. It then shows the encoding and the implementation of this generation. Before closing the section, the changes in this generation that resulted in the move to the next generation are listed.
- The third section describes the Fast Ethernet, the second generation, which can still be seen in many places. The section first describes the changes in the MAC sublayer. The section then discusses the physical layer and the implementation of this generation.
- The fourth section discusses the Gigabit Ethernet, with the rate of 1 gigabit per second. The section first describes the MAC sublayer. It then moves to the physical layer and implementation.
- The fifth section touches on the 10 Gigabit Ethernet. This is a new technology that can be used both for a backbone LAN or as a MAN (metropolitan area network). The section briefly describes the rationale and the implementation.

13.1 ETHERNET PROTOCOL

In Chapter 1, we mentioned that the TCP/IP protocol suite does not define any protocol for the data-link or the physical layer. In other words, TCP/IP accepts any protocol at these two layers that can provide services to the network layer. The data-link layer and the physical layer are actually the territory of the local and wide area networks. This means that when we discuss these two layers, we are talking about networks that are using them. As we see in this and the following two chapters, we can have wired or wireless networks. We discuss wired networks in this chapter and the next and postpone the discussion of wireless networks to Chapter 15.

In Chapter 1, we learned that a local area network (LAN) is a computer network that is designed for a limited geographic area such as a building or a campus. Although a LAN can be used as an isolated network to connect computers in an organization for the sole purpose of sharing resources, most LANs today are also linked to a wide area network (WAN) or the Internet.

In the 1980s and 1990s several different types of LANs were used. All of these LANs used a media-access method to solve the problem of sharing the media. The Ethernet used the CSMA/CD approach. The Token Ring, Token Bus, and FDDI (Fiber Distribution Data Interface) used the token-passing approach. During this period, another LAN technology, ATM LAN, which deployed the high speed WAN technology (ATM), appeared in the market.

Almost every LAN except Ethernet has disappeared from the marketplace because Ethernet was able to update itself to meet the needs of the time. Several reasons for this success have been mentioned in the literature, but we believe that the Ethernet protocol was designed so that it could evolve with the demand for higher transmission rates. It is natural that an organization that has used an Ethernet LAN in the past and now needs a higher data rate would update to the new generation instead of switching to another technology, which might cost more. This means that we confine our discussion of wired LANs to the discussion of Ethernet.

13.1.1 IEEE Project 802

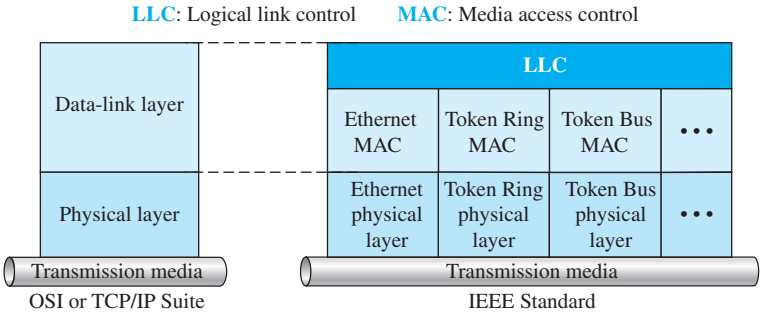
Before we discuss the Ethernet protocol and all its generations, we need to briefly discuss the IEEE standard that we often encounter in text or real life. In 1985, the Computer Society of the IEEE started a project, called ***Project 802,*** to set standards to enable intercommunication among equipment from a variety of manufacturers. Project 802 does not seek to replace any part of the OSI model or TCP/IP protocol suite. Instead, it is a way of specifying functions of the physical layer and the data-link layer of major LAN protocols.

The relationship of the 802 Standard to the TCP/IP protocol suite is shown in Figure 13.1. The IEEE has subdivided the data-link layer into two sublayers: **logical link control (LLC)** and **media access control (MAC).** IEEE has also created several physical-layer standards for different LAN protocols.

Logical Link Control (LLC)

Earlier we discussed *data link control*. We said that data link control handles framing, flow control, and error control. In IEEE Project 802, flow control, error control, and

Figure 13.1 *IEEE standard for LANs*



part of the framing duties are collected into one sublayer called the *logical link control* (LLC). Framing is handled in both the LLC sublayer and the MAC sublayer.

The LLC provides a single link-layer control protocol for all IEEE LANs. This means LLC protocol can provide interconnectivity between different LANs because it makes the MAC sublayer transparent.

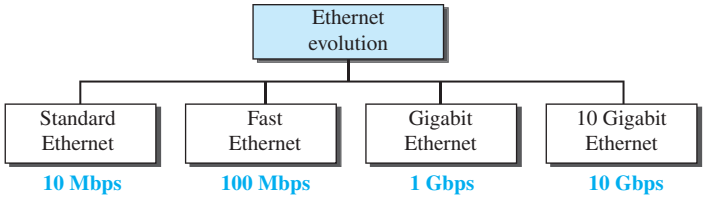
Media Access Control (MAC)

Earlier we discussed multiple access methods including random access, controlled access, and channelization. IEEE Project 802 has created a sublayer called ***media access control*** that defines the specific access method for each LAN. For example, it defines CSMA/CD as the media access method for Ethernet LANs and defines the token-passing method for Token Ring and Token Bus LANs. As we mentioned in the previous section, part of the framing function is also handled by the MAC layer.

13.1.2 Ethernet Evolution

The Ethernet LAN was developed in the 1970s by Robert Metcalfe and David Boggs. Since then, it has gone through four generations: **Standard Ethernet** (10 Mbps), **Fast Ethernet** (100 Mbps), **Gigabit Ethernet** (1 Gbps), and **10 Gigabit Ethernet** (10 Gbps), as shown in Figure 13.2. We briefly discuss all these generations.

Figure 13.2 *Ethernet evolution through four generations*



13.2 STANDARD ETHERNET

We refer to the original Ethernet technology with the data rate of 10 Mbps as the *Standard Ethernet*. Although most implementations have moved to other technologies in the Ethernet evolution, there are some features of the Standard Ethernet that have not changed during the evolution. We discuss this standard version to pave the way for understanding the other three technologies.

13.2.1 Characteristics

Let us first discuss some characteristics of the Standard Ethernet.

Connectionless and Unreliable Service

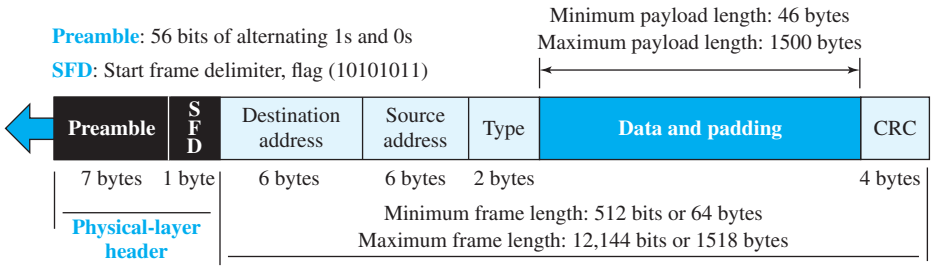
Ethernet provides a connectionless service, which means each frame sent is independent of the previous or next frame. Ethernet has no connection establishment or connection termination phases. The sender sends a frame whenever it has it; the receiver may or may not be ready for it. The sender may overwhelm the receiver with frames, which may result in dropping frames. If a frame drops, the sender will not know about it. Since IP, which is using the service of Ethernet, is also connectionless, it will not know about it either. If the transport layer is also a connectionless protocol, such as UDP, the frame is lost and salvation may only come from the application layer. However, if the transport layer is TCP, the sender TCP does not receive acknowledgment for its segment and sends it again.

Ethernet is also unreliable like IP and UDP. If a frame is corrupted during transmission and the receiver finds out about the corruption, which has a high level of probability of happening because of the CRC-32, the receiver drops the frame silently. It is the duty of high-level protocols to find out about it.

Frame Format

The Ethernet frame contains seven fields, as shown in Figure 13.3.

Figure 13.3 *Ethernet frame*



- ***Preamble.*** This field contains 7 bytes (56 bits) of alternating 0s and 1s that alert the receiving system to the coming frame and enable it to synchronize its clock if it's out of synchronization. The pattern provides only an alert and a timing pulse. The 56-bit

pattern allows the stations to miss some bits at the beginning of the frame. The *preamble* is actually added at the physical layer and is not (formally) part of the frame.

- ***Start frame delimiter (SFD).*** This field (1 byte: 10101011) signals the beginning of the frame. The SFD warns the station or stations that this is the last chance for synchronization. The last 2 bits are $(11)_2$ and alert the receiver that the next field is the destination address. This field is actually a flag that defines the beginning of the frame. We need to remember that an Ethernet frame is a variable-length frame. It needs a flag to define the beginning of the frame. The SFD field is also added at the physical layer.
- ***Destination address (DA).*** This field is six bytes (48 bits) and contains the link-layer address of the destination station or stations to receive the packet. We will discuss addressing shortly. When the receiver sees its own link-layer address, or a multicast address for a group that the receiver is a member of, or a broadcast address, it decapsulates the data from the frame and passes the data to the upper-layer protocol defined by the value of the type field.
- ***Source address (SA).*** This field is also six bytes and contains the link-layer address of the sender of the packet. We will discuss addressing shortly.
- ***Type.*** This field defines the upper-layer protocol whose packet is encapsulated in the frame. This protocol can be IP, ARP, OSPF, and so on. In other words, it serves the same purpose as the protocol field in a datagram and the port number in a segment or user datagram. It is used for multiplexing and demultiplexing.
- ***Data.*** This field carries data encapsulated from the upper-layer protocols. It is a minimum of 46 and a maximum of 1500 bytes. We discuss the reason for these minimum and maximum values shortly. If the data coming from the upper layer is more than 1500 bytes, it should be fragmented and encapsulated in more than one frame. If it is less than 46 bytes, it needs to be padded with extra 0s. A padded data frame is delivered to the upper-layer protocol as it is (without removing the padding), which means that it is the responsibility of the upper layer to remove or, in the case of the sender, to add the padding. The upper-layer protocol needs to know the length of its data. For example, a datagram has a field that defines the length of the data.
- ***CRC.*** The last field contains error detection information, in this case a CRC-32. The CRC is calculated over the addresses, types, and data field. If the receiver calculates the CRC and finds that it is not zero (corruption in transmission), it discards the frame.

Frame Length

Ethernet has imposed restrictions on both the minimum and maximum lengths of a frame. The minimum length restriction is required for the correct operation of CSMA/CD, as we will see shortly. An Ethernet frame needs to have a minimum length of 512 bits or 64 bytes. Part of this length is the header and the trailer. If we count 18 bytes of header and trailer (6 bytes of source address, 6 bytes of destination address, 2 bytes of length or type, and 4 bytes of CRC), then the minimum length of data from the upper layer is $64 - 18 = 46$ bytes. If the upper-layer packet is less than 46 bytes, padding is added to make up the difference.

The standard defines the maximum length of a frame (without preamble and SFD field) as 1518 bytes. If we subtract the 18 bytes of header and trailer, the maximum length of the payload is 1500 bytes. The maximum length restriction has two historical reasons. First, memory was very expensive when Ethernet was designed; a maximum length restriction helped to reduce the size of the buffer. Second, the maximum length restriction prevents one station from monopolizing the shared medium, blocking other stations that have data to send.

Minimum frame length: 64 bytes	**Minimum data length: 46 bytes**
Maximum frame length: 1518 bytes	**Maximum data length: 1500 bytes**

13.2.2 Addressing

Each station on an Ethernet network (such as a PC, workstation, or printer) has its own **network interface card (NIC).** The NIC fits inside the station and provides the station with a link-layer address. The Ethernet address is 6 bytes (48 bits), normally written in **hexadecimal notation,** with a colon between the bytes. For example, the following shows an Ethernet MAC address:

4A:30:10:21:10:1A

Transmission of Address Bits

The way the addresses are sent out online is different from the way they are written in hexadecimal notation. The transmission is left to right, byte by byte; however, for each byte, the least significant bit is sent first and the most significant bit is sent last. This means that the bit that defines an address as unicast or multicast arrives first at the receiver. This helps the receiver to immediately know if the packet is unicast or multicast.

Example 13.1

Show how the address 47:20:1B:2E:08:EE is sent out online.

Solution

The address is sent left to right, byte by byte; for each byte, it is sent right to left, bit by bit, as shown below:

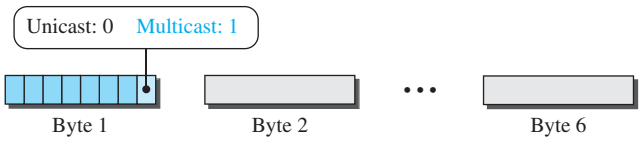
Hexadecimal	**47**	**20**	**1B**	**2E**	**08**	**EE**
Binary	01000111	00100000	00011011	00101110	00001000	11101110
Transmitted ←	11100010	00000100	11011000	01110100	00010000	01110111

Unicast, Multicast, and Broadcast Addresses

A source address is always a *unicast address*—the frame comes from only one station. The destination address, however, can be *unicast*, *multicast*, or *broadcast.* Figure 13.4 shows how to distinguish a unicast address from a multicast address. If the least significant bit of the first byte in a destination address is 0, the address is unicast; otherwise, it is multicast.

Note that with the way the bits are transmitted, the unicast/multicast bit is the first bit which is transmitted or received. The broadcast address is a special case of the

Figure 13.4 *Unicast and multicast addresses*



multicast address: the recipients are all the stations on the LAN. A broadcast destination address is forty-eight 1s.

Example 13.2

Define the type of the following destination addresses:

a. **4A:30:10:21:10:1A**
b. **47:20:1B:2E:08:EE**
c. **FF:FF:FF:FF:FF:FF**

Solution

To find the type of the address, we need to look at the second hexadecimal digit from the left. If it is even, the address is unicast. If it is odd, the address is multicast. If all digits are Fs, the address is broadcast. Therefore, we have the following:

a. This is a unicast address because A in binary is 1010 (even).
b. This is a multicast address because 7 in binary is 0111 (odd).
c. This is a broadcast address because all digits are Fs in hexadecimal.

Distinguish Between Unicast, Multicast, and Broadcast Transmission

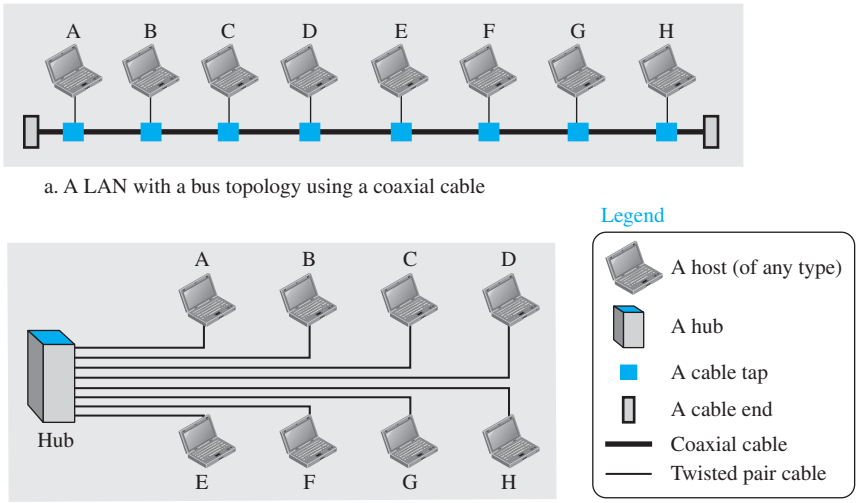
Standard Ethernet uses a coaxial cable (bus topology) or a set of twisted-pair cables with a hub (star topology) as shown in Figure 13.5.

We need to know that transmission in the standard Ethernet is always broadcast, no matter if the intention is unicast, multicast, or broadcast. In the bus topology, when station A sends a frame to station B, all stations will receive it. In the star topology, when station A sends a frame to station B, the hub will receive it. Since the hub is a passive element, it does not check the destination address of the frame; it regenerates the bits (if they have been weakened) and sends them to all stations except station A. In fact, it floods the network with the frame.

The question is, then, how the actual unicast, multicast, and broadcast transmissions are distinguished from each other. The answer is in the way the frames are kept or dropped.

- In a unicast transmission, all stations will receive the frame, the intended recipient keeps and handles the frame; the rest discard it.
- In a multicast transmission, all stations will receive the frame, the stations that are members of the group keep and handle it; the rest discard it.

Figure 13.5 *Implementation of standard Ethernet*



a. A LAN with a bus topology using a coaxial cable

b. A LAN with a star topology using a hub

- In a broadcast transmission, all stations (except the sender) will receive the frame and all stations (except the sender) keep and handle it.

13.2.3 Access Method

Since the network that uses the standard Ethernet protocol is a broadcast network, we need to use an access method to control access to the sharing medium. The standard Ethernet chose CSMA/CD with 1-persistent method, discussed earlier in Chapter 12, Section 1.3. Let us use a scenario to see how this method works for the Ethernet protocol.

- Assume station A in Figure 13.5 has a frame to send to station D. Station A first should check whether any other station is sending (carrier sense). Station A measures the level of energy on the medium (for a short period of time, normally less than 100 μs). If there is no signal energy on the medium, it means that no station is sending (or the signal has not reached station A). Station A interprets this situation as idle medium. It starts sending its frame. On the other hand, if the signal energy level is not zero, it means that the medium is being used by another station. Station A continuously monitors the medium until it becomes idle for 100 μs. It then starts sending the frame. However, station A needs to keep a copy of the frame in its buffer until it is sure that there is no collision. When station A is sure of this is the subject we discuss next.
- The medium sensing does not stop after station A has started sending the frame. Station A needs to send and receive continuously. Two cases may occur:

a. Station A has sent 512 bits and no collision is sensed (the energy level did not go above the regular energy level), the station then is sure that the frame will go through and stops sensing the medium. Where does the number 512 bits come from? If we consider the transmission rate of the Ethernet as 10 Mbps, this means that it takes the station 512/(10 Mbps) = 51.2 μs to send out 512 bits. With the speed of propagation in a cable (2×10^8 meters), the first bit could have gone 10,240 meters (one way) or only 5120 meters (round trip), have collided with a bit from the last station on the cable, and have gone back. In other words, if a collision were to occur, it should occur by the time the sender has sent out 512 bits (worst case) and the first bit has made a round trip of 5120 meters. We should know that if the collision happens in the middle of the cable, not at the end, station A hears the collision earlier and aborts the transmission. We also need to mention another issue. The above assumption is that the length of the cable is 5120 meters. The designer of the standard Ethernet actually put a restriction of 2500 meters because we need to consider the delays encountered throughout the journey. It means that they considered the worst case. The whole idea is that if station A does not sense the collision before sending 512 bits, there must have been no collision, because during this time, the first bit has reached the end of the line and all other stations know that a station is sending and refrain from sending. In other words, the problem occurs when another station (for example, the last station) starts sending before the first bit of station A has reached it. The other station mistakenly thinks that the line is free because the first bit has not yet reached it. The reader should notice that the restriction of 512 bits actually helps the sending station: The sending station is certain that no collision will occur if it is not heard during the first 512 bits, so it can discard the copy of the frame in its buffer.

b. Station A has sensed a collision before sending 512 bits. This means that one of the previous bits has collided with a bit sent by another station. In this case both stations should refrain from sending and keep the frame in their buffer for resending when the line becomes available. However, to inform other stations that there is a collision in the network, the station sends a 48-bit jam signal. The jam signal is to create enough signal (even if the collision happens after a few bits) to alert other stations about the collision. After sending the jam signal, the stations need to increment the value of K (number of attempts). If after increment $K = 15$, the experience has shown that the network is too busy, the station needs to abort its effort and try again. If $K < 15$, the station can wait a backoff time (T_B in Figure 12.13) and restart the process. As Figure 12.13 shows, the station creates a random number between 0 and $2^K - 1$, which means each time the collision occurs, the range of the random number increases exponentially. After the first collision ($K = 1$) the random number is in the range (0, 1). After the second collision ($K = 2$) it is in the range (0, 1, 2, 3). After the third collision ($K = 3$) it is in the range (0, 1, 2, 3, 4, 5, 6, 7). So after each collision, the probability increases that the backoff time becomes longer. This is due to the fact that if the collision happens even after the third or fourth attempt, it means that the network is really busy; a longer backoff time is needed.

13.2.4 Efficiency of Standard Ethernet

The efficiency of the Ethernet is defined as the ratio of the time used by a station to send data to the time the medium is occupied by this station. The practical efficiency of standard Ethernet has been measured to be

$$\textbf{Efficiency} = 1 / (1 + 6.4 \times a)$$

in which the parameter "a" is the number of frames that can fit on the medium. It can be calculated as a = (propagation delay)/(transmission delay) because the transmission delay is the time it takes a frame of average size to be sent out and the propagation delay is the time it takes to reach the end of the medium. Note that as the value of parameter a decreases, the efficiency increases. This means that if the length of the media is shorter or the frame size longer, the efficiency increases. In the ideal case, $a = 0$ and the efficiency is 1. We ask to calculate this efficiency in problems at the end of the chapter.

Example 13.3

In the Standard Ethernet with the transmission rate of 10 Mbps, we assume that the length of the medium is 2500 m and the size of the frame is 512 bits. The propagation speed of a signal in a cable is normally 2×10^8 m/s.

Propagation delay = $2500/(2 \times 10^8) = 12.5$ µs	**Transmission delay = $512/(10^7) = 51.2$ µs**
$a = 12.5/51.2 = 0.24$	**Efficiency = 39%**

The example shows that $a = 0.24$, which means only 0.24 of a frame occupies the whole medium in this case. The efficiency is 39 percent, which is considered moderate; it means that only 61 percent of the time the medium is occupied but not used by a station.

13.2.5 Implementation

The Standard Ethernet defined several implementations, but only four of them became popular during the 1980s. Table 13.1 shows a summary of Standard Ethernet implementations.

Table 13.1 *Summary of Standard Ethernet implementations*

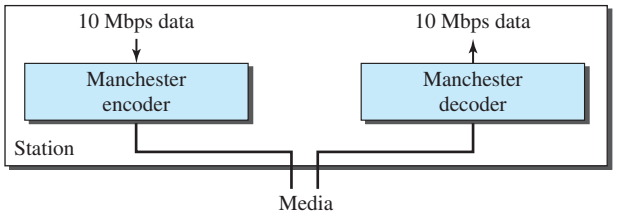
Implementation	*Medium*	*Medium Length*	*Encoding*
10Base5	Thick coax	500 m	Manchester
10Base2	Thin coax	185 m	Manchester
10Base-T	2 UTP	100 m	Manchester
10Base-F	2 Fiber	2000 m	Manchester

In the nomenclature 10BaseX, the number defines the data rate (10 Mbps), the term *Base* means baseband (digital) signal, and X approximately defines either the maximum size of the cable in 100 meters (for example 5 for 500 or 2 for 185 meters) or the type of cable, T for unshielded twisted pair cable (UTP) and F for fiber-optic. The standard Ethernet uses a baseband signal, which means that the bits are changed to a digital signal and directly sent on the line.

Encoding and Decoding

All standard implementations use digital signaling (baseband) at 10 Mbps. At the sender, data are converted to a digital signal using the Manchester scheme; at the receiver, the received signal is interpreted as Manchester and decoded into data. As we saw in Chapter 4, Manchester encoding is self-synchronous, providing a transition at each bit interval. Figure 13.6 shows the encoding scheme for Standard Ethernet.

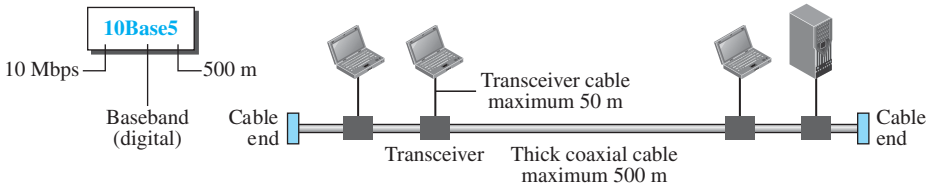
Figure 13.6 *Encoding in a Standard Ethernet implementation*



10Base5: Thick Ethernet

The first implementation is called ***10Base5, thick Ethernet,*** or ***Thicknet.*** The nickname derives from the size of the cable, which is roughly the size of a garden hose and too stiff to bend with your hands. 10Base5 was the first Ethernet specification to use a bus topology with an external **transceiver** (transmitter/receiver) connected via a tap to a thick coaxial cable. Figure 13.7 shows a schematic diagram of a 10Base5 implementation.

Figure 13.7 *10Base5 implementation*



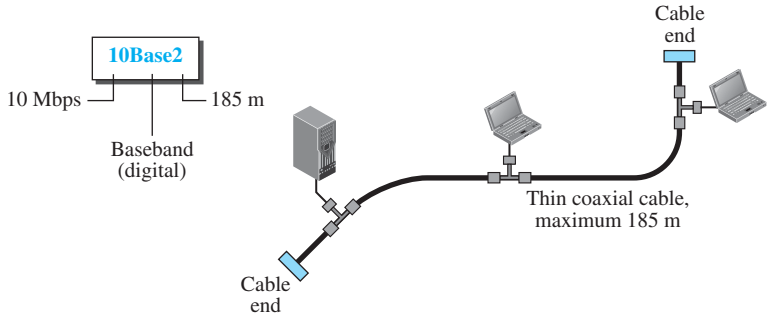
The transceiver is responsible for transmitting, receiving, and detecting collisions. The transceiver is connected to the station via a transceiver cable that provides separate paths for sending and receiving. This means that collision can only happen in the coaxial cable.

The maximum length of the coaxial cable must not exceed 500 m, otherwise, there is excessive degradation of the signal. If a length of more than 500 m is needed, up to five segments, each a maximum of 500 meters, can be connected using repeaters. Repeaters will be discussed in Chapter 17.

10Base2: Thin Ethernet

The second implementation is called ***10Base2, thin Ethernet,*** or ***Cheapernet.*** 10Base2 also uses a bus topology, but the cable is much thinner and more flexible. The cable can be bent to pass very close to the stations. In this case, the transceiver is normally part of the network interface card (NIC), which is installed inside the station. Figure 13.8 shows the schematic diagram of a 10Base2 implementation.

Figure 13.8 *10Base2 implementation*

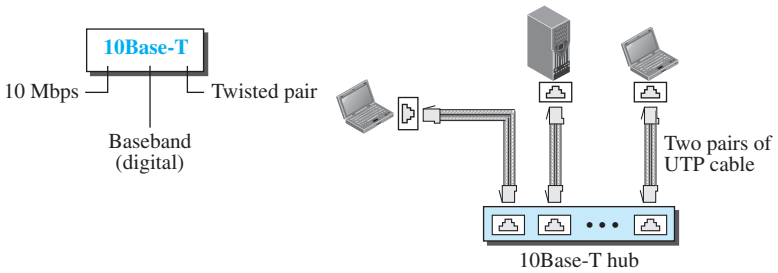


Note that the collision here occurs in the thin coaxial cable. This implementation is more cost effective than 10Base5 because thin coaxial cable is less expensive than thick coaxial and the tee connections are much cheaper than taps. Installation is simpler because the thin coaxial cable is very flexible. However, the length of each segment cannot exceed 185 m (close to 200 m) due to the high level of attenuation in thin coaxial cable.

10Base-T: Twisted-Pair Ethernet

The third implementation is called ***10Base-T*** or ***twisted-pair Ethernet.*** 10Base-T uses a physical star topology. The stations are connected to a hub via two pairs of twisted cable, as shown in Figure 13.9.

Figure 13.9 *10Base-T implementation*

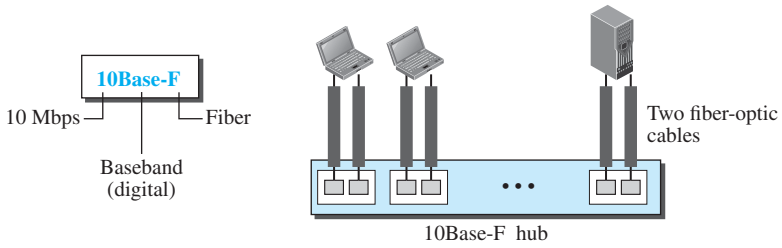


Note that two pairs of twisted cable create two paths (one for sending and one for receiving) between the station and the hub. Any collision here happens in the hub. Compared to 10Base5 or 10Base2, we can see that the hub actually replaces the coaxial cable as far as a collision is concerned. The maximum length of the twisted cable here is defined as 100 m, to minimize the effect of attenuation in the twisted cable.

10Base-F: Fiber Ethernet

Although there are several types of optical fiber 10-Mbps Ethernet, the most common is called ***10Base-F.*** 10Base-F uses a star topology to connect stations to a hub. The stations are connected to the hub using two fiber-optic cables, as shown in Figure 13.10.

Figure 13.10 *10Base-F implementation*



13.2.6 Changes in the Standard

Before we discuss higher-rate Ethernet protocols, we need to discuss the changes that occurred to the 10-Mbps Standard Ethernet. These changes actually opened the road to the evolution of the Ethernet to become compatible with other high-data-rate LANs.

Bridged Ethernet

The first step in the Ethernet evolution was the division of a LAN by **bridges.** Bridges have two effects on an Ethernet LAN: They raise the bandwidth and they separate collision domains. We discuss bridges in Chapter 17.

Raising the Bandwidth

In an unbridged Ethernet network, the total capacity (10 Mbps) is shared among all stations with a frame to send; the stations share the bandwidth of the network. If only one station has frames to send, it benefits from the total capacity (10 Mbps). But if more than one station needs to use the network, the capacity is shared. For example, if two stations have a lot of frames to send, they probably alternate in usage. When one station is sending, the other one refrains from sending. We can say that, in this case, each station on average sends at a rate of 5 Mbps. Figure 13.11 shows the situation.

The bridge, as we will learn in Chapter 17, can help here. A bridge divides the network into two or more networks. Bandwidthwise, each network is independent. For example, in Figure 13.12, a network with 12 stations is divided into two networks, each with 6 stations. Now each network has a capacity of 10 Mbps. The 10-Mbps capacity in each segment is now shared between 6 stations (actually 7 because the bridge acts as a

Figure 13.11 *Sharing bandwidth*

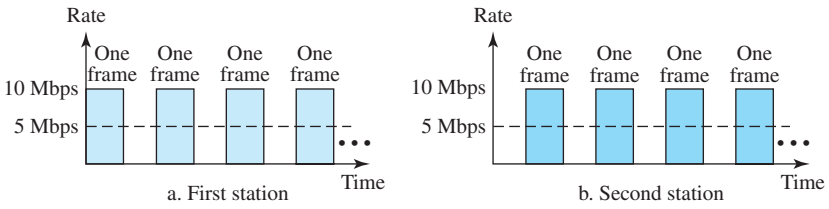
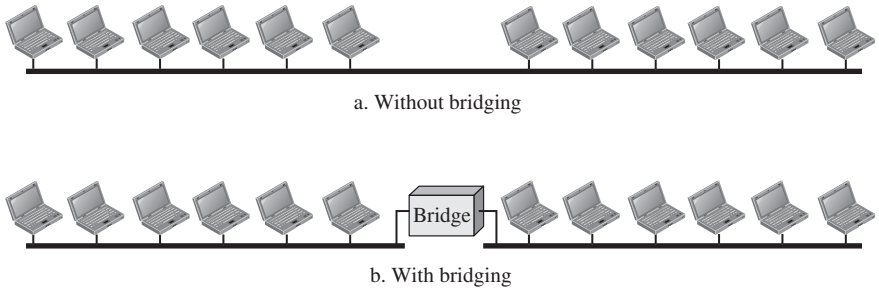


Figure 13.12 *A network with and without a bridge*



station in each segment), not 12 stations. In a network with a heavy load, eacsh station theoretically is offered 10/7 Mbps instead of 10/12 Mbps.

It is obvious that if we further divide the network, we can gain more bandwidth for each segment. For example, if we use a four-port bridge, each station is now offered 10/4 Mbps, which is 3 times more than an unbridged network.

Separating Collision Domains

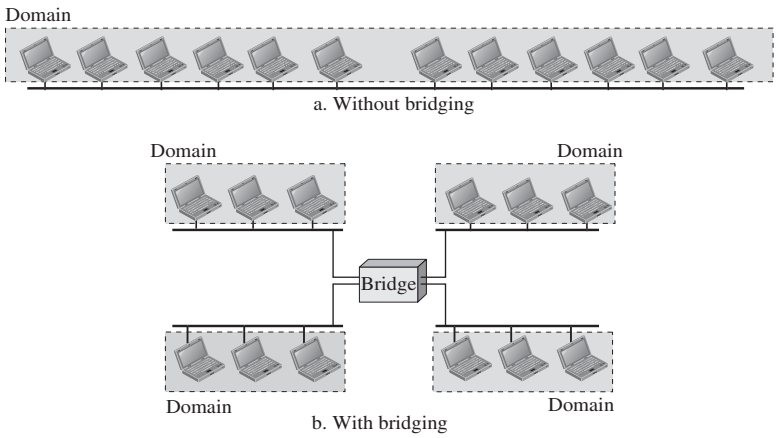
Another advantage of a bridge is the separation of the **collision domain.** Figure 13.13 shows the collision domains for an unbridged and a bridged network. You can see that the collision domain becomes much smaller and the probability of collision is reduced tremendously. Without bridging, 12 stations contend for access to the medium; with bridging only 3 stations contend for access to the medium.

Switched Ethernet

The idea of a bridged LAN can be extended to a switched LAN. Instead of having two to four networks, why not have *N* networks, where *N* is the number of stations on the LAN? In other words, if we can have a multiple-port bridge, why not have an *N*-port switch? In this way, the bandwidth is shared only between the station and the switch (5 Mbps each). In addition, the collision domain is divided into *N* domains.

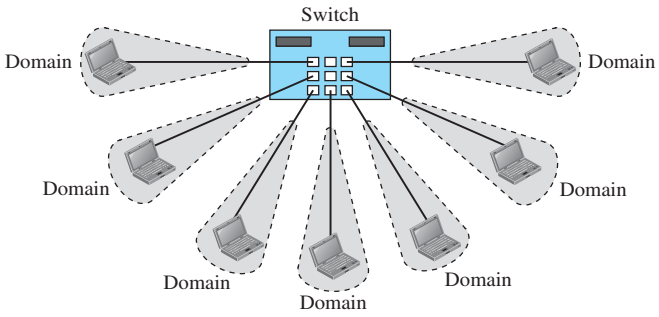
A layer-2 **switch** is an *N*-port bridge with additional sophistication that allows faster handling of the packets. Evolution from a bridged Ethernet to a **switched Ethernet** was

Figure 13.13 *Collision domains in an unbridged network and a bridged network*



a big step that opened the way to an even faster Ethernet, as we will see. Figure 13.14 shows a switched LAN.

Figure 13.14 *Switched Ethernet*



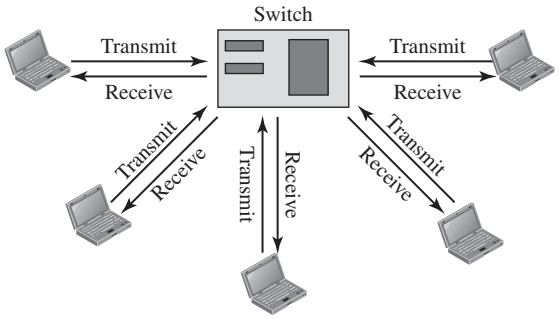
Full-Duplex Ethernet

One of the limitations of 10Base5 and 10Base2 is that communication is half-duplex (10Base-T is always full-duplex); a station can either send or receive, but may not do both at the same time. The next step in the evolution was to move from switched Ethernet to **full-duplex switched Ethernet.** The full-duplex mode increases the capacity of each domain from 10 to 20 Mbps. Figure 13.15 shows a switched Ethernet in full-duplex mode. Note that instead of using one link between the station and the switch, the configuration uses two links: one to transmit and one to receive.

No Need for CSMA/CD

In full-duplex switched Ethernet, there is no need for the CSMA/CD method. In a full-duplex switched Ethernet, each station is connected to the switch via two separate links.

Figure 13.15 *Full-duplex switched Ethernet*



Each station or switch can send and receive independently without worrying about collision. Each link is a point-to-point dedicated path between the station and the switch. There is no longer a need for carrier sensing; there is no longer a need for collision detection. The job of the MAC layer becomes much easier. The carrier sensing and collision detection functionalities of the MAC sublayer can be turned off.

MAC Control Layer

Standard Ethernet was designed as a connectionless protocol at the MAC sublayer. There is no explicit flow control or error control to inform the sender that the frame has arrived at the destination without error. When the receiver receives the frame, it does not send any positive or negative acknowledgment.

To provide for flow and error control in full-duplex switched Ethernet, a new sublayer, called the *MAC control,* is added between the LLC sublayer and the MAC sublayer.

13.3 FAST ETHERNET (100 MBPS)

In the 1990s, some LAN technologies with transmission rates higher than 10 Mbps, such as FDDI and Fiber Channel, appeared on the market. If the Standard Ethernet wanted to survive, it had to compete with these technologies. Ethernet made a big jump by increasing the transmission rate to 100 Mbps, and the new generation was called the *Fast Ethernet*. The designers of the Fast Ethernet needed to make it compatible with the Standard Ethernet. The MAC sublayer was left unchanged, which meant the frame format and the maximum and minimum size could also remain unchanged. By increasing the transmission rate, features of the Standard Ethernet that depend on the transmission rate, access method, and implementation had to be reconsidered. The goals of Fast Ethernet can be summarized as follows:

1. Upgrade the data rate to 100 Mbps.
2. Make it compatible with Standard Ethernet.
3. Keep the same 48-bit address.
4. Keep the same frame format.

13.3.1 Access Method

We remember that the proper operation of the CSMA/CD depends on the transmission rate, the minimum size of the frame, and the maximum network length. If we want to keep the minimum size of the frame, the maximum length of the network should be changed. In other words, if the minimum frame size is still 512 bits, and it is transmitted 10 times faster, the collision needs to be detected 10 times sooner, which means the maximum length of the network should be 10 times shorter (the propagation speed does not change). So the Fast Ethernet came with two solutions (it can work with either choice):

1. The first solution was to totally drop the bus topology and use a passive hub and star topology but make the maximum size of the network 250 meters instead of 2500 meters as in the Standard Ethernet. This approach is kept for compatibility with the Standard Ethernet.
2. The second solution is to use a link-layer switch (discussed later in the chapter) with a buffer to store frames and a full-duplex connection to each host to make the transmission medium private for each host. In this case, there is no need for CSMA/CD because the hosts are not competing with each other. The link-layer switch receives a frame from a source host and stores it in the buffer (queue) waiting for processing. It then checks the destination address and sends the frame out of the corresponding interface. Since the connection to the switch is full-duplex, the destination address can even send a frame to another station at the same time that it is receiving a frame. In other words, the shared medium is changed to many point-to-point media, and there is no need for contention.

Autonegotiation

A new feature added to Fast Ethernet is called ***autonegotiation***. It allows a station or a hub a range of capabilities. Autonegotiation allows two devices to negotiate the mode or data rate of operation. It was designed particularly to allow incompatible devices to connect to one another. For example, a device with a maximum data rate of 10 Mbps can communicate with a device with a 100 Mbps data rate (but which can work at a lower rate). We can summarize the goal of autonegotiation as follows. It was designed particularly for these purposes:

- To allow incompatible devices to connect to one another. For example, a device with a maximum capacity of 10 Mbps can communicate with a device with a 100 Mbps capacity (but which can work at a lower rate).
- To allow one device to have multiple capabilities.
- To allow a station to check a hub's capabilities.

13.3.2 Physical Layer

To be able to handle a 100 Mbps data rate, several changes need to be made at the physical layer.

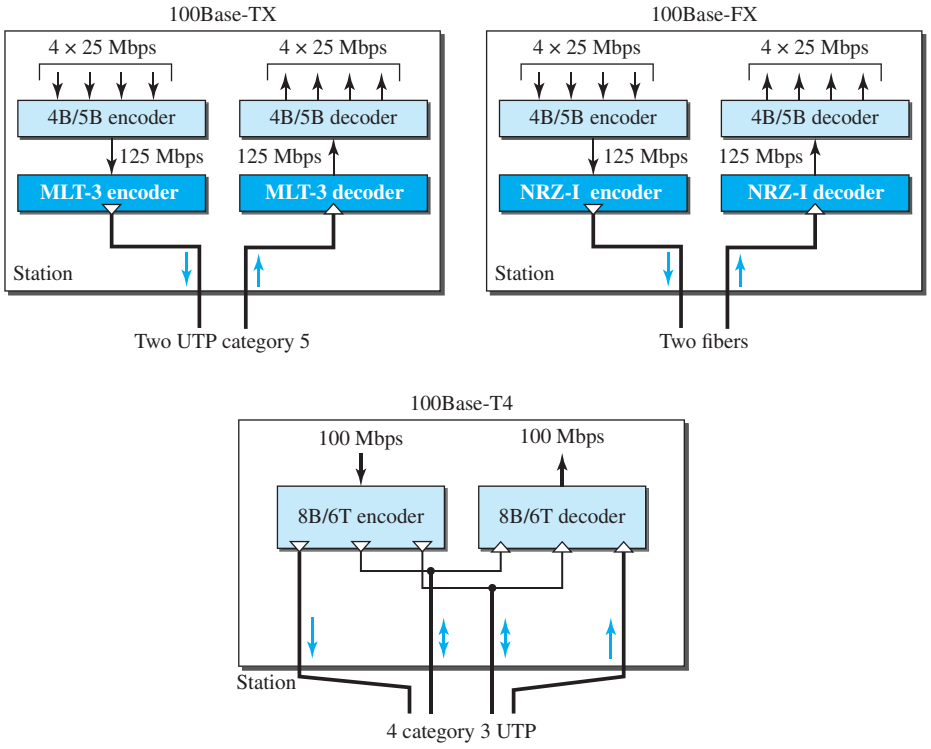
Topology

Fast Ethernet is designed to connect two or more stations. If there are only two stations, they can be connected point-to-point. Three or more stations need to be connected in a star topology with a hub or a switch at the center.

Encoding

Manchester encoding needs a 200-Mbaud bandwidth for a data rate of 100 Mbps, which makes it unsuitable for a medium such as twisted-pair cable. For this reason, the Fast Ethernet designers sought some alternative encoding/decoding scheme. However, it was found that one scheme would not perform equally well for all three implementations. Therefore, three different encoding schemes were chosen (see Figure 13.16).

Figure 13.16 *Encoding for Fast Ethernet implementation*



100Base-TX uses two pairs of twisted-pair cable (either category 5 UTP or STP). For this implementation, the MLT-3 scheme was selected since it has good bandwidth performance (see Chapter 4). However, since MLT-3 is not a self-synchronous line coding scheme, 4B/5B block coding is used to provide bit synchronization by preventing

the occurrence of a long sequence of 0s and 1s. This creates a data rate of 125 Mbps, which is fed into MLT-3 for encoding.

100Base-FX uses two pairs of fiber-optic cables. Optical fiber can easily handle high bandwidth requirements by using simple encoding schemes. The designers of 100Base-FX selected the NRZ-I encoding scheme (see Chapter 4) for this implementation. However, NRZ-I has a bit synchronization problem for long sequences of 0s (or 1s, based on the encoding), as we saw in Chapter 4. To overcome this problem, the designers used 4B/5B block encoding, as we described for 100Base-TX. The block encoding increases the bit rate from 100 to 125 Mbps, which can easily be handled by fiber-optic cable.

A 100Base-TX network can provide a data rate of 100 Mbps, but it requires the use of category 5 UTP or STP cable. This is not cost-efficient for buildings that have already been wired for voice-grade twisted-pair (category 3). A new standard, called ***100Base-T4,*** was designed to use category 3 or higher UTP. The implementation uses four pairs of UTP for transmitting 100 Mbps. Encoding/decoding in 100Base-T4 is more complicated. As this implementation uses category 3 UTP, each twisted-pair cannot easily handle more than 25 Mbaud. In this design, one pair switches between sending and receiving. Three pairs of UTP category 3, however, can handle only 75 Mbaud (25 Mbaud) each. We need to use an encoding scheme that converts 100 Mbps to a 75 Mbaud signal. As we saw in Chapter 4, 8B/6T satisfies this requirement. In 8B/6T, eight data elements are encoded as six signal elements. This means that 100 Mbps uses only (6/8) × 100 Mbps, or 75 Mbaud.

Summary

Fast Ethernet implementation at the physical layer can be categorized as either two-wire or four-wire. The two-wire implementation can be either shielded twisted pair (STP), which is called *100Base-TX,* or fiber-optic cable, which is called *100Base-FX*. The four-wire implementation is designed only for unshielded twisted pair (UTP), which is called *100Base-T4*. Table 13.2 is a summary of the Fast Ethernet implementations. We discussed encoding in Chapter 4.

Table 13.2 *Summary of Fast Ethernet implementations*

Implementation	*Medium*	*Medium Length*	*Wires*	*Encoding*
100Base-TX	UTP or STP	100 m	2	4B5B + MLT-3
100Base-FX	Fiber	185 m	2	4B5B + NRZ-I
100Base-T4	UTP	100 m	4	Two 8B/6T

13.4 GIGABIT ETHERNET

The need for an even higher data rate resulted in the design of the Gigabit Ethernet Protocol (1000 Mbps). The IEEE committee calls it the Standard 802.3z. The goals of the Gigabit Ethernet were to upgrade the data rate to 1 Gbps, but keep the address length, the frame format, and the maximum and minimum frame length the same. The goals of the Gigabit Ethernet design can be summarized as follows:

1. Upgrade the data rate to 1 Gbps.
2. Make it compatible with Standard or Fast Ethernet.

3. Use the same 48-bit address.
4. Use the same frame format.
5. Keep the same minimum and maximum frame lengths.
6. Support autonegotiation as defined in Fast Ethernet.

13.4.1 MAC Sublayer

A main consideration in the evolution of Ethernet was to keep the MAC sublayer untouched. However, to achieve a data rate of 1 Gbps, this was no longer possible. Gigabit Ethernet has two distinctive approaches for medium access: half-duplex and full-duplex. Almost all implementations of Gigabit Ethernet follow the full-duplex approach, so we mostly ignore the half-duplex mode.

Full-Duplex Mode

In full-duplex mode, there is a central switch connected to all computers or other switches. In this mode, for each input port, each switch has buffers in which data are stored until they are transmitted. Since the switch uses the destination address of the frame and sends a frame out of the port connected to that particular destination, there is no collision. This means that CSMA/CD is not used. Lack of collision implies that the maximum length of the cable is determined by the signal attenuation in the cable, not by the collision detection process.

In the full-duplex mode of Gigabit Ethernet, there is no collision; the maximum length of the cable is determined by the signal attenuation in the cable.

Half-Duplex Mode

Gigabit Ethernet can also be used in half-duplex mode, although it is rare. In this case, a switch can be replaced by a hub, which acts as the common cable in which a collision might occur. The half-duplex approach uses CSMA/CD. However, as we saw before, the maximum length of the network in this approach is totally dependent on the minimum frame size. Three methods have been defined: traditional, carrier extension, and frame bursting.

Traditional

In the traditional approach, we keep the minimum length of the frame as in traditional Ethernet (512 bits). However, because the length of a bit is 1/100 shorter in Gigabit Ethernet than in 10-Mbps Ethernet, the slot time for Gigabit Ethernet is 512 bits × 1/1000 μs, which is equal to 0.512 μs. The reduced slot time means that collision is detected 100 times earlier. This means that the maximum length of the network is 25 m. This length may be suitable if all the stations are in one room, but it may not even be long enough to connect the computers in one single office.

Carrier Extension

To allow for a longer network, we increase the minimum frame length. The **carrier extension** approach defines the minimum length of a frame as 512 bytes (4096 bits). This means that the minimum length is 8 times longer. This method forces a station to add

extension bits (padding) to any frame that is less than 4096 bits. In this way, the maximum length of the network can be increased 8 times to a length of 200 m. This allows a length of 100 m from the hub to the station.

Frame Bursting

Carrier extension is very inefficient if we have a series of short frames to send; each frame carries redundant data. To improve efficiency, **frame bursting** was proposed. Instead of adding an extension to each frame, multiple frames are sent. However, to make these multiple frames look like one frame, padding is added between the frames (the same as that used for the carrier extension method) so that the channel is not idle. In other words, the method deceives other stations into thinking that a very large frame has been transmitted.

13.4.2 Physical Layer

The physical layer in Gigabit Ethernet is more complicated than that in Standard or Fast Ethernet. We briefly discuss some features of this layer.

Topology

Gigabit Ethernet is designed to connect two or more stations. If there are only two stations, they can be connected point-to-point. Three or more stations need to be connected in a star topology with a hub or a switch at the center. Another possible configuration is to connect several star topologies or let one star topology be part of another.

Implementation

Gigabit Ethernet can be categorized as either a two-wire or a four-wire implementation. The two-wire implementations use fiber-optic cable (**1000Base-SX,** short-wave, or **1000Base-LX,** long-wave), or STP (**1000Base-CX**). The four-wire version uses category 5 twisted-pair cable (**1000Base-T**). In other words, we have four implementations. 1000Base-T was designed in response to those users who had already installed this wiring for other purposes such as Fast Ethernet or telephone services.

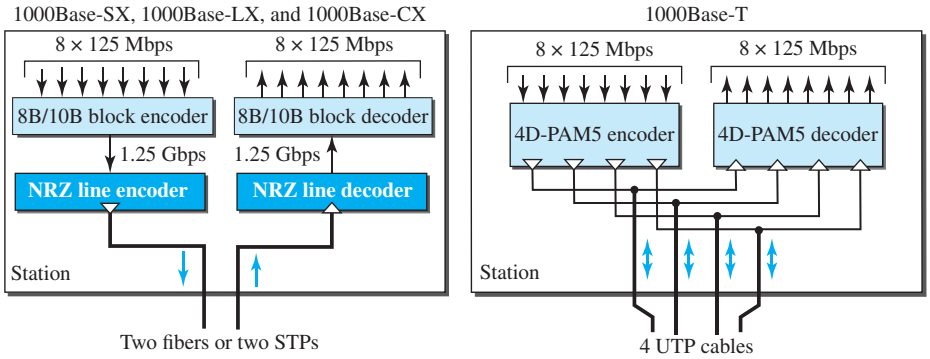
Encoding

Figure 13.17 shows the encoding/decoding schemes for the four implementations. Gigabit Ethernet cannot use the Manchester encoding scheme because it involves a very high bandwidth (2 GBaud). The two-wire implementations use an NRZ scheme, but NRZ does not self-synchronize properly. To synchronize bits, particularly at this high data rate, 8B/10B block encoding, discussed in Chapter 4, is used.

This block encoding prevents long sequences of 0s or 1s in the stream, but the resulting stream is 1.25 Gbps. Note that in this implementation, one wire (fiber or STP) is used for sending and one for receiving.

In the four-wire implementation it is not possible to have 2 wires for input and 2 for output, because each wire would need to carry 500 Mbps, which exceeds the capacity for category 5 UTP. As a solution, 4D-PAM5 encoding, as discussed in Chapter 4, is used to reduce the bandwidth. Thus, all four wires are involved in both input and output; each wire carries 250 Mbps, which is in the range for category 5 UTP cable.

Figure 13.17 *Encoding in Gigabit Ethernet implementations*



Implementation Summary

Table 13.3 is a summary of the Gigabit Ethernet implementations. S-W and L-W mean short-wave and long-wave respectively.

Table 13.3 *Summary of Gigabit Ethernet implementations*

Implementation	*Medium*	*Medium Length*	*Wires*	*Encoding*
1000Base-SX	Fiber S-W	550 m	2	8B/10B + NRZ
1000Base-LX	Fiber L-W	5000 m	2	8B/10B + NRZ
1000Base-CX	STP	25 m	2	8B/10B + NRZ
1000Base-T4	UTP	100 m	4	4D-PAM5

13.5 10 GIGABIT ETHERNET

In recent years, there has been another look into the Ethernet for use in metropolitan areas. The idea is to extend the technology, the data rate, and the coverage distance so that the Ethernet can be used as LAN and MAN (metropolitan area network). The IEEE committee created 10 Gigabit Ethernet and called it Standard 802.3ae. The goals of the 10 Gigabit Ethernet design can be summarized as upgrading the data rate to 10 Gbps, keeping the same frame size and format, and allowing the interconnection of LANs, MANs, and WAN possible. This data rate is possible only with fiber-optic technology at this time. The standard defines two types of physical layers: LAN PHY and WAN PHY. The first is designed to support existing LANs; the second actually defines a WAN with links connected through SONET OC-192.

13.5.1 Implementation

10 Gigabit Ethernet operates only in full-duplex mode, which means there is no need for contention; CSMA/CD is not used in 10 Gigabit Ethernet. Four implementations are the most common: **10GBase-SR, 10GBase-LR, 10GBase-EW,** and **10GBase-X4.** Table 13.4 shows a summary of the 10 Gigabit Ethernet implementations. We discussed the encoding in Chapter 4.

Table 13.4 *Summary of 10 Gigabit Ethernet implementations*

Implementation	*Medium*	*Medium Length*	*Number of wires*	*Encoding*
10GBase-SR	Fiber 850 nm	300 m	2	64B66B
10GBase-LR	Fiber 1310 nm	10 Km	2	64B66B
10GBase-EW	Fiber 1350 nm	40 Km	2	SONET
10GBase-X4	Fiber 1310 nm	300 m to 10 Km	2	8B10B

13.6 END-CHAPTER MATERIALS

13.6.1 Recommended Reading

For more details about subjects discussed in this chapter, we recommend the following books and RFCs. The items in brackets […] refer to the reference list at the end of the text.

Books

Several books discuss Ethernet. Among them we recommend [Ham 80], [Zar 02], [Ror 96], [Tan 03], [GW 04], [For 03], [KMK 04], [Sta 04], [Kes 02], [PD 03], [Kei 02], [Spu 00], [KCK 98], [Sau 98], [Izz 00], [Per 00], and [WV 00].

RFCs

A discussion of the use of the checksum in the Internet can be found in RFC 1141.

13.6.2 Key Terms

10 Gigabit Ethernet
1000Base-CX
1000Base-LX
1000Base-SX
1000Base-T
100Base-FX
100Base-T4
100Base-TX
10Base2
10Base5
10Base-F
10Base-T
10GBase-EW
10GBase-LR
10GBase-SR
10GBase-X4
autonegotiation
bridge
carrier extension
Cheapernet
collision domain
Fast Ethernet
frame bursting
full-duplex switched Ethernet
Gigabit Ethernet
hexadecimal notation
logical link control (LLC)
media access control (MAC)
network interface card (NIC)
Project 802
Standard Ethernet
switch
switched Ethernet
thick Ethernet
Thicknet
thin Ethernet
transceiver
twisted-pair Ethernet

13.6.3 Summary

Ethernet is the most widely used local area network protocol. The IEEE 802.3 Standard defines 1-persistent CSMA/CD as the access method for first-generation 10-Mbps Ethernet. The data-link layer of Ethernet consists of the LLC sublayer and the MAC sublayer. The MAC sublayer is responsible for the operation of the CSMA/CD access

method and framing. Each station on an Ethernet network has a unique 48-bit address imprinted on its network interface card (NIC). The minimum frame length for 10-Mbps Ethernet is 64 bytes; the maximum is 1518 bytes.

The common implementations of 10-Mbps Ethernet are 10Base5 (thick Ethernet), 10Base2 (thin Ethernet), 10Base-T (twisted-pair Ethernet), and 10Base-F (fiber Ethernet). The 10Base5 implementation of Ethernet uses thick coaxial cable. 10Base2 uses thin coaxial cable. 10Base-T uses four twisted-pair cables that connect each station to a common hub. 10Base-F uses fiber-optic cable. A bridge can increase the bandwidth and separate the collision domains on an Ethernet LAN. A switch allows each station on an Ethernet LAN to have the entire capacity of the network to itself. Full-duplex mode doubles the capacity of each domain and removes the need for the CSMA/CD method.

Fast Ethernet has a data rate of 100 Mbps. In Fast Ethernet, autonegotiation allows two devices to negotiate the mode or data rate of operation. The common Fast Ethernet implementations are 100Base-TX (two pairs of twisted-pair cable), 100Base-FX (two fiber-optic cables), and 100Base-T4 (four pairs of voice-grade, or higher, twisted-pair cable).

Gigabit Ethernet has a data rate of 1000 Mbps. Gigabit Ethernet access methods include half-duplex mode using traditional CSMA/CD (not common) and full-duplex mode (most popular method). The common Gigabit Ethernet implementations are 1000Base-SX (two optical fibers and a short-wave laser source), 1000Base-LX (two optical fibers and a long-wave laser source), and 1000Base-T (four twisted pairs).

The latest Ethernet standard is 10 Gigabit Ethernet, which operates at 10 Gbps. The four common implementations are 10GBase-SR, 10GBase-LR, 10GBase-EW, and 10GBase-X4. These implementations use fiber-optic cables in full-duplex mode.

13.7 PRACTICE SET

13.7.1 Quizzes

A set of interactive quizzes for this chapter can be found on the book website. It is strongly recommended that the student take the quizzes to check his/her understanding of the materials before continuing with the practice set.

13.7.2 Questions

Q13-1. Why is there no need for CSMA/CD on a full-duplex Ethernet LAN?

Q13-2. Compare the data rates for Standard Ethernet, Fast Ethernet, Gigabit Ethernet, and 10 Gigabit Ethernet.

Q13-3. What are the common Standard Ethernet implementations?

Q13-4. What are the common Fast Ethernet implementations?

Q13-5. What are the common Gigabit Ethernet implementations?

Q13-6. What are the common 10 Gigabit implementations?

Q13-7. How is the preamble field different from the SFD field?

Q13-8. What is the difference between unicast, multicast, and broadcast addresses?

Q13-9. What are the advantages of dividing an Ethernet LAN with a bridge?

Q13-10. What is the relationship between a switch and a bridge?

13.7.3 Problems

P13-1. What is the hexadecimal equivalent of the following Ethernet address?

```
01011010 00010001 01010101 00011000 10101010 00001111
```

P13-2. How does the Ethernet address 1A:2B:3C:4D:5E:6F appear on the line in binary?

P13-3. If an Ethernet destination address is 07:01:02:03:04:05, what is the type of the address (unicast, multicast, or broadcast)?

P13-4. An Ethernet MAC sublayer receives 42 bytes of data from the upper layer. How many bytes of padding must be added to the data?

P13-5. What is the ratio of useful data to the entire packet for the smallest Ethernet frame?

P13-6. Suppose the length of a 10Base5 cable is 2500 m. If the speed of propagation in a thick coaxial cable is 200,000,000 m/s, how long does it take for a bit to travel from the beginning to the end of the network? Assume there is a 10 μs delay in the equipment.

P13-7. Suppose you are to design a LAN for a company that has 100 employees, each given a desktop computer attached to the LAN. What should be the data rate of the LAN if the typical use of the LAN is shown below:

- **a.** Each employee needs to retrieve a file of average size of 10 megabytes in a second. An employee may do this on average 10 times during the eight-hour working time.
- **b.** Each employee needs to access the Internet at 250 Kbps. This can happen for 10 employees simultaneously.
- **c.** Each employee may receive 10 e-mails per hour with an average size of 100 kilobytes. Half of the employees may receive e-mails simultaneously.

P13-8. In a Standard Ethernet LAN, the average size of a frame is 1000 bytes. If a noise of 2 ms occurs on the LAN, how many frames are destroyed?

P13-9. Repeat Problem P13-8 for a Fast Ethernet LAN.

P13-10. Repeat Problem P13-8 for a Gigabit Ethernet LAN.

P13-11. Repeat Problem P13-8 for a 10 Gigabit Ethernet LAN.

13.8 SIMULATION EXPERIMENTS

13.8.1 Applets

We have created some Java applets to show some of the main concepts discussed in this chapter. It is strongly recommended that the students activate these applets on the book website and carefully examine the protocols in action.

13.8.2 Lab Assignments

In this section, we use Wireshark to simulate two protocols: Ethernet and ARP. Full descriptions of these lab assignments are on the book website.

Lab13-1. In this lab we need to examine the contents of a frame sent by the data-link layer. We want to find the value of different fields such as destination and source MAC addresses, the value of CRC, the value of the protocol field which shows which payload is being carried by the frame, and so on.

CHAPTER 15

Wireless LANs

We discussed wired LANs and wired WANs in the two previous chapters. We concentrate on wireless LANs in this chapter and wireless WANs in the next.

In this chapter, we cover two types of wireless LANs. The first is the wireless LAN defined by the IEEE 802.11 project (sometimes called *wireless Ethernet*); the second is a personal wireless LAN, Bluetooth, that is sometimes called *personal area network* or *PAN*.

This chapter is divided into three sections:

- The first section introduces the general issues behind wireless LANs and compares wired and wireless networks. The section describes the characteristics of the wireless networks and the way access is controlled in these types of networks.
- The second section discusses a wireless LAN defined by the IEEE 802.11 Project, which is sometimes called *wireless Ethernet*. This section defines the architecture of this type of LAN and describes the MAC sublayer, which uses the CSMA/CA access method discussed in Chapter 12. The section then shows the addressing mechanism used in this network and gives the format of different packets used at the data-link layer. Finally, the section discusses different physical-layer protocols that are used by this type of network.
- The third section discusses the Bluetooth technology as a personal area network (PAN). The section describes the architecture of the network, the addressing mechanism, and the packet format. Different layers used in this protocol are also briefly described and compared with the ones in the other wired and wireless LANs.

15.1 INTRODUCTION

Wireless communication is one of the fastest-growing technologies. The demand for connecting devices without the use of cables is increasing everywhere. Wireless LANs can be found on college campuses, in office buildings, and in many public areas. Before we discuss a specific protocol related to wireless LANs, let us talk about them in general.

15.1.1 Architectural Comparison

Let us first compare the architecture of wired and wireless LANs to give some idea of what we need to look for when we study wireless LANs.

Medium

The first difference we can see between a wired and a wireless LAN is the medium. In a wired LAN, we use wires to connect hosts. In Chapter 7, we saw that we moved from multiple access to point-to-point access through the generation of the Ethernet. In a switched LAN, with a link-layer switch, the communication between the hosts is point-to-point and full-duplex (bidirectional). In a wireless LAN, the medium is air, the signal is generally broadcast. When hosts in a wireless LAN communicate with each other, they are sharing the same medium (multiple access). In a very rare situation, we may be able to create a point-to-point communication between two wireless hosts by using a very limited bandwidth and two-directional antennas. Our discussion in this chapter, however, is about the multiple-access medium, which means we need to use MAC protocols.

Hosts

In a wired LAN, a host is always connected to its network at a point with a fixed link-layer address related to its network interface card (NIC). Of course, a host can move from one point in the Internet to another point. In this case, its link-layer address remains the same, but its network-layer address will change, as we see later in Chapter 19, Section 19.3 (Mobile IP section). However, before the host can use the services of the Internet, it needs to be physically connected to the Internet. In a wireless LAN, a host is not physically connected to the network; it can move freely (as we'll see) and can use the services provided by the network. Therefore, mobility in a wired network and wireless network are totally different issues, which we try to clarify in this chapter.

Isolated LANs

The concept of a wired isolated LAN also differs from that of a wireless isolated LAN. A wired isolated LAN is a set of hosts connected via a link-layer switch (in the recent generation of Ethernet). A wireless isolated LAN, called an ***ad hoc network*** in wireless LAN terminology, is a set of hosts that communicate freely with each other. The concept of a link-layer switch does not exist in wireless LANs. Figure 15.1 shows two isolated LANs, one wired and one wireless.

Connection to Other Networks

A wired LAN can be connected to another network or an internetwork such as the Internet using a router. A wireless LAN may be connected to a wired infrastructure network,

to a wireless infrastructure network, or to another wireless LAN. The first situation is the one that we discuss in this section: connection of a wireless LAN to a wired infrastructure network. Figure 15.2 shows the two environments.

Figure 15.1 *Isolated LANs: wired versus wireless*

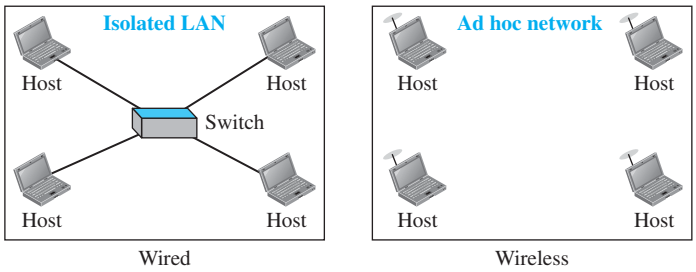
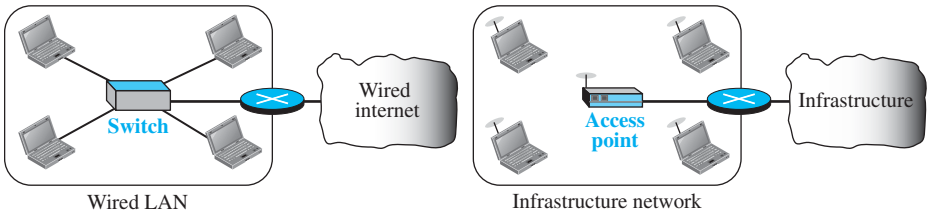


Figure 15.2 *Connection of a wired LAN and a wireless LAN to other networks*



In this case, the wireless LAN is referred to as an *infrastructure network,* and the connection to the wired infrastructure, such as the Internet, is done via a device called an ***access point (AP)***. Note that the role of the access point is completely different from the role of a link-layer switch in the wired environment. An access point is gluing two different environments together: one wired and one wireless. Communication between the AP and the wireless host occurs in a wireless environment; communication between the AP and the infrastructure occurs in a wired environment.

Moving between Environments

The discussion above confirms what we learned in Chapters 2 and 9: a wired LAN or a wireless LAN operates only in the lower two layers of the TCP/IP protocol suite. This means that if we have a wired LAN in a building that is connected via a router or a modem to the Internet, all we need in order to move from the wired environment to a wireless environment is to change the network interface cards designed for wired environments to the ones designed for wireless environments and replace the link-layer switch with an access point. In this change, the link-layer addresses will change (because of changing NICs), but the network-layer addresses (IP addresses) will remain the same; we are moving from wired links to wireless links.

15.1.2 Characteristics

There are several characteristics of wireless LANs that either do not apply to wired LANs or the existence of which is negligible and can be ignored. We discuss some of these characteristics here to pave the way for discussing wireless LAN protocols.

Attenuation

The strength of electromagnetic signals decreases rapidly because the signal disperses in all directions; only a small portion of it reaches the receiver. The situation becomes worse with mobile senders that operate on batteries and normally have small power supplies.

Interference

Another issue is that a receiver may receive signals not only from the intended sender, but also from other senders if they are using the same frequency band.

Multipath Propagation

A receiver may receive more than one signal from the same sender because electromagnetic waves can be reflected back from obstacles such as walls, the ground, or objects. The result is that the receiver receives some signals at different phases (because they travel different paths). This makes the signal less recognizable.

Error

With the above characteristics of a wireless network, we can expect that errors and error detection are more serious issues in a wireless network than in a wired network. If we think about the error level as the measurement of **signal-to-noise ratio (SNR),** we can better understand why error detection and error correction and retransmission are more important in a wireless network. We discussed SNR in more detail in Chapter 3, but it is enough to say that it measures the ratio of good stuff to bad stuff (signal to noise). If SNR is high, it means that the signal is stronger than the noise (unwanted signal), so we may be able to convert the signal to actual data. On the other hand, when SNR is low, it means that the signal is corrupted by the noise and the data cannot be recovered.

15.1.3 Access Control

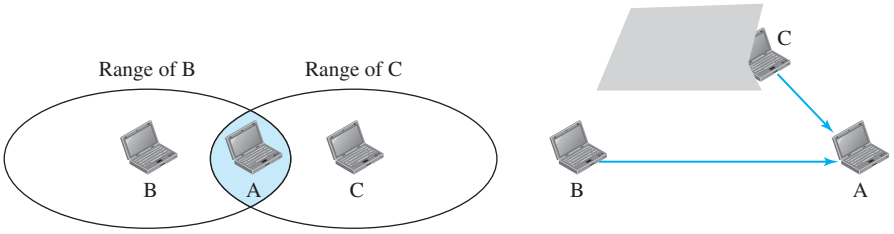
Maybe the most important issue we need to discuss in a wireless LAN is access control—how a wireless host can get access to the shared medium (air). We discussed in Chapter 12 that the Standard Ethernet uses the CSMA/CD algorithm. In this method, each host contends to access the medium and sends its frame if it finds the medium idle. If a collision occurs, it is detected and the frame is sent again. Collision detection in CSMA/CD serves two purposes. If a collision is detected, it means that the frame has not been received and needs to be resent. If a collision is not detected, it is a kind of acknowledgment that the frame was received. The CSMA/CD algorithm does not work in wireless LANs for three reasons:

1. To detect a collision, a host needs to send and receive at the same time (sending the frame and receiving the collision signal), which means the host needs to work in a

duplex mode. Wireless hosts do not have enough power to do so (the power is supplied by batteries). They can only send or receive at one time.

2. Because of the hidden station problem, in which a station may not be aware of another station's transmission due to some obstacles or range problems, collision may occur but not be detected. Figure 15.3 shows an example of the hidden station problem. Station B has a transmission range shown by the left oval (sphere in space); every station in this range can hear any signal transmitted by station B. Station C has a transmission range shown by the right oval (sphere in space); every station located in this range can hear any signal transmitted by C. Station C is

Figure 15.3 *Hidden station problem*



outside the transmission range of B; likewise, station B is outside the transmission range of C. Station A, however, is in the area covered by both B and C; it can hear any signal transmitted by B or C. The figure also shows that the hidden station problem may also occur due to an obstacle.

Assume that station B is sending data to station A. In the middle of this transmission, station C also has data to send to station A. However, station C is out of B's range and transmissions from B cannot reach C. Therefore C thinks the medium is free. Station C sends its data to A, which results in a collision at A because this station is receiving data from both B and C. In this case, we say that stations B and C are hidden from each other with respect to A. Hidden stations can reduce the capacity of the network because of the possibility of collision.

3. The distance between stations can be great. Signal fading could prevent a station at one end from hearing a collision at the other end.

To overcome the above three problems, Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) was invented for wireless LANs, which we discussed in Chapter 12.

15.2 IEEE 802.11 PROJECT

IEEE has defined the specifications for a wireless LAN, called IEEE 802.11, which covers the physical and data-link layers. It is sometimes called *wireless Ethernet*. In

some countries, including the United States, the public uses the term *WiFi* (short for wireless fidelity) as a synonym for *wireless LAN*. WiFi, however, is a wireless LAN that is certified by the WiFi Alliance, a global, nonprofit industry association of more than 300 member companies devoted to promoting the growth of wireless LANs.

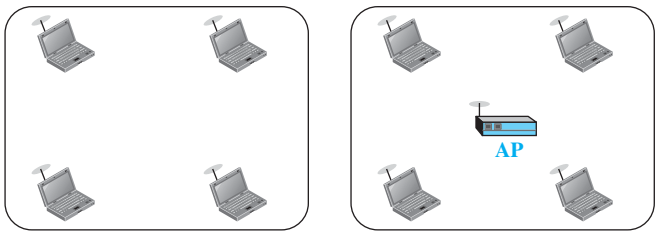
15.2.1 Architecture

The standard defines two kinds of services: the basic service set (BSS) and the extended service set (ESS).

Basic Service Set

IEEE 802.11 defines the **basic service set (BSS)** as the building blocks of a wireless LAN. A basic service set is made of stationary or mobile wireless stations and an optional central base station, known as the *access point (AP)*. Figure 15.4 shows two sets in this standard.

Figure 15.4 *Basic service sets (BSSs)*



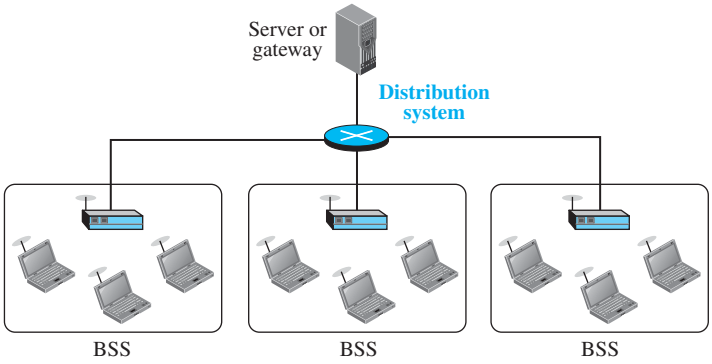
The BSS without an AP is a stand-alone network and cannot send data to other BSSs. It is called an *ad hoc architecture*. In this architecture, stations can form a network without the need of an AP; they can locate one another and agree to be part of a BSS. A BSS with an AP is sometimes referred to as an *infrastructure BSS*.

Extended Service Set

An **extended service set (ESS)** is made up of two or more BSSs with APs. In this case, the BSSs are connected through a *distribution system,* which is a wired or a wireless network. The distribution system connects the APs in the BSSs. IEEE 802.11 does not restrict the distribution system; it can be any IEEE LAN such as an Ethernet. Note that the extended service set uses two types of stations: mobile and stationary. The mobile stations are normal stations inside a BSS. The stationary stations are AP stations that are part of a wired LAN. Figure 15.5 shows an ESS.

When BSSs are connected, the stations within reach of one another can communicate without the use of an AP. However, communication between a station in a BSS and the outside BSS occurs via the AP. The idea is similar to communication in a cellular network (discussed in Chapter 16) if we consider each BSS to be a cell and each AP to be a base station. Note that a mobile station can belong to more than one BSS at the same time.

Figure 15.5 *Extended service set (ESS)*



Station Types

IEEE 802.11 defines three types of stations based on their mobility in a wireless LAN: no-transition, BSS-transition, and ESS-transition mobility. A station with **no-transition mobility** is either stationary (not moving) or moving only inside a BSS. A station with **BSS-transition mobility** can move from one BSS to another, but the movement is confined inside one ESS. A station with **ESS-transition mobility** can move from one ESS to another. However, IEEE 802.11 does not guarantee that communication is continuous during the move.

15.2.2 MAC Sublayer

IEEE 802.11 defines two MAC sublayers: the distributed coordination function (DCF) and point coordination function (PCF). Figure 15.6 shows the relationship between the two MAC sublayers, the LLC sublayer, and the physical layer. We discuss the physical layer implementations later in the chapter and will now concentrate on the MAC sublayer.

Distributed Coordination Function

One of the two protocols defined by IEEE at the MAC sublayer is called the ***distributed coordination function (DCF)***. DCF uses CSMA/CA as the access method (see Chapter 12).

Frame Exchange Time Line

Figure 15.7 shows the exchange of data and control frames in time.

1. Before sending a frame, the source station senses the medium by checking the energy level at the carrier frequency.
   a. The channel uses a persistence strategy with backoff until the channel is idle.
   b. After the station is found to be idle, the station waits for a period of time called the ***distributed interframe space (DIFS);*** then the station sends a control frame called the *request to send (RTS)*.

Figure 15.6 *MAC layers in IEEE 802.11 standard*

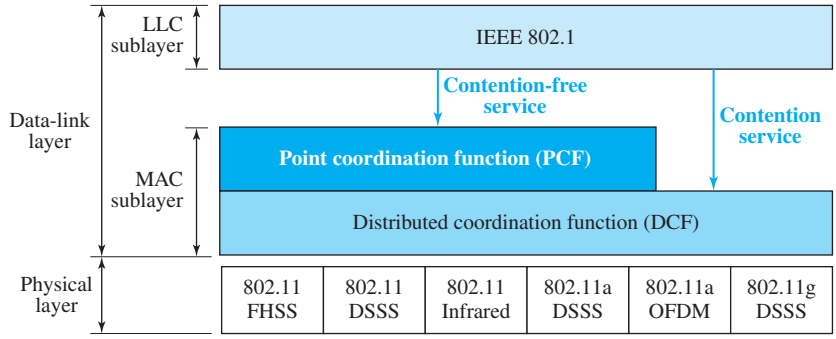
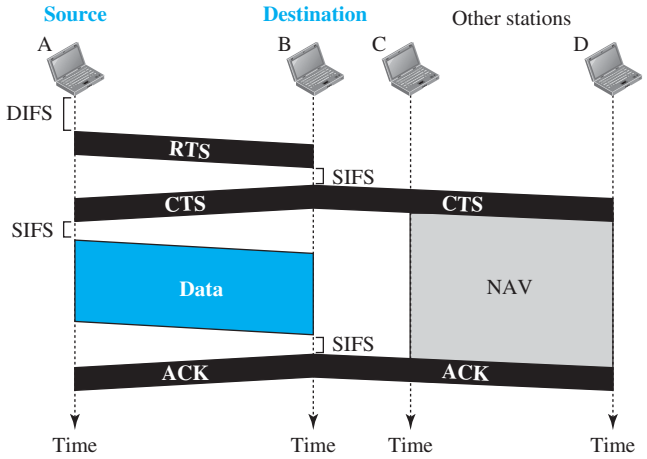


Figure 15.7 *CSMA/CA and NAV*



2. After receiving the RTS and waiting a period of time called the ***short interframe space (SIFS),*** the destination station sends a control frame, called the *clear to send (CTS),* to the source station. This control frame indicates that the destination station is ready to receive data.
3. The source station sends data after waiting an amount of time equal to SIFS.
4. The destination station, after waiting an amount of time equal to SIFS, sends an acknowledgment to show that the frame has been received. Acknowledgment is needed in this protocol because the station does not have any means to check for the successful arrival of its data at the destination. On the other hand, the lack of collision in CSMA/CD is a kind of indication to the source that data have arrived.

Network Allocation Vector

How do other stations defer sending their data if one station acquires access? In other words, how is the *collision avoidance* aspect of this protocol accomplished*?* The key is a feature called NAV.

When a station sends an RTS frame, it includes the duration of time that it needs to occupy the channel. The stations that are affected by this transmission create a timer called a ***network allocation vector (NAV)*** that shows how much time must pass before these stations are allowed to check the channel for idleness. Each time a station accesses the system and sends an RTS frame, other stations start their NAV. In other words, each station, before sensing the physical medium to see if it is idle, first checks its NAV to see if it has expired. Figure 15.7 shows the idea of NAV.

Collision During Handshaking

What happens if there is a collision during the time when RTS or CTS control frames are in transition, often called the *handshaking period*? Two or more stations may try to send RTS frames at the same time. These control frames may collide. However, because there is no mechanism for collision detection, the sender assumes there has been a collision if it has not received a CTS frame from the receiver. The backoff strategy is employed, and the sender tries again.

Hidden-Station Problem

The solution to the hidden station problem is the use of the handshake frames (RTS and CTS). Figure 15.7 also shows that the RTS message from B reaches A, but not C. However, because both B and C are within the range of A, the CTS message, which contains the duration of data transmission from B to A, reaches C. Station C knows that some hidden station is using the channel and refrains from transmitting until that duration is over.

Point Coordination Function (PCF)

The **point coordination function (PCF)** is an optional access method that can be implemented in an infrastructure network (not in an ad hoc network). It is implemented on top of the DCF and is used mostly for time-sensitive transmission.

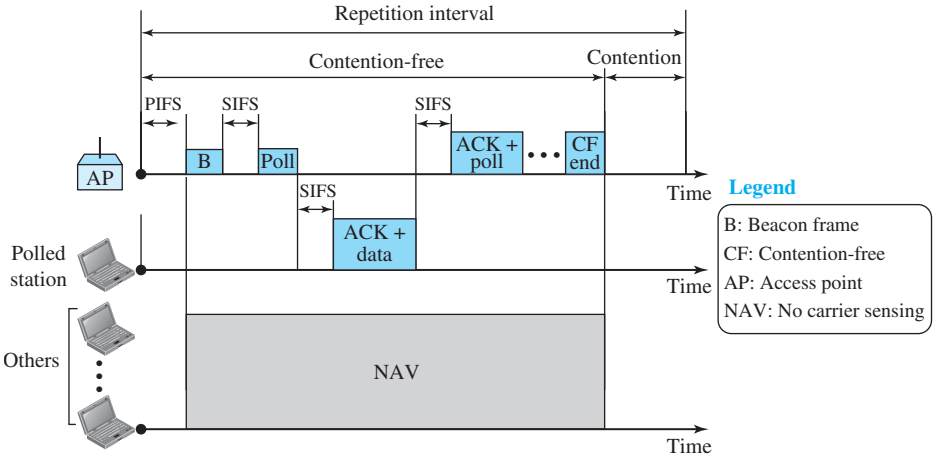
PCF has a centralized, contention-free polling access method, which we discussed in Chapter 12. The AP performs polling for stations that are capable of being polled. The stations are polled one after another, sending any data they have to the AP.

To give priority to PCF over DCF, another interframe space, PIFS, has been defined. PIFS (PCF IFS) is shorter than DIFS. This means that if, at the same time, a station wants to use only DCF and an AP wants to use PCF, the AP has priority.

Due to the priority of PCF over DCF, stations that only use DCF may not gain access to the medium. To prevent this, a repetition interval has been designed to cover both contention-free PCF and contention-based DCF traffic. The *repetition interval,* which is repeated continuously, starts with a special control frame, called a ***beacon frame*.** When the stations hear the beacon frame, they start their NAV for the duration of the contention-free period of the repetition interval. Figure 15.8 shows an example of a repetition interval.

During the repetition interval, the PC (point controller) can send a poll frame, receive data, send an ACK, receive an ACK, or do any combination of these (802.11

Figure 15.8 *Example of repetition interval*



uses piggybacking). At the end of the contention-free period, the PC sends a CF end (contention-free end) frame to allow the contention-based stations to use the medium.

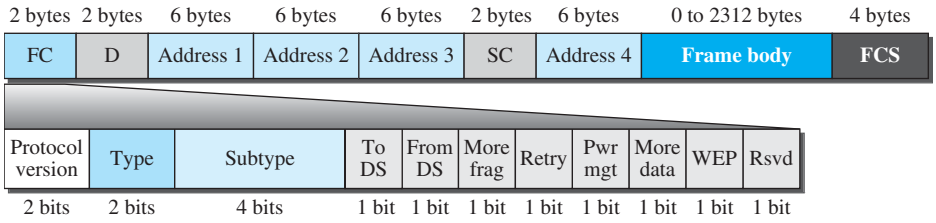
Fragmentation

The wireless environment is very noisy, so frames are often corrupted. A corrupt frame has to be retransmitted. The protocol, therefore, recommends fragmentation—the division of a large frame into smaller ones. It is more efficient to resend a small frame than a large one.

Frame Format

The MAC layer frame consists of nine fields, as shown in Figure 15.9.

Figure 15.9 *Frame format*



- ***Frame control (FC).*** The FC field is 2 bytes long and defines the type of frame and some control information. Table 15.1 describes the subfields. We will discuss each frame type later in this chapter.

Table 15.1 *Subfields in FC field*

Field	*Explanation*
Version	Current version is 0
Type	Type of information: management (00), control (01), or data (10)
Subtype	Subtype of each type (see Table 15.2)
To DS	Defined later
From DS	Defined later
More frag	When set to 1, means more fragments
Retry	When set to 1, means retransmitted frame
Pwr mgt	When set to 1, means station is in power management mode
More data	When set to 1, means station has more data to send
WEP	Wired equivalent privacy (encryption implemented)
Rsvd	Reserved

- ***D.*** This field defines the duration of the transmission that is used to set the value of NAV. In one control frame, it defines the ID of the frame.
- ***Addresses.*** There are four address fields, each 6 bytes long. The meaning of each address field depends on the value of the *To DS* and *From DS* subfields and will be discussed later.
- ***Sequence control.*** This field, often called the *SC* field, defines a 16-bit value. The first four bits define the fragment number; the last 12 bits define the sequence number, which is the same in all fragments.
- ***Frame body.*** This field, which can be between 0 and 2312 bytes, contains information based on the type and the subtype defined in the FC field.
- ***FCS.*** The FCS field is 4 bytes long and contains a CRC-32 error-detection sequence.

Frame Types

A wireless LAN defined by IEEE 802.11 has three categories of frames: management frames, control frames, and data frames.

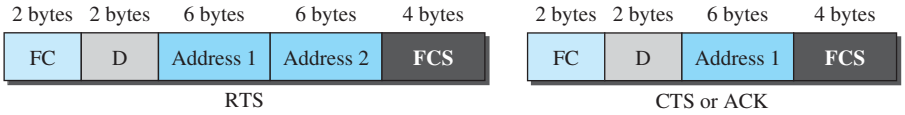
Management Frames

Management frames are used for the initial communication between stations and access points.

Control Frames

Control frames are used for accessing the channel and acknowledging frames. Figure 15.10 shows the format.

Figure 15.10 *Control frames*



For control frames the value of the type field is 01; the values of the subtype fields for frames we have discussed are shown in Table 15.2.

Table 15.2 *Values of subtype fields in control frames*

Subtype	*Meaning*
1011	Request to send (RTS)
1100	Clear to send (CTS)
1101	Acknowledgment (ACK)

Data Frames

Data frames are used for carrying data and control information.

15.2.3 Addressing Mechanism

The IEEE 802.11 addressing mechanism specifies four cases, defined by the value of the two flags in the FC field, *To DS* and *From DS*. Each flag can be either 0 or 1, resulting in four different situations. The interpretation of the four addresses (address 1 to address 4) in the MAC frame depends on the value of these flags, as shown in Table 15.3.

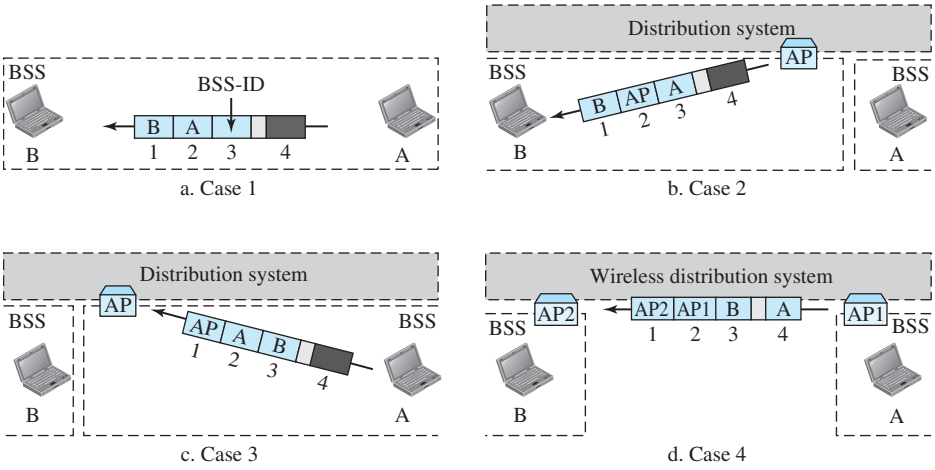
Table 15.3 *Addresses*

To DS	*From DS*	*Address 1*	*Address 2*	*Address 3*	*Address 4*
0	0	Destination	Source	BSS ID	N/A
0	1	Destination	Sending AP	Source	N/A
1	0	Receiving AP	Source	Destination	N/A
1	1	Receiving AP	Sending AP	Destination	Source

Note that address 1 is always the address of the next device that the frame will visit. Address 2 is always the address of the previous device that the frame has left. Address 3 is the address of the final destination station if it is not defined by address 1 or the original source station if it is not defined by address 2. Address 4 is the original source when the distribution system is also wireless.

- **Case 1: 00** In this case, *To DS* = 0 and *From DS* = 0. This means that the frame is not going to a distribution system (*To DS* = 0) and is not coming from a distribution system (*From DS* = 0). The frame is going from one station in a BSS to another without passing through the distribution system. The addresses are shown in Figure 15.11.
- **Case 2: 01** In this case, *To DS* = 0 and *From DS* = 1. This means that the frame is coming from a distribution system (*From DS* = 1). The frame is coming from an

Figure 15.11 *Addressing mechanisms*



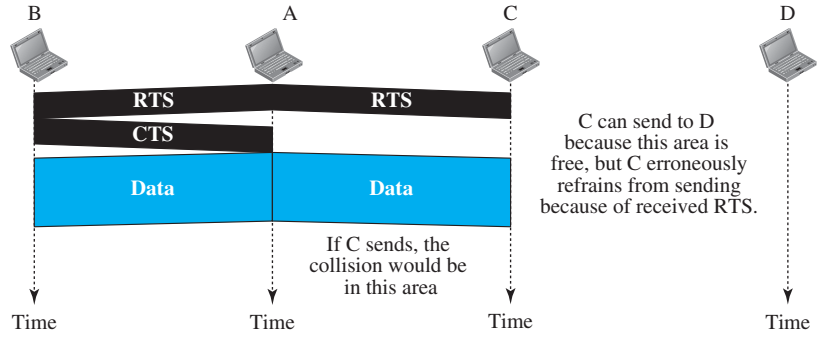
AP and going to a station. The addresses are as shown in Figure 15.11. Note that address 3 contains the original sender of the frame (in another BSS).

- **Case 3: 10** In this case, *To DS* = 1 and *From DS* = 0. This means that the frame is going to a distribution system (*To DS* = 1). The frame is going from a station to an AP. The ACK is sent to the original station. The addresses are as shown in Figure 15.11. Note that address 3 contains the final destination of the frame in the distribution system.
- **Case 4: 11** In this case, *To DS* = 1 and *From DS* = 1. This is the case in which the distribution system is also wireless. The frame is going from one AP to another AP in a wireless distribution system. Here, we need four addresses to define the original sender, the final destination, and two intermediate APs. Figure 15.11 shows the situation.

Exposed Station Problem

We discussed how to solve the hidden station problem. A similar problem is called the *exposed station problem*. In this problem a station refrains from using a channel when it is, in fact, available. In Figure 15.12, station A is transmitting to station B. Station C has some data to send to station D, which can be sent without interfering with the transmission from A to B. However, station C is exposed to transmission from A; it hears what A is sending and thus refrains from sending. In other words, C is too conservative and wastes the capacity of the channel. The handshaking messages RTS and CTS cannot help in this case. Station C hears the RTS from A and refrains from sending, even though the communication between C and D cannot cause a collision in the zone between A and C; station C cannot know that station A's transmission does not affect the zone between C and D.

Figure 15.12 *Exposed station problem*



15.2.4 Physical Layer

We discuss six specifications, as shown in Table 15.4. All implementations, except the infrared, operate in the *industrial, scientific,* and *medical (ISM)* band, which defines three unlicensed bands in the three ranges 902–928 MHz, 2.400–4.835 GHz, and 5.725–5.850 GHz.

Table 15.4 *Specifications*

IEEE	*Technique*	*Band*	*Modulation*	*Rate (Mbps)*
802.11	FHSS	2.400–4.835 GHz	FSK	1 and 2
	DSSS	2.400–4.835 GHz	PSK	1 and 2
	None	Infrared	PPM	1 and 2
802.11a	OFDM	5.725–5.850 GHz	PSK or QAM	6 to 54
802.11b	DSSS	2.400–4.835 GHz	PSK	5.5 and 11
802.11g	OFDM	2.400–4.835 GHz	Different	22 and 54
802.11n	OFDM	5.725–5.850 GHz	Different	600

IEEE 802.11 FHSS

IEEE 802.11 FHSS uses the **frequency-hopping spread spectrum (FHSS)** method, as discussed in Chapter 6. FHSS uses the 2.400–4.835 GHz ISM band. The band is divided into 79 subbands of 1 MHz (and some guard bands). A pseudorandom number generator selects the hopping sequence. The modulation technique in this specification is either two-level FSK or four-level FSK with 1 or 2 bits/baud, which results in a data rate of 1 or 2 Mbps, as shown in Figure 15.13.

IEEE 802.11 DSSS

IEEE 802.11 DSSS uses the **direct-sequence spread spectrum (DSSS)** method, as discussed in Chapter 6. DSSS uses the 2.400–4.835 GHz ISM band. The modulation technique in this specification is PSK at 1 Mbaud/s. The system allows 1 or 2 bits/baud (BPSK or QPSK), which results in a data rate of 1 or 2 Mbps, as shown in Figure 15.14.

Figure 15.13 *Physical layer of IEEE 802.11 FHSS*

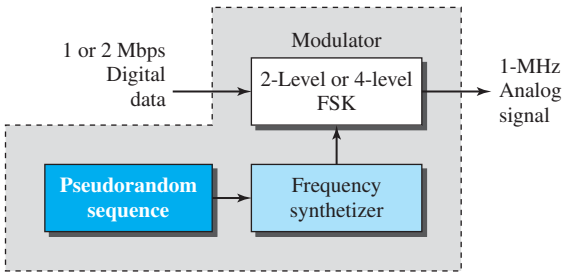
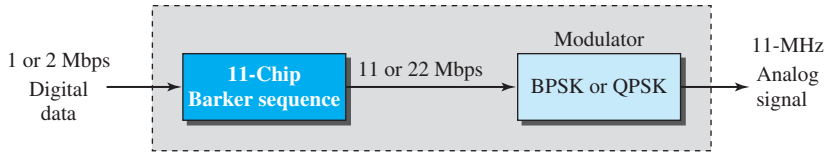


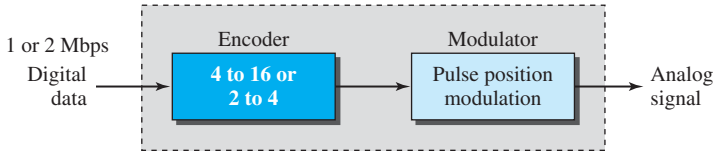
Figure 15.14 *Physical layer of IEEE 802.11 DSSS*



IEEE 802.11 Infrared

IEEE 802.11 infrared uses infrared light in the range of 800 to 950 nm. The modulation technique is called ***pulse position modulation (PPM).*** For a 1-Mbps data rate, a 4-bit sequence is first mapped into a 16-bit sequence in which only one bit is set to 1 and the rest are set to 0. For a 2-Mbps data rate, a 2-bit sequence is first mapped into a 4-bit sequence in which only one bit is set to 1 and the rest are set to 0. The mapped sequences are then converted to optical signals; the presence of light specifies 1, the absence of light specifies 0. See Figure 15.15.

Figure 15.15 *Physical layer of IEEE 802.11 infrared*



IEEE 802.11a OFDM

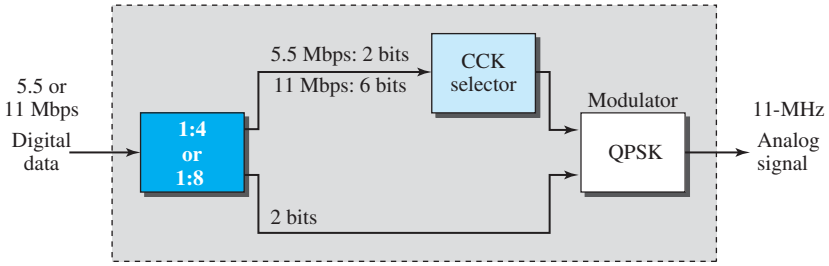
IEEE 802.11a OFDM describes the **orthogonal frequency-division multiplexing (OFDM)** method for signal generation in a 5.725–5.850 GHz ISM band. OFDM is similar to FDM, as discussed in Chapter 6, with one major difference: All the

subbands are used by one source at a given time. Sources contend with one another at the data-link layer for access. The band is divided into 52 subbands, with 48 subbands for sending 48 groups of bits at a time and 4 subbands for control information. Dividing the band into subbands diminishes the effects of interference. If the subbands are used randomly, security can also be increased. OFDM uses PSK and QAM for modulation. The common data rates are 18 Mbps (PSK) and 54 Mbps (QAM).

IEEE 802.11b DSSS

IEEE 802.11b DSSS describes the **high-rate direct-sequence spread spectrum (HR-DSSS)** method for signal generation in the 2.400–4.835 GHz ISM band. HR-DSSS is similar to DSSS except for the encoding method, which is called ***complementary code keying (CCK).*** CCK encodes 4 or 8 bits to one CCK symbol. To be backward compatible with DSSS, HR-DSSS defines four data rates: 1, 2, 5.5, and 11 Mbps. The first two use the same modulation techniques as DSSS. The 5.5-Mbps version uses BPSK and transmits at 1.375 Mbaud/s with 4-bit CCK encoding. The 11-Mbps version uses QPSK and transmits at 1.375 Mbps with 8-bit CCK encoding. Figure 15.16 shows the modulation technique for this standard.

Figure 15.16 *Physical layer of IEEE 802.11b*



IEEE 802.11g

This new specification defines forward error correction and OFDM using the 2.400–4.835 GHz ISM band. The modulation technique achieves a 22- or 54-Mbps data rate. It is backward-compatible with 802.11b, but the modulation technique is OFDM.

IEEE 802.11n

An upgrade to the 802.11 project is called 802.11n (the next generation of wireless LAN). The goal is to increase the throughput of 802.11 wireless LANs. The new standard emphasizes not only the higher bit rate but also eliminating some unnecessary overhead. The standard uses what is called ***MIMO (multiple-input multiple-output antenna)*** to overcome the noise problem in wireless LANs. The idea is that if we can send multiple output signals and receive multiple input signals, we are in a better

position to eliminate noise. Some implementations of this project have reached up to 600 Mbps data rate.

15.3 BLUETOOTH

Bluetooth is a wireless LAN technology designed to connect devices of different functions such as telephones, notebooks, computers (desktop and laptop), cameras, printers, and even coffee makers when they are at a short distance from each other. A Bluetooth LAN is an ad hoc network, which means that the network is formed spontaneously; the devices, sometimes called gadgets, find each other and make a network called a piconet. A Bluetooth LAN can even be connected to the Internet if one of the gadgets has this capability. A Bluetooth LAN, by nature, cannot be large. If there are many gadgets that try to connect, there is chaos.

Bluetooth technology has several applications. Peripheral devices such as a wireless mouse or keyboard can communicate with the computer through this technology. Monitoring devices can communicate with sensor devices in a small health care center. Home security devices can use this technology to connect different sensors to the main security controller. Conference attendees can synchronize their laptop computers at a conference.

Bluetooth was originally started as a project by the Ericsson Company. It is named for Harald Blaatand, the king of Denmark (940-981) who united Denmark and Norway. *Blaatand* translates to *Bluetooth* in English.

Today, Bluetooth technology is the implementation of a protocol defined by the IEEE 802.15 standard. The standard defines a wireless personal-area network (PAN) operable in an area the size of a room or a hall.

15.3.1 Architecture

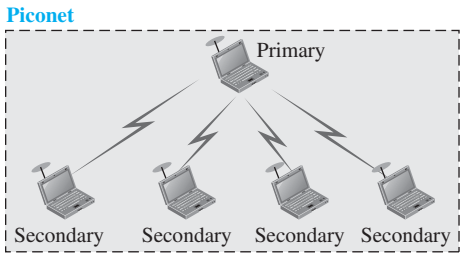
Bluetooth defines two types of networks: piconet and scatternet.

Piconets

A Bluetooth network is called a ***piconet,*** or a small net. A piconet can have up to eight stations, one of which is called the *primary;* the rest are called *secondaries.* All the secondary stations synchronize their clocks and hopping sequence with the primary. Note that a piconet can have only one primary station. The communication between the primary and secondary stations can be one-to-one or one-to-many. Figure 15.17 shows a piconet.

Although a piconet can have a maximum of seven secondaries, additional secondaries can be in the *parked state*. A secondary in a parked state is synchronized with the primary, but cannot take part in communication until it is moved from the parked state to the active state. Because only eight stations can be active in a piconet, activating a station from the parked state means that an active station must go to the parked state.

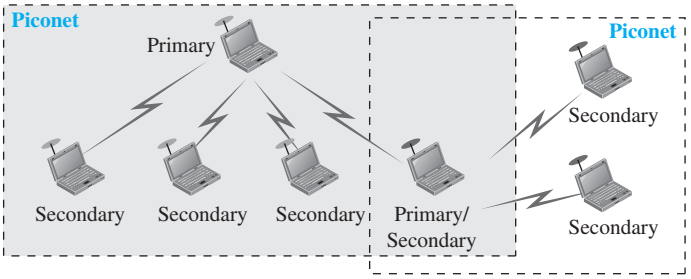
Figure 15.17 *Piconet*



Scatternet

Piconets can be combined to form what is called a ***scatternet.*** A secondary station in one piconet can be the primary in another piconet. This station can receive messages from the primary in the first piconet (as a secondary) and, acting as a primary, deliver them to secondaries in the second piconet. A station can be a member of two piconets. Figure 15.18 illustrates a scatternet.

Figure 15.18 *Scatternet*



Bluetooth Devices

A Bluetooth device has a built-in short-range radio transmitter. The current data rate is 1 Mbps with a 2.4-GHz bandwidth. This means that there is a possibility of interference between the IEEE 802.11b wireless LANs and Bluetooth LANs.

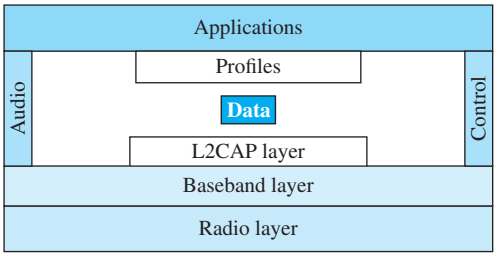
15.3.2 Bluetooth Layers

Bluetooth uses several layers that do not exactly match those of the Internet model we have defined in this book. Figure 15.19 shows these layers.

L2CAP

The **Logical Link Control and Adaptation Protocol,** or **L2CAP** (L2 here means LL), is roughly equivalent to the LLC sublayer in LANs. It is used for data exchange on an

Figure 15.19 *Bluetooth layers*



ACL link; SCO channels do not use L2CAP. Figure 15.20 shows the format of the data packet at this level.

Figure 15.20 *L2CAP data packet format*

The 16-bit length field defines the size of the data, in bytes, coming from the upper layers. Data can be up to 65,535 bytes. The channel ID (CID) defines a unique identifier for the virtual channel created at this level (see below).

The L2CAP has specific duties: multiplexing, segmentation and reassembly, quality of service (QoS), and group management.

Multiplexing

The L2CAP can do multiplexing. At the sender site, it accepts data from one of the upper-layer protocols, frames them, and delivers them to the baseband layer. At the receiver site, it accepts a frame from the baseband layer, extracts the data, and delivers them to the appropriate protocol layer. It creates a kind of virtual channel that we will discuss in later chapters on higher-level protocols.

Segmentation and Reassembly

The maximum size of the payload field in the baseband layer is 2774 bits, or 343 bytes. This includes 4 bytes to define the packet and packet length. Therefore, the size of the packet that can arrive from an upper layer can only be 339 bytes. However, application layers sometimes need to send a data packet that can be up to 65,535 bytes (an Internet packet, for example). The L2CAP divides these large packets into segments and adds extra information to define the location of the segments in the original packet. The L2CAP segments the packets at the source and reassembles them at the destination.

QoS

Bluetooth allows the stations to define a quality-of-service level. We discuss quality of service in Chapter 30. For the moment, it is sufficient to know that if no quality-of-service

level is defined, Bluetooth defaults to what is called *best-effort* service; it will do its best under the circumstances.

Group Management

Another functionality of L2CAP is to allow devices to create a type of logical addressing between themselves. This is similar to multicasting. For example, two or three secondary devices can be part of a multicast group to receive data from the primary.

Baseband Layer

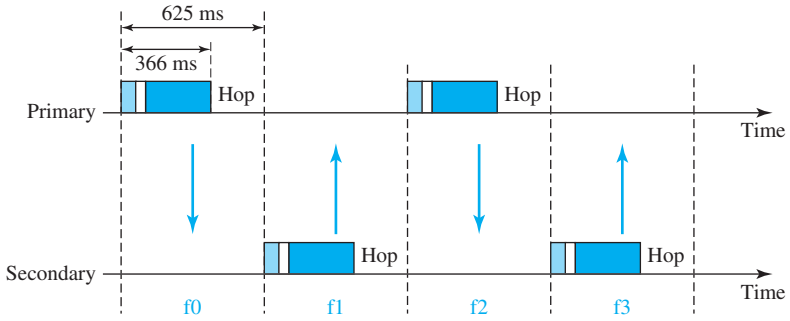
The baseband layer is roughly equivalent to the MAC sublayer in LANs. The access method is TDMA (discussed later). The primary and secondary stations communicate with each other using time slots. The length of a time slot is exactly the same as the dwell time, 625 μs. This means that during the time that one frequency is used, a primary sends a frame to a secondary, or a secondary sends a frame to the primary. Note that the communication is only between the primary and a secondary; secondaries cannot communicate directly with one another.

TDMA

Bluetooth uses a form of TDMA that is called ***TDD-TDMA (time-division duplex TDMA).*** TDD-TDMA is a kind of half-duplex communication in which the sender and receiver send and receive data, but not at the same time (half-duplex); however, the communication for each direction uses different hops. This is similar to walkie-talkies using different carrier frequencies.

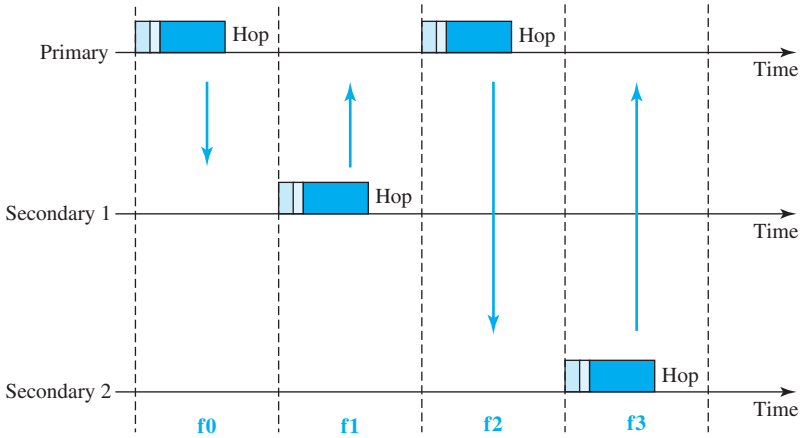
- ***Single-Secondary Communication*** If the piconet has only one secondary, the TDMA operation is very simple. The time is divided into slots of 625 μs. The primary uses even-numbered slots (0, 2, 4, . . .); the secondary uses odd-numbered slots (1, 3, 5, . . .). TDD-TDMA allows the primary and the secondary to communicate in half-duplex mode. In slot 0, the primary sends and the secondary receives; in slot 1, the secondary sends and the primary receives. The cycle is repeated. Figure 15.21 shows the concept.

Figure 15.21 *Single-secondary communication*



- ***Multiple-Secondary Communication*** The process is a little more involved if there is more than one secondary in the piconet. Again, the primary uses the even-numbered slots, but a secondary sends in the next odd-numbered slot if the packet in the previous slot was addressed to it. All secondaries listen on even-numbered slots, but only one secondary sends in any odd-numbered slot. Figure 15.22 shows a scenario.

Figure 15.22 *Multiple-secondary communication*



Let us elaborate on the figure.

1. In slot 0, the primary sends a frame to secondary 1.
2. In slot 1, only secondary 1 sends a frame to the primary because the previous frame was addressed to secondary 1; other secondaries are silent.
3. In slot 2, the primary sends a frame to secondary 2.
4. In slot 3, only secondary 2 sends a frame to the primary because the previous frame was addressed to secondary 2; other secondaries are silent.
5. The cycle continues.

We can say that this access method is similar to a poll/select operation with reservations. When the primary selects a secondary, it also polls it. The next time slot is reserved for the polled station to send its frame. If the polled secondary has no frame to send, the channel is silent.

Links

Two types of links can be created between a primary and a secondary: SCO links and ACL links.

- ***SCO*** A **synchronous connection-oriented (SCO) link** is used when avoiding latency (delay in data delivery) is more important than integrity (error-free delivery). In an SCO link, a physical link is created between the primary and a secondary by reserving specific slots at regular intervals. The basic unit of connection is two slots, one for each direction. If a packet is damaged, it is never retransmitted.

SCO is used for real-time audio where avoiding delay is all-important. A secondary can create up to three SCO links with the primary, sending digitized audio (PCM) at 64 kbps in each link.

- ***ACL*** An **asynchronous connectionless link (ACL)** is used when data integrity is more important than avoiding latency. In this type of link, if a payload encapsulated in the frame is corrupted, it is retransmitted. A secondary returns an ACL frame in the available odd-numbered slot if the previous slot has been addressed to it. ACL can use one, three, or more slots and can achieve a maximum data rate of 721 kbps.

Frame Format

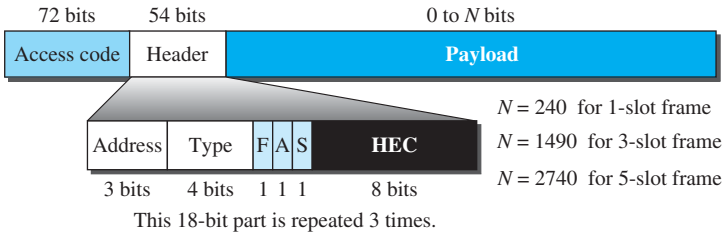
A frame in the baseband layer can be one of three types: one-slot, three-slot, or five-slot. A slot, as we said before, is 625 μs. However, in a one-slot frame exchange, 259 μs is needed for hopping and control mechanisms. This means that a one-slot frame can last only 625 – 259, or 366 μs. With a 1-MHz bandwidth and 1 bit/Hz, the size of a one-slot frame is 366 bits.

A three-slot frame occupies three slots. However, since 259 μs is used for hopping, the length of the frame is $3 \times 625 - 259 = 1616$ μs or 1616 bits. A device that uses a three-slot frame remains at the same hop (at the same carrier frequency) for three slots. Even though only one hop number is used, three hop numbers are consumed. That means the hop number for each frame is equal to the first slot of the frame.

A five-slot frame also uses 259 bits for hopping, which means that the length of the frame is $5 \times 625 - 259 = 2866$ bits.

Figure 15.23 shows the format of the three frame types.

Figure 15.23 *Frame format types*



The following describes each field:

- ***Access code.*** This 72-bit field normally contains synchronization bits and the identifier of the primary to distinguish the frame of one piconet from that of another.
- ***Header.*** This 54-bit field is a repeated 18-bit pattern. Each pattern has the following subfields:

a. ***Address.*** The 3-bit address subfield can define up to seven secondaries (1 to 7). If the address is zero, it is used for broadcast communication from the primary to all secondaries.
b. ***Type.*** The 4-bit type subfield defines the type of data coming from the upper layers. We discuss these types later.
c. ***F.*** This 1-bit subfield is for flow control. When set (1), it indicates that the device is unable to receive more frames (buffer is full).
d. ***A.*** This 1-bit subfield is for acknowledgment. Bluetooth uses Stop-and-Wait ARQ; 1 bit is sufficient for acknowledgment.
e. ***S.*** This 1-bit subfield holds a sequence number. Bluetooth uses Stop-and-Wait ARQ; 1 bit is sufficient for sequence numbering.
f. ***HEC.*** The 8-bit header error correction subfield is a checksum to detect errors in each 18-bit header section. The header has three identical 18-bit sections. The receiver compares these three sections, bit by bit. If each of the corresponding bits is the same, the bit is accepted; if not, the majority opinion rules. This is a form of forward error correction (for the header only). This double error control is needed because the nature of the communication, via air, is very noisy. Note that there is no retransmission in this sublayer.

- ***Payload.*** This subfield can be 0 to 2740 bits long. It contains data or control information coming from the upper layers.

Radio Layer

The radio layer is roughly equivalent to the physical layer of the Internet model. Bluetooth devices are low-power and have a range of 10 m.

Band

Bluetooth uses a 2.4-GHz ISM band divided into 79 channels of 1 MHz each.

FHSS

Bluetooth uses the **frequency-hopping spread spectrum (FHSS)** method in the physical layer to avoid interference from other devices or other networks. Bluetooth hops 1600 times per second, which means that each device changes its modulation frequency 1600 times per second. A device uses a frequency for only 625 μs (1/1600 s) before it hops to another frequency; the dwell time is 625 μs.

Modulation

To transform bits to a signal, Bluetooth uses a sophisticated version of FSK, called GFSK (FSK with Gaussian bandwidth filtering; a discussion of this topic is beyond the scope of this book). GFSK has a carrier frequency. Bit 1 is represented by a frequency deviation above the carrier; bit 0 is represented by a frequency deviation below the carrier. The frequencies, in megahertz, are defined according to the following formula for each channel.

$$f_c = 2402 + n \text{ MHz} \qquad n = 0, 1, 2, 3, \ldots, 78$$

For example, the first channel uses carrier frequency 2402 MHz (2.402 GHz), and the second channel uses carrier frequency 2403 MHz (2.403 GHz).

15.4 END-CHAPTER MATERIALS

15.4.1 Further Reading

For more details about subjects discussed in this chapter, we recommend the following books. The items in brackets [. . .] refer to the reference list at the end of the text.

Books

Several books cover materials discussed in this chapter, including [Sch 03], [Gas 02], [For 03], [Sta 04], [Sta 02], [Kei 02], [Jam 03], [AZ 03], [Tan 03], [Cou 01], [Com 06], [GW 04], and [PD 03].

15.4.2 Key Terms

access point (AP)
ad hoc network
asynchronous connectionless link (ACL)
basic service set (BSS)
beacon frame
Bluetooth
BSS-transition mobility
complementary code keying (CCK)
direct sequence spread spectrum (DSSS)
distributed coordination function (DCF)
distributed interframe space (DIFS)
ESS-transition mobility
extended service set (ESS)
frequency-hopping spread spectrum (FHSS)
high-rate direct-sequence spread spectrum (HR-DSSS)
Logical Link Control and Adaptation Protocol (L2CAP)
multiple-input multiple-output (MIMO) antenna
network allocation vector (NAV)
no transition mobility
orthogonal frequency-division multiplexing (OFDM)
piconet
point coordination function (PCF)
pulse position modulation (PPM)
scatternet
short interframe space (SIFS)
signal to noise ration (SNR)
synchronous connection-oriented (SCO) link
TDD-TDMA (time-division duplex TDMA)

15.4.3 Summary

The nature and characteristics of a wireless network are different from those of a wired network. There are some issues in a wireless network that are negligible in a wired network. Access control in a wireless LAN is also different from that in a wired LAN because of some issues such as the hidden station problem.

Wireless LANs became formalized with the IEEE 802.11 standard, which defines two services: basic service set (BSS) and extended service set (ESS). The access method used in the distributed coordination function (DCF) MAC sublayer is CSMA/CA. The access method used in the point coordination function (PCF) MAC sublayer is polling. A frame in this network carries four addresses to define the original and previous sources and immediate and final destinations. There are other frame types in this network to handle access control and data transfer.

Bluetooth is a wireless LAN technology that connects devices (called gadgets) in a small area. A Bluetooth network is called a piconet. Piconets can be combined to form what is called a scatternet. Bluetooth uses several layers that do not exactly match those of the Internet model we have defined in this book. L2CAP is roughly equivalent to the LLC sublayer in LANs. The baseband layer is roughly equivalent to the MAC sublayer in LANs. The access method is TDMA.

15.5 PRACTICE SET

15.5.1 Quizzes

A set of interactive quizzes for this chapter can be found on the book website. It is strongly recommended that the student take the quizzes to check his/her understanding of the materials before continuing with the practice set.

15.5.2 Questions

Q15-1. Compare the medium of a wired LAN with that of a wireless LAN in today's communication environment.

Q15-2. Explain why the MAC protocol is more important in wireless LANs than in wired LANs.

Q15-3. Explain why there is more attenuation in a wireless LAN than in a wired LAN, ignoring the noise and the interference.

Q15-4. Why is SNR in a wireless LAN normally lower than SNR in a wired LAN?

Q15-5. What is multipath propagation? What is its effect on wireless networks?

Q15-6. What are some reasons that CSMA/CD cannot be used in a wireless LAN?

Q15-7. Explain why fragmentation is recommended in a wireless LAN.

Q15-8. Explain why we have only one frame type in a wired LAN, but four frame types in a wireless LAN.

Q15-9. Do the MAC addresses used in an 802.3 (Wired Ethernet) and the MAC addresses used in an 802.11 (Wireless Ethernet) belong to two different address spaces?

Q15-10. An AP may connect a wireless network to a wired network. Does the AP need to have two MAC addresses in this case?

Q15-11. An AP in a wireless network plays the same role as a link-layer switch in a wired network. However, a link-layer switch has no MAC address, but an AP normally needs a MAC address. Explain the reason.

Q15-12. What is the reason that Bluetooth is normally called a wireless personal area network (WPAN) instead of a wireless local area network (WLAN)?

Q15-13. Compare a piconet and a scatternet in the Bluetooth architecture.

Q15-14. Can a piconet have more than eight stations? Explain.

Q15-15. What is the actual bandwidth used for communication in a Bluetooth network?

Q15-16. What is the role of the *radio* layer in Bluetooth?

Q15-17. Fill in the blanks. The 83.5 MHz bandwidth in Bluetooth is divided into_____ channels, each of _____ MHz.

Q15-18. What is the spread spectrum technique used by Bluetooth?

Q15-19. What is the modulation technique in the radio layer of Bluetooth? In other words, how are digital data (bits) changed to analog signals (radio waves)?

Q15-20. What MAC protocol is used in the baseband layer of Bluetooth?

Q15-21. What is the role of the *L2CAP* layer in Bluetooth?

15.5.3 Problems

P15-1. In an 802.11, give the value of the address 1 field in each of the following situations (left bit defines *To DS* and right bit defines *From DS*).

a. 00 **b.** 01 **c.** 10 **d.** 11

P15-2. In an 802.11, give the value of the address 2 field in each of the following situations (left bit defines *To DS* and right bit defines *From DS*).

a. 00 **b.** 01 **c.** 10 **d.** 11

P15-3. In an 802.11, give the value of the address 3 field in each of the following situations (left bit defines *To DS* and right bit defines *From DS*).

a. 00 **b.** 01 **c.** 10 **d.** 11

P15-4. In an 802.11, give the value of the address 4 field in each of the following situations (left bit defines *To DS* and right bit defines *From DS*).

a. 00 **b.** 01 **c.** 10 **d.** 11

P15-5. In a BSS with no AP (ad hoc network), we have five stations: A, B, C, D, and E. Station A needs to send a message to station B. Answer the following questions for the situation where the network is using the DCF protocol:

a. What are the values of the *To DS* and *From DS* bits in the frames exchanged?

b. Which station sends the RTS frame and what is (are) the value(s) of the address field(s) in this frame?

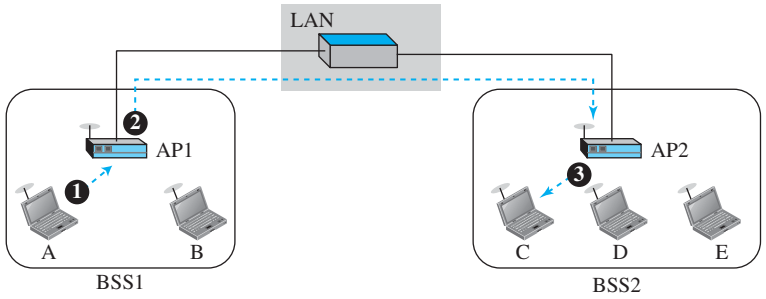
c. Which station sends the CTS frame and what is (are) the value(s) of the address field(s) in this frame?

d. Which station sends the data frame and what is (are) the value(s) of the address field(s) in this frame?

e. Which station sends the ACK frame and what is (are) the value(s) of the address field(s) in this frame?

P15-6. In Figure 15.24, two wireless networks, BSS1 and BSS2, are connected through a wired distribution system (DS), an Ethernet LAN. Assume station A

Figure 15.24 *Problem P15-6*



in BSS1 needs to send a data frame to station C in BSS2. Show the value of addresses in 802.11 and 802.3 frames for three transmissions: from station A to AP1, from AP1 to AP2, and from AP2 to station C. Note that the communication between AP1 and AP2 occurs in a wired environment.

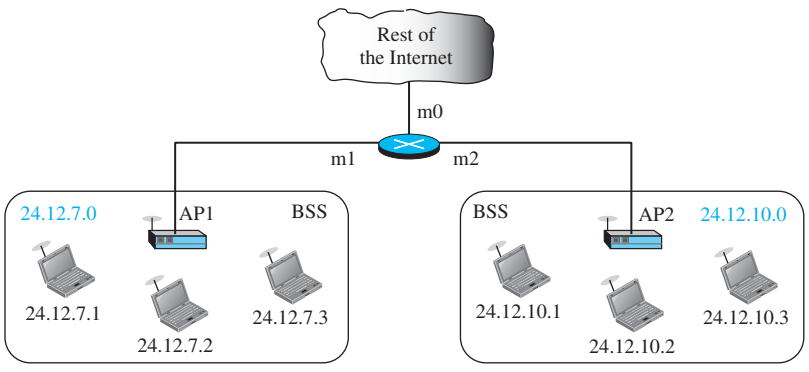
P15-7. Repeat the previous problem (Figure 15.24), but assume that the distribution system is also wireless. AP1 is connected to AP2 through a wireless channel. Show the value of addresses in all communication sections: from station A to AP1, from AP1 to AP2, and from AP2 to station C.

P15-8. Assume that a frame moves from a wired network using the 802.3 protocol to a wireless network using the 802.11 protocol. Show how the field values in the 802.11 frame are filled with the values of the 802.3 frame. Assume that the transformation occurs at the AP that is on the boundary between the two networks.

P15-9. Assume a frame moves from a wireless network using the 802.11 protocol to a wired network using the 802.3 protocol. Show how the field values in the 802.3 frame are filled with the values of the 802.11 frame. Assume that the transformation occurs at the AP that is on the boundary between the two networks.

P15-10. Assume two 802.11 wireless networks are connected to the rest of the Internet via a router as shown in Figure 15.25. The router has received an IP datagram with the destination IP address 24.12.7.1 and needs to send it to the corresponding wireless host. Explain the process and describe how the values of address 1, address 2, address 3, and address 4 (see Figure 15.25) are determined in this case.

Figure 15.25 *Problem P15-10*



P15-11. In Figure 15.25 (previous problem), assume that the host with IP address 24.12.10.3 needs to send an IP datagram to the host with IP address 128.41.23.12 somewhere in the world (not shown in the figure). Explain the process and show how the values of address 1, address 2, address 3, and address 4 (see Figure 15.25) are determined in this case.

P15-12. A BSS ID (BSSID) is a 48-bit address assigned to a BSS in an 802.11 network. Do some research and find what the use of the BSSID is and how BSSIDs are assigned in ad hoc and infrastructure networks.

P15-13. Do some research and find out how flow and error control are accomplished in an 802.11 network using the DCF MAC sublayer.

P15-14. In an 802.11 communication, the size of the payload (frame body) is 1200 bytes. The station decides to fragment the frame into three fragments, each of 400 payload bytes. Answer the following questions:

- **a.** What would be the size of the data frame if no fragmentations were done?
- **b.** What is the size of each frame after fragmentation?
- **c.** How many total bytes are sent after fragmentation (ignoring the extra control frames)?
- **d.** How many extra bytes are sent because of fragmentation (again ignoring extra control frames)?

P15-15. Both the IP protocol and the 802.11 project fragment their packets. IP fragments a datagram at the network layer; 802.11 fragments a frame at the data-link layer. Compare and contrast the two fragmentation schemes, using the different fields and subfields used in each protocol.

P15-16. In an 802.11 network, assume station A has four fragments to send to station B. If the sequence number of the first fragment is selected as 3273, what are the values of the more fragment flag, fragment number, and sequence number?

P15-17. In an 802.11 network, station A sends one data frame (not fragmented) to station B. What would be the value of the D field (in microseconds) that needs to be set for the NAV period in each of the following frames: RTS, CTS, data, and ACK? Assume that the transmission time for RTS, CTS, and ACK is 4 μs each. The transmission time for the data frame is 40 μs and the SIFS duration is set to 1 μs. Ignore the propagation time. Note that each frame needs to set the duration of NAV for the rest of the time the medium needs to be reserved to complete the transaction.

P15-18. In an 802.11 network, station A sends two data fragments to station B. What would be the value of the D field (in microseconds) that needs to be set for the NAV period in each of the following frames: RTS, CTS, data, and ACK? Assume that the transmission time for RTS, CTS, and ACK is 4 μs each. The transmission time for each fragment is 20 μs and the SIFS duration is set to 1 μs. Ignore the propagation time. Note that each frame needs to set the duration of NAV for the rest of the time the medium needs to be reserved to complete the transaction.

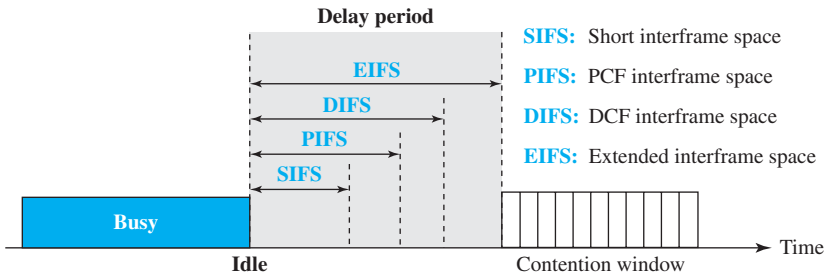
P15-19. In an 802.11 network, three stations (A, B, and C) are contending to access the medium. The contention window for each station has 31 slots. Station A randomly picks up the first slot; station B picks up the fifth slot; and station C picks up the twenty-first slot. Show the procedure each station should follow.

P15-20. In an 802.11 network, there are three stations, A, B, and C. Station C is hidden from A, but can be seen (electronically) by B. Now assume that station A

needs to send data to station B. Since C is hidden from A, the RTS frame cannot reach C. Explain how station C can find out that the channel is locked by A and that it should refrain from transmitting.

P15-21. An 802.11 network may use four different interframe spaces (IFSs) to delay the transmission of a frame in different situations. This allows low-priority traffic to wait for high-priority traffic when the channel becomes idle. Normally, four different IFSs are used in different implementations, as shown in Figure 15.26. Explain the purpose of these IFSs (you may need to do some research using the Internet).

Figure 15.26 *Problem P15-21*



P15-22. Although an RTS frame defines the value of time that NAV can be effective for the rest of the session, why does the 802.11 project define that other frames used in the session should redefine the rest of the period for NAV?

P15-23. Figure 15.23 shows the frame format of the baseband layer in Bluetooth (802.15). Based on this format, answer the following questions:

a. What is the range of the address domain in a Bluetooth network?

b. How many stations can be active at the same time in a piconet based on the information in the above figure?

15.6 SIMULATION EXPERIMENTS

15.6.1 Applets

We have created some Java applets to show some of the main concepts discussed in this chapter. It is strongly recommended that the students activate these applets on the book website and carefully examine the protocols in action.

15.6.2 Lab Assignments

Use the book website to find how you can use a simulator for wireless LANs.

Lab15-1. In this lab we capture and study wireless frames that are exchanged between a wireless host and the access point. See the book website for a detailed description of this lab.

https://hemanthrajhemu.github.io

CHAPTER 16

Other Wireless Networks

We discussed wired LANs in Chapter 13 and other wired networks (access networks and wide area networks) in Chapter 14. We discussed wireless LANs in Chapter 15. It is time now to discuss other wireless networks (access networks and wide area networks) in this chapter.

This chapter is divided into three sections:

- The first section discusses the WiMAX, a wireless access network that can replace the wired access networks we discussed in Chapter 14. The section first describes services provided by this network. It then describes the IEEE 802.16 project as the basis of the network. The section finally defines the link-layer and the physical layer of WiMAX.
- The second section discusses cellular telephone networks. It explains the frequency reuse principle. It then describes the general operations of this network. The next four sections each discuss one of the four generations of the cellular telephony network.
- The third section discusses satellite networks. It first describes the operations of the all types of satellites. The section then defines GEO satellites and their characteristics, then moves to MEO satellites and shows their applications. Finally, the section discusses the LEO satellites and their features and applications.

16.1 WiMAX

We first need to discuss a wireless access network. In Chapter 14, we discussed that the telephone and cable companies provide wired access to the Internet for homes and offices. Today, the tendency is to move to wireless technology for this purpose. There are two reasons for this tendency. First, people want to have access to the Internet from home or office (fixed) where the wired access to the Internet is either not available or is expensive. Second, people need to access the Internet when they are using their cellular phones (mobiles). The **Worldwide Interoperability for Microwave Access (WiMAX)** has been designed for these types of applications. It provides the "last mile" broadband wireless access.

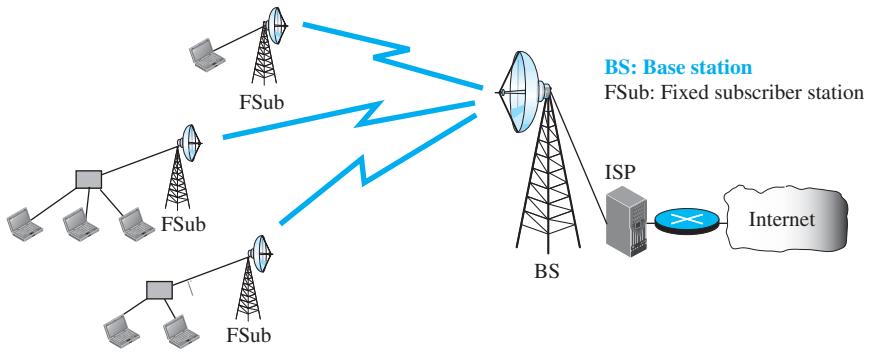
16.1.1 Services

WiMAX provides two types of services to subscribers: fixed and mobile.

Fixed WiMAX

Figure 16.1 shows the idea behind a fixed service. A base station can use different types of antenna (omnidirectional, sector, or panel) to optimize the performance. WiMAX uses a beamsteering **adaptive antenna system (AAS).** While transmitting, it can focus its energy in the direction of the subscriber; while receiving, it can focus in the direction of the subscriber station to receive maximum energy sent by the subscriber.

Figure 16.1 *Fixed WiMAX*

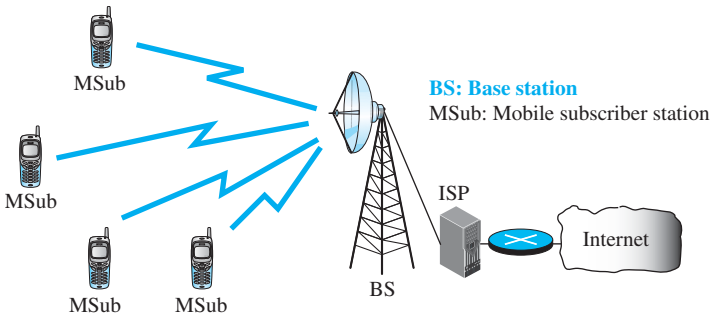


The fixed service can be compared with the service provided by the telephone and the network companies using wired connections. WiMAX also uses a MIMO antenna system, which can provide simultaneous transmitting and receiving.

Mobile WiMAX

Figure 16.2 shows the idea behind mobile service. It is the same as fixed service except the subscribers are mobile stations that move from one place to another. The same issues involved in the cellular telephone system, such as roaming, are present here.

Figure 16.2 *Mobile WiMAX*



16.1.2 IEEE Project 802.16

WiMAX is the result of the IEEE 802.16 project, which was an effort to standardize the proprietary broadband wireless system in 2002. The standard is sometimes referred to as *wireless local loop*, in contrast with wired local loop (dial-up, DLS, or cable). Before we discuss this standard, let us compare the 802.16 and 802.11 projects. First, 802.11 is a standard for a wireless LAN; 802.16 is a standard for a wireless WAN (or MAN). The distance between a base station and a host in the first is very limited; the base station and subscriber station in the second may be separated by tens of kilometers. Project 802.11 defines a connectionless communication; project 802.16 defines a connection-oriented service.

A later revision of IEEE 802.16 created two new standards called IEEE 802.16d, which concentrates on the fixed WiMAX, and IEEE 802.16e, which defines the mobile WiMAX. The two new standards do not change the main idea behind the original 802.16, but concentrate on the nature of two services.

16.1.3 Layers in Project 802.16

Figure 16.3 shows the layers in the 802.16 project. IEEE has divided the data-link layer into three sublayers and the physical layer into two sublayers.

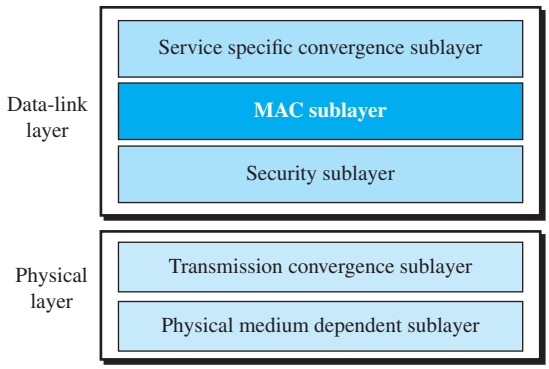
Service Specific Convergence Sublayer

This is actually the DLC sublayer revised for broadband wireless communication. It has been devised for a connection-oriented service in which each connection may benefit from a specific quality of service (QoS), as discussed in Chapter 30.

MAC Sublayer

The MAC sublayer defines the access method and the format of the frame. It is a sublayer designed for connection-oriented service. The packets are routed from the base station to the subscriber station using a connection identifier which is the same during the duration of the communication. For more about connection-oriented communication, see Chapters 8 and 18.

Figure 16.3 *Data-link and physical layers*



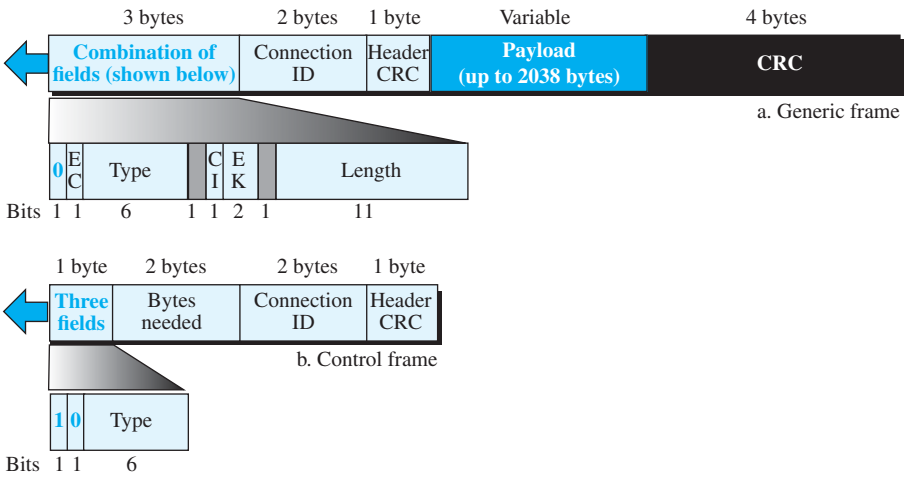
Access Method

WiMAX uses the reservation (scheduling) access method we discussed in Chapter 12 (Figure 12.18). The base station needs to make a slot reservation before sending a slot-size of data to a subscriber station; each subscriber station needs to make a reservation before sending a slot-size of data to the base station.

Frame Format

The frame format is shown in Figure 16.4. We distinguish two types of frames: generic and control. The first is used to send and receive payload; the second is used only during the connection establishment. Both frame types use a 6-byte generic header. However, some bytes have different interpretations in different frame types.

Figure 16.4 *WiMAX MAC frame format*



A brief purpose of each field follows.

- The first bit in a frame is the frame identifier. If it is 0, the frame is a generic frame; if it is 1, it is a control frame.
- ***EC.*** The *encryption control* field uses one bit to define whether the frame should be encrypted for security purpose (See Chapters 31 and 32). If the bit is 0, it means no encryption; if it is 1, it means the frame needs to be encrypted at the *security sublayer,* described later.
- ***Type.*** The *type* field uses six bits to define the type of the frame. This field is only present in the generic frame and normally is used to define the type of the payload. The payload can be a packed load, a fragmented load, and so on.
- ***CI.*** The *checksum ID* field uses one bit to define whether the frame checksum field should be present or not. If the payload is multimedia, forward error correction is applied (at the physical layer) to the frame and there is no need for checksum.
- ***EK.*** The *encryption key* field uses two bits to define one of the four keys for encryption if encryption is required (see EC field).
- ***Length.*** The *length* field uses eleven bits to define the total length of the frame. Note that this field is present in a *generic* frame and is replaced by the *bytes needed* field in the control frame.
- ***Bytes Needed.*** The *bytes needed* field uses sixteen bits to define the number of bytes needed for allocated slots in the physical layer.
- ***Connection ID.*** The *connection ID* field uses sixteen bits to define the connection identifier for the current connection. Note that the IEEE 802.16 and WiMAX define a connection-oriented protocol, as we discussed before.
- ***Header CRC.*** Both types of frames need to have an 8-bit header CRC field. The header CRC is used to check whether the header itself is corrupted. It uses the polynomial $(x^8 + x^2 + x + 1)$ as the divisor.
- ***Payload.*** This variable-length field defines the payload, the data that is encapsulated in the frame from the *service specific convergence* sublayer. The field is not needed in the control frame.
- ***CRC.*** The last field, if present, is used for error detection over the whole frame. It uses the same divisor discussed for the Ethernet.

Addressing

Each subscriber and base station typically has a 48-bit MAC address as defined in Chapter 9 (link-layer addressing) because each station is a node in the global Internet. However, there is no source or destination address field in Figure 16.4. The reason is that the combination of source and destination addresses are mapped to a *connection identifier* during the connection-establishing phase. This protocol is a connection-oriented protocol that uses a connection identifier or *virtual connection identifier* (VCI) as we discussed in Chapters 8 and 18. Each frame then uses the same connection identifier for the duration of data transfer.

Security Sublayer

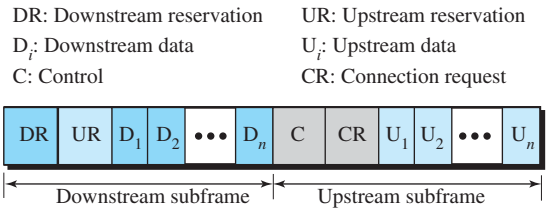
The last sublayer in the data-link layer provides security for communication using WiMAX. The nature of broadband wireless communication requires security to provide encryption for the information exchanged between a subscriber station and the base station. We discuss these issues in Chapters 31 and 32 in detail.

Transmission Convergence Sublayer

The transmission convergence sublayer uses TDD (time-division duplex), a variation of time-division multiplexing, discussed in Chapter 6, designed for duplex (bidirectional) communication. The physical layer packs the frames received from the data-link layer into two subframes at the physical layer. However, we need to distinguish between the term *frame* at the physical layer and the one at the data-link layer. A data-link layer frame in the network layer may be encapsulated in several slots in the corresponding subframe at the physical layer.

Each frame at the physical layer is made of two subframes that carry data from the base station to the subscribers (downstream) and from the subscribers to the base station (upstream). Each subframe is divided into slots as shown in Figure 16.5.

Figure 16.5 *WiMAX frame structure at the physical layer*



Physical Medium Dependent Sublayer

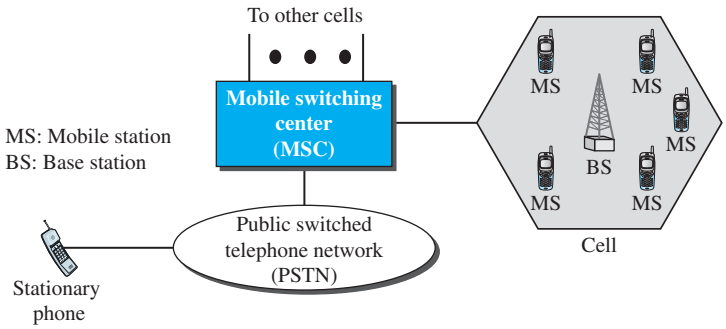
This sublayer is in continuous revision. Originally, 802.16 defined the band 10-66 GHz and modulations QPSK (2 bits/baud), QAM-16 (4 bits/baud), and QAM-64 (6 bits/baud) for long-, medium-, and short-distance communication from the base station. Later IEEE defined 802.16d (fixed WiMAX), which added the band 2-11 GHz (compatible with wireless LANs) using the orthogonal frequency-division multiplexing (OFDM). Sometime later, IEEE defined 802.16e (mobile WiMAX) and added scalable orthogonal frequency-division multiplexing (SOFDM).

16.2 CELLULAR TELEPHONY

Cellular telephony is designed to provide communications between two moving units, called *mobile stations (MSs),* or between one mobile unit and one stationary unit, often called a *land unit*. A service provider must be able to locate and track a caller, assign a channel to the call, and transfer the channel from base station to base station as the caller moves out of range.

To make this tracking possible, each cellular service area is divided into small regions called *cells*. Each cell contains an antenna and is controlled by a solar- or AC-powered network station, called the *base station* (BS). Each base station, in turn, is controlled by a switching office, called a ***mobile switching center (MSC).*** The MSC coordinates communication between all the base stations and the telephone central office. It is a computerized center that is responsible for connecting calls, recording call information, and billing (see Figure 16.6).

Figure 16.6 *Cellular system*



Cell size is not fixed and can be increased or decreased depending on the population of the area. The typical radius of a cell is 1 to 12 mi. High-density areas require more, geographically smaller cells to meet traffic demands than do low-density areas. Once determined, cell size is optimized to prevent the interference of adjacent cell signals. The transmission power of each cell is kept low to prevent its signal from interfering with those of other cells.

16.2.1 Operation

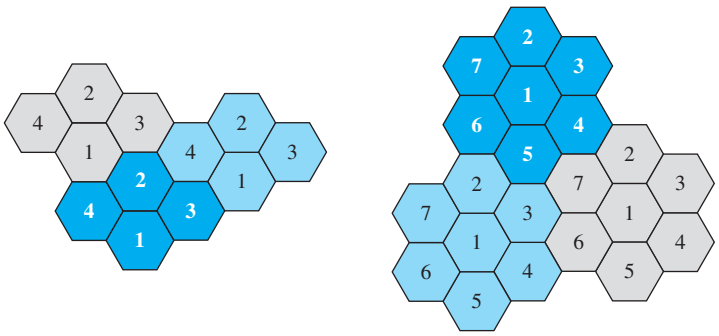
Let us first briefly discuss the operation of the cellular telephony.

Frequency-Reuse Principle

In general, neighboring cells cannot use the same set of frequencies for communication because doing so may create interference for the users located near the cell boundaries. However, the set of frequencies available is limited, and frequencies need to be reused. A frequency reuse pattern is a configuration of N cells, N being the **reuse factor,** in which each cell uses a unique set of frequencies. When the pattern is repeated, the frequencies can be reused. There are several different patterns. Figure 16.7 shows two of them.

The numbers in the cells define the pattern. The cells with the same number in a pattern can use the same set of frequencies. We call these cells the *reusing cells*. As Figure 16.7 shows, in a pattern with reuse factor 4, only one cell separates the cells using the same set of frequencies. In a pattern with reuse factor 7, two cells separate the reusing cells.

Figure 16.7 *Frequency reuse patterns*



a. Reuse factor of 4 b. Reuse factor of 7

Transmitting

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

Receiving

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

Handoff

It may happen that, during a conversation, the mobile station moves from one cell to another. When it does, the signal may become weak. To solve this problem, the MSC monitors the level of the signal every few seconds. If the strength of the signal diminishes, the MSC seeks a new cell that can better accommodate the communication. The MSC then changes the channel carrying the call (hands the signal off from the old channel to a new one).

Hard Handoff

Early systems used a hard **handoff.** In a hard handoff, a mobile station only communicates with one base station. When the MS moves from one cell to another, communication must first be broken with the previous base station before communication can be established with the new one. This may create a rough transition.

Soft Handoff

New systems use a soft handoff. In this case, a mobile station can communicate with two base stations at the same time. This means that, during handoff, a mobile station may continue with the new base station before breaking off from the old one.

Roaming

One feature of cellular telephony is called ***roaming.*** Roaming means, in principle, that a user can have access to communication or can be reached where there is coverage. A service provider usually has limited coverage. Neighboring service providers can provide extended coverage through a roaming contract. The situation is similar to snail mail between countries. The charge for delivery of a letter between two countries can be divided upon agreement by the two countries.

16.2.2 First Generation (1G)

Cellular telephony is now in its fourth generation. The first generation was designed for voice communication using analog signals. We discuss one first-generation mobile system used in North America, AMPS.

AMPS

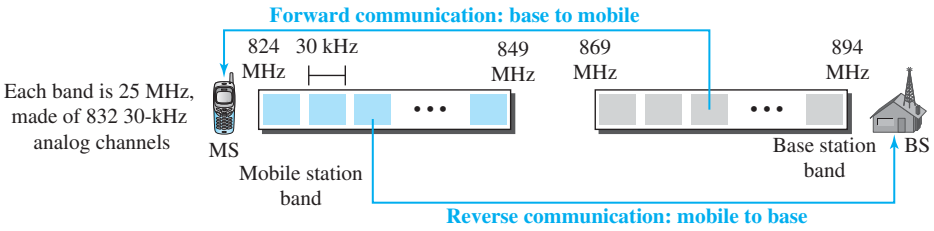
Advanced Mobile Phone System (AMPS) is one of the leading analog cellular systems in North America. It uses FDMA (see Chapter 12) to separate channels in a link.

AMPS is an analog cellular phone system using FDMA.

Bands

AMPS operates in the ISM 800-MHz band. The system uses two separate analog channels, one for forward (base station to mobile station) communication and one for reverse (mobile station to base station) communication. The band between 824 and 849 MHz carries reverse communication; the band between 869 and 894 MHz carries forward communication (see Figure 16.8).

Figure 16.8 *Cellular bands for AMPS*

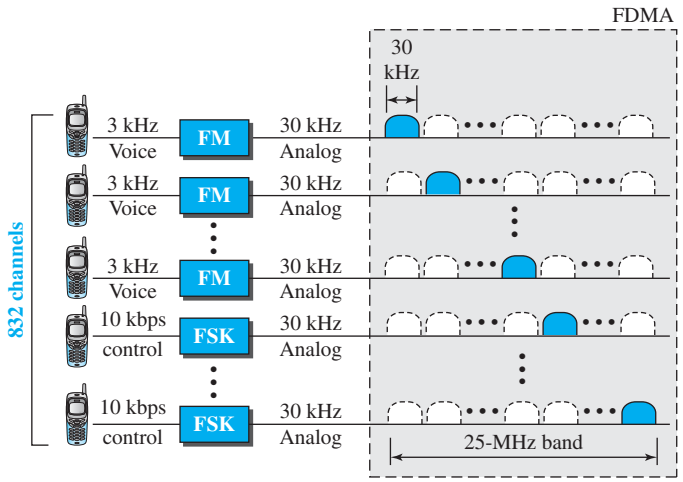


Each band is divided into 832 channels. However, two providers can share an area, which means 416 channels in each cell for each provider. Out of these 416, 21 channels are used for control, which leaves 395 channels. AMPS has a frequency reuse factor of 7; this means only one-seventh of these 395 traffic channels are actually available in a cell.

Transmission

AMPS uses FM and FSK for modulation. Figure 16.9 shows the transmission in the reverse direction. Voice channels are modulated using FM, and control channels use FSK to create 30-kHz analog signals. AMPS uses FDMA to divide each 25-MHz band into 30-kHz channels.

Figure 16.9 *AMPS reverse communication band*



16.2.3 Second Generation (2G)

To provide higher-quality (less noise-prone) mobile voice communications, the second generation of the cellular phone network was developed. While the first generation was designed for analog voice communication, the second generation was mainly designed for digitized voice. Three major systems evolved in the second generation: D-AMPS, GSM, and IS-95.

D-AMPS

The product of the evolution of the analog AMPS into a digital system is **digital AMPS (D-AMPS).** D-AMPS was designed to be backward-compatible with AMPS. This means that in a cell, one telephone can use AMPS and another D-AMPS. D-AMPS was first defined by IS-54 (Interim Standard 54) and later revised by IS-136.

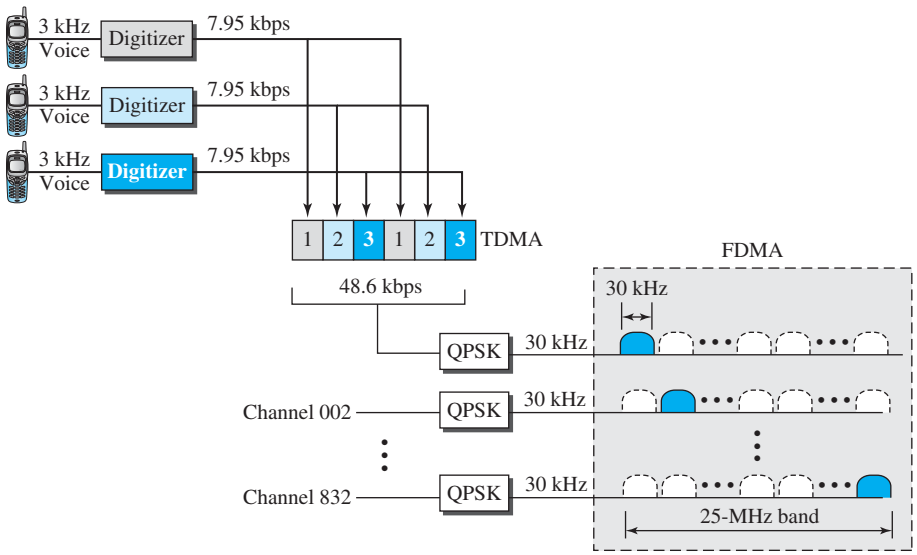
Band

D-AMPS uses the same bands and channels as AMPS.

Transmission

Each voice channel is digitized using a very complex PCM and compression technique. A voice channel is digitized to 7.95 kbps. Three 7.95-kbps digital voice channels are combined using TDMA. The result is 48.6 kbps of digital data; much of this is overhead. As Figure 16.10 shows, the system sends 25 frames per second, with 1944 bits per

Figure 16.10 *D-AMPS*



frame. Each frame lasts 40 ms (1/25) and is divided into six slots shared by three digital channels; each channel is allotted two slots.

Each slot holds 324 bits. However, only 159 bits come from the digitized voice; 64 bits are for control and 101 bits are for error correction. In other words, each channel drops 159 bits of data into each of the two channels assigned to it. The system adds 64 control bits and 101 error-correcting bits.

The resulting 48.6 kbps of digital data modulates a carrier using QPSK; the result is a 30-kHz analog signal. Finally, the 30-kHz analog signals share a 25-MHz band (FDMA). D-AMPS has a frequency reuse factor of 7.

D-AMPS, or IS-136, is a digital cellular phone system using TDMA and FDMA.

GSM

The **Global System for Mobile Communication (GSM)** is a European standard that was developed to provide a common second-generation technology for all Europe. The aim was to replace a number of incompatible first-generation technologies.

Bands

GSM uses two bands for duplex communication. Each band is 25 MHz in width, shifted toward 900 MHz, as shown in Figure 16.11. Each band is divided into 124 channels of 200 kHz separated by guard bands.

Transmission

Figure 16.12 shows a GSM system. Each voice channel is digitized and compressed to a 13-kbps digital signal. Each slot carries 156.25 bits. Eight slots share a frame (TDMA).

Figure 16.11 *GSM bands*

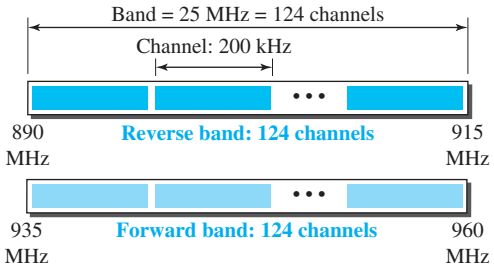
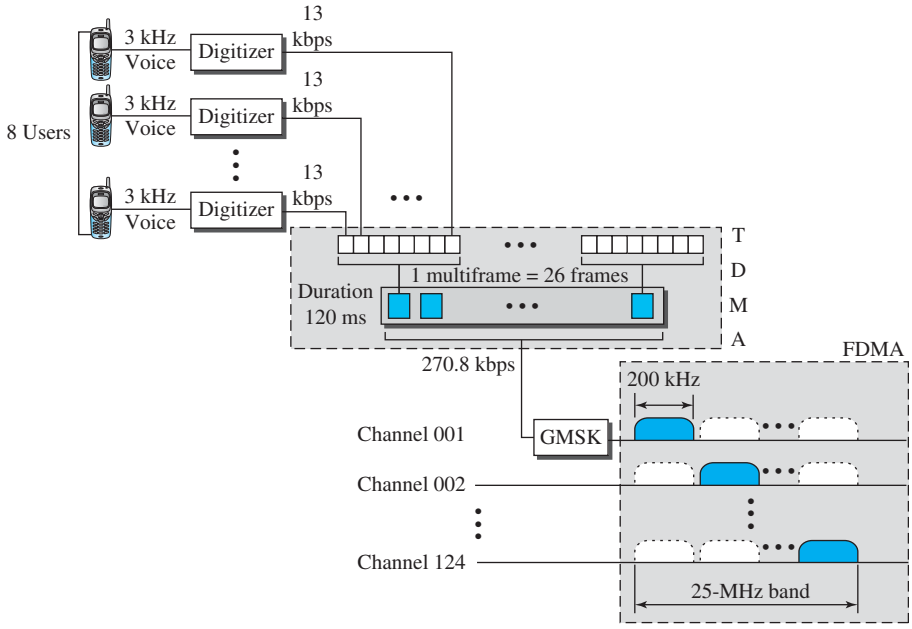


Figure 16.12 *GSM*

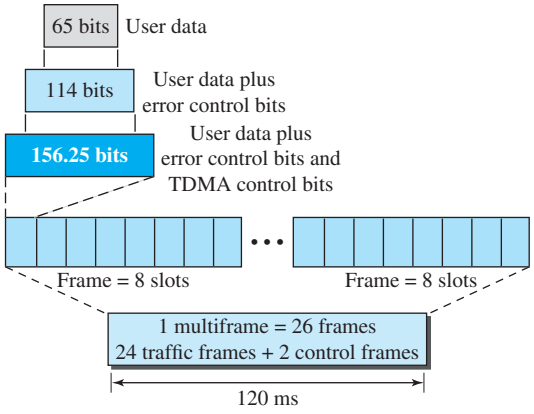


Twenty-six frames also share a multiframe (TDMA). We can calculate the bit rate of each channel as follows.

$$\text{Channel data rate} = (1/120\ \text{ms}) \times 26 \times 8 \times 156.25 = 270.8\ \text{kbps}$$

Each 270.8-kbps digital channel modulates a carrier using GMSK (a form of FSK used mainly in European systems); the result is a 200-kHz analog signal. Finally 124 analog channels of 200 kHz are combined using FDMA. The result is a 25-MHz band. Figure 16.13 shows the user data and overhead in a multiframe.

Figure 16.13 *Multiframe components*



The reader may have noticed the large amount of overhead in TDMA. The user data are only 65 bits per slot. The system adds extra bits for error correction to make it 114 bits per slot. To this, control bits are added to bring it up to 156.25 bits per slot. Eight slots are encapsulated in a frame. Twenty-four traffic frames and two additional control frames make a multiframe. A multiframe has a duration of 120 ms. However, the architecture does define superframes and hyperframes that do not add any overhead; we will not discuss them here.

Reuse Factor

Because of the complex error correction mechanism, GSM allows a reuse factor as low as 3.

GSM is a digital cellular phone system using TDMA and FDMA.

IS-95

One of the dominant second-generation standards in North America is **Interim Standard 95 (IS-95).** It is based on CDMA and DSSS.

Bands and Channels

IS-95 uses two bands for duplex communication. The bands can be the traditional ISM 800-MHz band or the ISM 1900-MHz band. Each band is divided into 20 channels of 1.228 MHz separated by guard bands. Each service provider is allotted 10 channels. IS-95 can be used in parallel with AMPS. Each IS-95 channel is equivalent to 41 AMPS channels (41×30 kHz = 1.23 MHz).

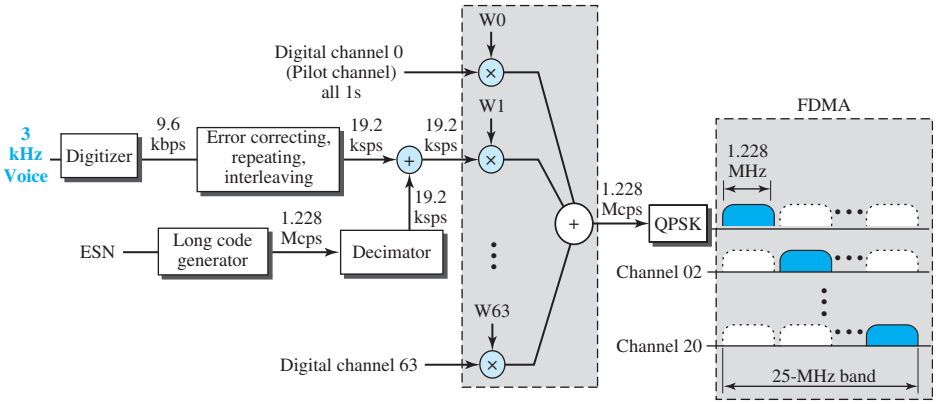
Synchronization

All base channels need to be synchronized to use CDMA. To provide synchronization, bases use the services of GPS (Global Positioning System), a satellite system that we discuss in the next section.

Forward Transmission

IS-95 has two different transmission techniques: one for use in the forward (base to mobile) direction and another for use in the reverse (mobile to base) direction. In the forward direction, communications between the base and all mobiles are synchronized; the base sends synchronized data to all mobiles. Figure 16.14 shows a simplified diagram for the forward direction.

Figure 16.14 *IS-95 forward transmission*



Each voice channel is digitized, producing data at a basic rate of 9.6 kbps. After adding error-correcting and repeating bits and interleaving, the result is a signal of 19.2 ksps (kilo signals per second). This output is now scrambled using a 19.2-ksps signal. The scrambling signal is produced from a long code generator that uses the electronic serial number (ESN) of the mobile station and generates 2^{42} pseudorandom chips, each chip having 42 bits. Note that the chips are generated pseudorandomly, not randomly, because the pattern repeats itself. The output of the long code generator is fed to a decimator, which chooses 1 bit out of 64 bits. The output of the decimator is used for scrambling. The scrambling is used to create privacy; the ESN is unique for each station.

The result of the scrambler is combined using CDMA. For each traffic channel, one Walsh 64 × 64 row chip is selected. The result is a signal of 1.228 Mcps (megachips per second).

$$\textbf{19.2 ksps} \times \textbf{64 cps} = \textbf{1.228 Mcps}$$

The signal is fed into a QPSK modulator to produce a signal of 1.228 MHz. The resulting bandwidth is shifted appropriately, using FDMA. An analog channel creates 64 digital channels, of which 55 channels are traffic channels (carrying digitized voice). Nine channels are used for control and synchronization:

a. Channel 0 is a pilot channel. This channel sends a continuous stream of 1s to mobile stations. The stream provides bit synchronization, serves as a phase

reference for demodulation, and allows the mobile station to compare the signal strength of neighboring bases for handoff decisions.

b. Channel 32 gives information about the system to the mobile station.

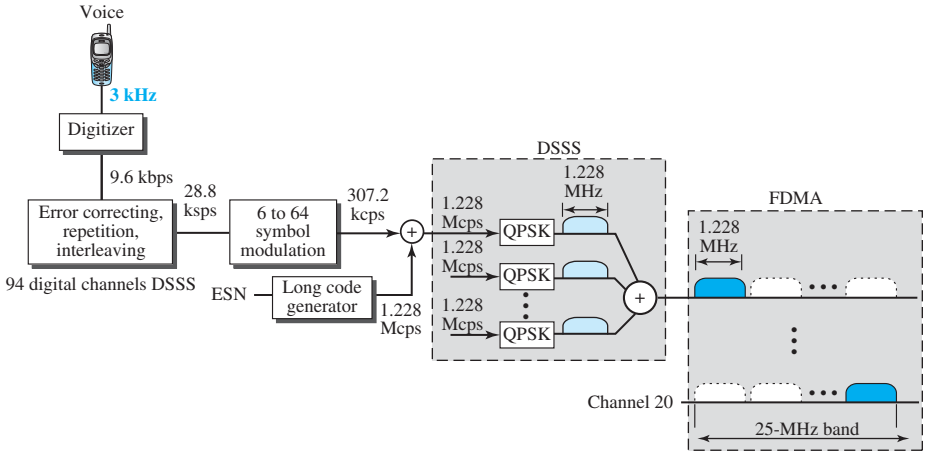
c. Channels 1 to 7 are used for paging, to send messages to one or more mobile stations.

d. Channels 8 to 31 and 33 to 63 are traffic channels carrying digitized voice from the base station to the corresponding mobile station.

Reverse Transmission

The use of CDMA in the forward direction is possible because the pilot channel sends a continuous sequence of 1s to synchronize transmission. The synchronization is not used in the reverse direction because we need an entity to do that, which is not feasible. Instead of CDMA, the reverse channels use DSSS (direct sequence spread spectrum), which we discussed in Chapter 6. Figure 16.15 shows a simplified diagram for reverse transmission.

Figure 16.15 *IS-95 reverse transmission*



Each voice channel is digitized, producing data at a rate of 9.6 kbps. However, after adding error-correcting and repeating bits plus interleaving, the result is a signal of 28.8 ksps. The output is now passed through a 6/64 symbol modulator. The symbols are divided into six-symbol chunks, and each chunk is interpreted as a binary number (from 0 to 63). The binary number is used as the index to a 64×64 Walsh matrix for selection of a row of chips. Note that this procedure is not CDMA; each bit is not multiplied by the chips in a row. Each six-symbol chunk is replaced by a 64-chip code. This is done to provide a kind of orthogonality; it differentiates the streams of chips from the different mobile stations. The result creates a signal of 307.2 kcps or $(28.8/6) \times 64$.

Spreading is the next step; each chip is spread into 4. Again the ESN of the mobile station creates a long code of 42 bits at a rate of 1.228 Mcps, which is 4 times 307.2. After spreading, each signal is modulated using QPSK, which is slightly different from the one used in the forward direction; we do not go into details here. Note that there is no multiple-access mechanism here; all reverse channels send their analog signal into the air, but the correct chips will be received by the base station due to spreading.

Although we can create $2^{42} - 1$ digital channels in the reverse direction (because of the long code generator), normally 94 channels are used; 62 are traffic channels, and 32 are channels used to gain access to the base station.

IS-95 is a digital cellular phone system using CDMA/DSSS and FDMA.

Two Data Rate Sets

IS-95 defines two data rate sets, with four different rates in each set. The first set defines 9600, 4800, 2400, and 1200 bps. If, for example, the selected rate is 1200 bps, each bit is repeated 8 times to provide a rate of 9600 bps. The second set defines 14,400, 7200, 3600, and 1800 bps. This is possible by reducing the number of bits used for error correction. The bit rates in a set are related to the activity of the channel. If the channel is silent, only 1200 bits can be transferred, which improves the spreading by repeating each bit 8 times.

Frequency-Reuse Factor

In an IS-95 system, the frequency-reuse factor is normally 1 because the interference from neighboring cells cannot affect CDMA or DSSS transmission.

Soft Handoff

Every base station continuously broadcasts signals using its pilot channel. This means a mobile station can detect the pilot signal from its cell and neighboring cells. This enables a mobile station to do a soft handoff in contrast to a hard handoff.

16.2.4 Third Generation (3G)

The third generation of cellular telephony refers to a combination of technologies that provide both digital data and voice communication. Using a small portable device, a person is able to talk to anyone else in the world with a voice quality similar to that of the existing fixed telephone network. A person can download and watch a movie, download and listen to music, surf the Internet or play games, have a video conference, and do much more. One of the interesting characteristics of a third-generation system is that the portable device is always connected; you do not need to dial a number to connect to the Internet.

The third-generation concept started in 1992, when ITU issued a blueprint called the **Internet Mobile Communication 2000 (IMT-2000).** The blueprint defines some criteria for third-generation technology as outlined below:

a. Voice quality comparable to that of the existing public telephone network.

b. Data rate of 144 kbps for access in a moving vehicle (car), 384 kbps for access as the user walks (pedestrians), and 2 Mbps for the stationary user (office or home).

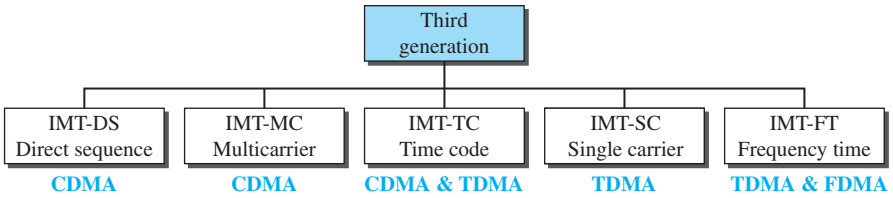
c. Support for packet-switched and circuit-switched data services.
d. A band of 2 GHz.
e. Bandwidths of 2 MHz.
f. Interface to the Internet.

The main goal of third-generation cellular telephony is to provide universal personal communication.

IMT-2000 Radio Interfaces

Figure 16.16 shows the radio interfaces (wireless standards) adopted by IMT-2000. All five are developed from second-generation technologies. The first two evolve from CDMA technology. The third evolves from a combination of CDMA and TDMA. The fourth evolves from TDMA, and the last evolves from both FDMA and TDMA.

Figure 16.16 *IMT-2000 radio interfaces*



IMT-DS

This approach uses a version of CDMA called *wideband CDMA* or *W-CDMA*. W-CDMA uses a 5-MHz bandwidth. It was developed in Europe, and it is compatible with the CDMA used in IS-95.

IMT-MC

This approach was developed in North America and is known as *CDMA 2000*. It is an evolution of CDMA technology used in IS-95 channels. It combines the new wideband (15-MHz) spread spectrum with the narrowband (1.25-MHz) CDMA of IS-95. It is backward-compatible with IS-95. It allows communication on multiple 1.25-MHz channels (1, 3, 6, 9, 12 times), up to 15 MHz. The use of the wider channels allows it to reach the 2-Mbps data rate defined for the third generation.

IMT-TC

This standard uses a combination of W-CDMA and TDMA. The standard tries to reach the IMT-2000 goals by adding TDMA multiplexing to W-CDMA.

IMT-SC

This standard uses only TDMA.

IMT-FT

This standard uses a combination of FDMA and TDMA.

16.2.5 Fourth Generation (4G)

The fourth generation of cellular telephony is expected to be a complete evolution in wireless communications. Some of the objectives defined by the 4G working group are as follows:

- **a.** A spectrally efficient system.
- **b.** High network capacity.
- **c.** Data rate of 100 Mbit/s for access in a moving car and 1 Gbit/s for stationary users.
- **d.** Data rate of at least 100 Mbit/s between any two points in the world.
- **e.** Smooth handoff across heterogeneous networks.
- **f.** Seamless connectivity and global roaming across multiple networks.
- **g.** High quality of service for next generation multimedia support (quality of service will be discussed in Chapter 30).
- **h.** Interoperability with existing wireless standards.
- **i.** All IP, packet-switched, networks.

The fourth generation is only packet-based (unlike 3G) and supports IPv6. This provides better multicast, security, and route optimization capabilities.

Access Scheme

To increase efficiency, capacity, and scalability, new access techniques are being considered for 4G. For example, **orthogonal FDMA** (**OFDMA**) and **interleaved FDMA** (**IFDMA**) are being considered respectively for the downlink and uplink of the next generation **Universal Mobile Telecommunications System** (**UMTS**). Similarly, **multicarrier code division multiple access** (**MC-CDMA**) is proposed for the IEEE 802.20 standard.

Modulation

More efficient quadrature amplitude modulation (64-QAM) is being proposed for use with the Long Term Evolution (LTE) standards.

Radio System

The fourth generation uses a **Software Defined Radio** (**SDR**) system. Unlike a common radio, which uses hardware, the components of an SDR are pieces of software and thus flexible. The SDR can change its program to shift its frequencies to mitigate frequency interference.

Antenna

The **multiple-input multiple-output** (**MIMO**) and **multiuser MIMO** (**MU-MIMO**) antenna system, a branch of intelligent antenna, is proposed for 4G. Using this antenna system together with special multiplexing, 4G allows independent streams to be transmitted simultaneously from all the antennas to increase the data rate into multiple folds. MIMO also allows the transmitter and receiver coordinates to move to an open frequency when interference occurs.

Applications

At the present rates of 15-30 Mbit/s, 4G is capable of providing users with streaming high-definition television. At rates of 100 Mbit/s, the content of a DVD-5 can be downloaded within about 5 minutes for offline access.

16.3 SATELLITE NETWORKS

A *satellite network* is a combination of nodes, some of which are satellites, that provides communication from one point on the Earth to another. A node in the network can be a satellite, an Earth station, or an end-user terminal or telephone. Although a natural satellite, such as the moon, can be used as a relaying node in the network, the use of artificial satellites is preferred because we can install electronic equipment on the satellite to regenerate the signal that has lost its energy during travel. Another restriction on using natural satellites is their distances from the Earth, which create a long delay in communication.

Satellite networks are like cellular networks in that they divide the planet into cells. Satellites can provide transmission capability to and from any location on Earth, no matter how remote. This advantage makes high-quality communication available to undeveloped parts of the world without requiring a huge investment in ground-based infrastructure.

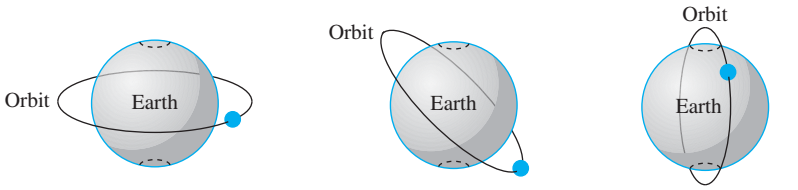
16.3.1 Operation

Let us first discuss some general issues related to the operation of satellites.

Orbits

An artificial satellite needs to have an *orbit,* the path in which it travels around the Earth. The orbit can be equatorial, inclined, or polar, as shown in Figure 16.17.

Figure 16.17 *Satellite orbits*



a. Equatorial-orbit satellite b. Inclined-orbit satellite c. Polar-orbit satellite

The period of a satellite, the time required for a satellite to make a complete trip around the Earth, is determined by Kepler's law, which defines the period as a function of the distance of the satellite from the center of the Earth.

Example 16.1

What is the period of the moon, according to Kepler's law?

$$\text{Period} = C \times \text{distance}^{1.5}$$

Here C is a constant approximately equal to 1/100. The period is in seconds and the distance in kilometers.

Solution

The moon is located approximately 384,000 km above the Earth. The radius of the Earth is 6378 km. Applying the formula, we get the following.

$$\text{Period} = (1/100) \times (384{,}000 + 6378)^{1.5} = 2{,}439{,}090 \text{ s} = 1 \text{ month}$$

Example 16.2

According to Kepler's law, what is the period of a satellite that is located at an orbit approximately 35,786 km above the Earth?

Solution

Applying the formula, we get the following.

$$\text{Period} = (1/100) \times (35{,}786 + 6378)^{1.5} = 86{,}579 \text{ s} = 24 \text{ h}$$

This means that a satellite located at 35,786 km has a period of 24 h, which is the same as the rotation period of the Earth. A satellite like this is said to be *stationary* to the Earth. The orbit, as we will see, is called a *geostationary orbit*.

Footprint

Satellites process microwaves with bidirectional antennas (line-of-sight). Therefore, the signal from a satellite is normally aimed at a specific area called the ***footprint.*** The signal power at the center of the footprint is maximum. The power decreases as we move out from the footprint center. The boundary of the footprint is the location where the power level is at a predefined threshold.

Frequency Bands for Satellite Communication

The frequencies reserved for satellite microwave communication are in the gigahertz (GHz) range. Each satellite sends and receives over two different bands. Transmission from the Earth to the satellite is called the *uplink.* Transmission from the satellite to the Earth is called the *downlink.* Table 16.1 gives the band names and frequencies for each range.

Table 16.1 *Satellite frequency bands*

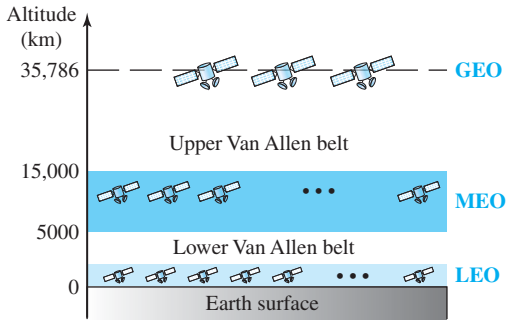
Band	*Downlink, GHz*	*Uplink, GHz*	*Bandwidth, MHz*
L	1.5	1.6	15
S	1.9	2.2	70
C	4.0	6.0	500
Ku	11.0	14.0	500
Ka	20.0	30.0	3500

Three Categories of Satellites

Based on the location of the orbit, satellites can be divided into three categories: **geostationary Earth orbit (GEO), low-Earth-orbit (LEO),** and **medium-Earth-orbit (MEO).**

Figure 16.18 shows the satellite altitudes with respect to the surface of the Earth. There is only one orbit, at an altitude of 35,786 km, for the GEO satellite. MEO satellites are located at altitudes between 5000 and 15,000 km. LEO satellites are normally below an altitude of 2000 km.

Figure 16.18 *Satellite orbit altitudes*



One reason for having different orbits is the existence of two Van Allen belts. A Van Allen belt is a layer that contains charged particles. A satellite orbiting in one of these two belts would be totally destroyed by the energetic charged particles. The MEO orbits are located between these two belts.

16.3.2 GEO Satellites

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

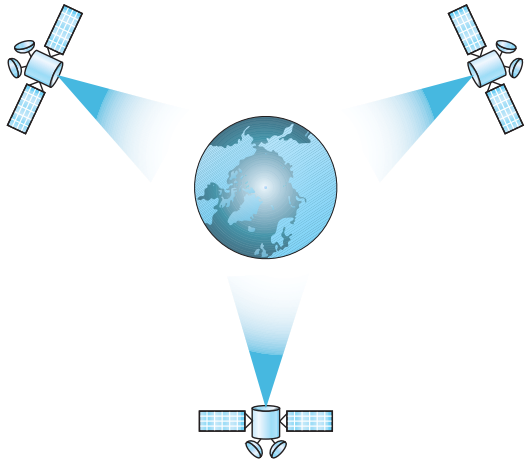
Because orbital speed is based on the distance from the planet, only one orbit can be geostationary. This orbit occurs at the equatorial plane and is approximately 22,000 mi from the surface of the Earth.

But one geostationary satellite cannot cover the whole Earth. One satellite in orbit has line-of-sight contact with a vast number of stations, but the curvature of the Earth still keeps much of the planet out of sight. It takes a minimum of three satellites equidistant from each other in geostationary Earth orbit (GEO) to provide full global transmission. Figure 16.19 shows three satellites, each 120° from another in geosynchronous orbit around the equator. The view is from the North Pole.

16.3.3 MEO Satellites

Medium-Earth-orbit (MEO) satellites are positioned between the two Van Allen belts. A satellite at this orbit takes approximately 6 to 8 hours to circle the Earth.

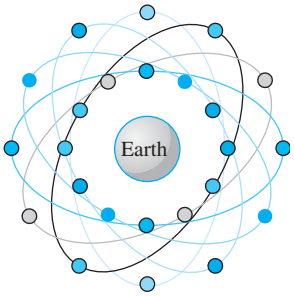
Figure 16.19 *Satellites in geostationary orbit*



Global Positioning System

One example of a MEO satellite system is the **Global Positioning System (GPS),** contracted and operated by the U.S. Department of Defense, orbiting at an altitude about 18,000 km (11,000 mi) above the Earth. The system consists of 24 satellites and is used for land, sea, and air navigation to provide time and location for vehicles and ships. GPS uses 24 satellites in six orbits, as shown in Figure 16.20. The orbits and the locations of the satellites in each orbit are designed in such a way that, at any time, four satellites are visible from any point on Earth. A GPS receiver has an almanac that tells the current position of each satellite.

Figure 16.20 *Orbits for global positioning system (GPS) satellites*

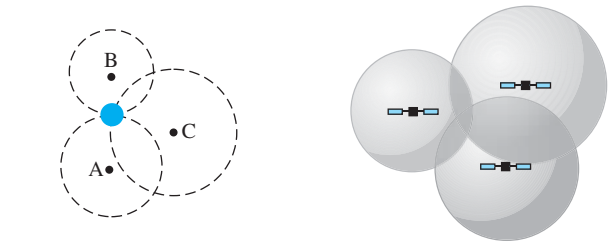


Trilateration

GPS is based on a principle called *trilateration.* The terms *trilateration* and *triangulation* are normally used interchangeably. We use the word ***trilateration,*** which means

using three distances, instead of ***triangulation,*** which may mean using three angles. On a plane, if we know our distance from three points, we know exactly where we are. Let us say that we are 10 miles away from point A, 12 miles away from point B, and 15 miles away from point C. If we draw three circles with the centers at A, B, and C, we must be somewhere on circle A, somewhere on circle B, and somewhere on circle C. These three circles meet at one single point (if our distances are correct); this is our position. Figure 16.21a shows the concept.

Figure 16.21 *Trilateration on a plane*



a. Two-dimensional trilateration b. Three-dimensional trilateration

In three-dimensional space, the situation is different. Three spheres meet in two points, as shown in Figure 16.21b. We need at least four spheres to find our exact position in space (longitude, latitude, and altitude). However, if we have additional facts about our location (for example, we know that we are not inside the ocean or somewhere in space), three spheres are enough, because one of the two points, where the spheres meet, is so improbable that the other can be selected without a doubt.

Measuring the distance

The trilateration principle can find our location on the Earth if we know our distance from three satellites and know the position of each satellite. The position of each satellite can be calculated by a GPS receiver (using the predetermined path of the satellites). The GPS receiver, then, needs to find its distance from at least three GPS satellites (center of the spheres). Measuring the distance is done using a principle called *one-way ranging*. For the moment, let us assume that all GPS satellites and the receiver on the Earth are synchronized. Each of 24 satellites synchronously transmits a complex signal, each satellite's signal having a unique pattern. The computer on the receiver measures the delay between the signals from the satellites and its copy of the signals to determine the distances to the satellites.

Synchronization

The previous discussion was based on the assumption that the satellites' clocks are synchronized with each other and with the receiver's clock. Satellites use atomic clocks, which are precise and can function synchronously with each other. The receiver's clock, however, is a normal quartz clock (an atomic clock costs more than $50,000), and there is no way to synchronize it with the satellite clocks. There is an unknown offset between the satellite clocks and the receiver clock that introduces a corresponding

offset in the distance calculation. Because of this offset, the measured distance is called a *pseudorange*.

GPS uses an elegant solution to the clock offset problem, by recognizing that the offset's value is the same for all satellites being used. The calculation of position becomes finding four unknowns: the x_r, y_r, z_r coordinates of the receiver, and common clock offset dt. For finding these four unknown values, we need at least four equations. This means that we need to measure pseudoranges from four satellites instead of three. If we call the four measured pseudoranges PR_1, PR_2, PR_3, and PR_4 and the coordinates of each satellite x_i, y_i, and z_i (for $i = 1$ to 4), we can find the four previously mentioned unknown values using the following four equations (the four unknown values are shown in color).

$$\mathbf{PR_1} = [(x_1 - \mathbf{x_r})^2 + (y_1 - \mathbf{y_r})^2 + (z_1 - \mathbf{z_r})^2]^{1/2} + \mathbf{c} \times \mathbf{dt}$$
$$\mathbf{PR_2} = [(x_2 - \mathbf{x_r})^2 + (y_2 - \mathbf{y_r})^2 + (z_2 - \mathbf{z_r})^2]^{1/2} + \mathbf{c} \times \mathbf{dt}$$
$$\mathbf{PR_3} = [(x_3 - \mathbf{x_r})^2 + (y_3 - \mathbf{y_r})^2 + (z_3 - \mathbf{z_r})^2]^{1/2} + \mathbf{c} \times \mathbf{dt}$$
$$\mathbf{PR_4} = [(x_4 - \mathbf{x_r})^2 + (y_4 - \mathbf{y_r})^2 + (z_4 - \mathbf{z_r})^2]^{1/2} + \mathbf{c} \times \mathbf{dt}$$

The coordinates used in the above formulas are in an Earth-Centered Earth-Fixed (ECEF) reference frame, which means that the origin of the coordinate space is at the center of the Earth and the coordinate space rotates with the Earth. This implies that the ECEF coordinates of a fixed point on the surface of the earth do not change.

Application

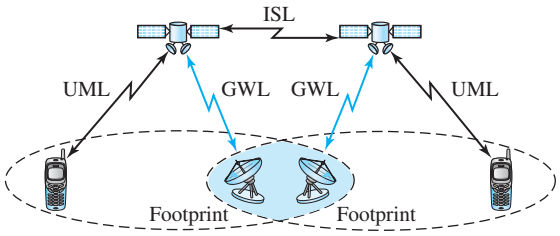
GPS is used by military forces. For example, thousands of portable GPS receivers were used during the Persian Gulf war by foot soldiers, vehicles, and helicopters. Another use of GPS is in navigation. The driver of a car can find the location of the car. The driver can then consult a database in the memory of the automobile to be directed to the destination. In other words, GPS gives the location of the car, and the database uses this information to find a path to the destination. A very interesting application is clock synchronization. As we mentioned previously, the IS-95 cellular telephone system uses GPS to create time synchronization between the base stations.

16.3.4 LEO Satellites

Low-Earth-orbit (LEO) satellites have polar orbits. The altitude is between 500 and 2000 km, with a rotation period of 90 to 120 min. The satellite has a speed of 20,000 to 25,000 km/h. A LEO system usually has a cellular type of access, similar to the cellular telephone system. The footprint normally has a diameter of 8000 km. Because LEO satellites are close to Earth, the round-trip time propagation delay is normally less than 20 ms, which is acceptable for audio communication.

A LEO system is made of a constellation of satellites that work together as a network; each satellite acts as a switch. Satellites that are close to each other are connected through intersatellite links (ISLs). A mobile system communicates with the satellite through a user mobile link (UML). A satellite can also communicate with an Earth station (gateway) through a gateway link (GWL). Figure 16.22 shows a typical LEO satellite network.

Figure 16.22 *LEO satellite system*



LEO satellites can be divided into three categories: little LEOs, big LEOs, and broadband LEOs. The little LEOs operate under 1 GHz. They are mostly used for low-data-rate messaging. The big LEOs operate between 1 and 3 GHz. **Globalstar** is one of the examples of a big LEO satellite system. It uses 48 satellites in 6 polar orbits with each orbit hosting 8 satellites. The orbits are located at an altitude of almost 1400 km. Iridium systems are also examples of big LEOs. The **Iridium** system has 66 satellites divided into 6 orbits, with 11 satellites in each orbit. The orbits are at an altitude of 750 km. The satellites in each orbit are separated from one another by approximately 32° of latitude. The broadband LEOs provide communication similar to fiber-optic networks. The first broadband LEO system was **Teledesic.** Teledesic is a system of satellites that provides fiber-optic-like communication (broadband channels, low error rate, and low delay). Its main purpose is to provide broadband Internet access for users all over the world. It is sometimes called "Internet in the sky." The project was started in 1990 by Craig McCaw and Bill Gates; later, other investors joined the consortium. The project is scheduled to be fully functional in the near future.

16.4 END-CHAPTER MATERIALS

16.4.1 Recommended Reading

For more details about subjects discussed in this chapter, we recommend the following books. The items in brackets […] refer to the reference list at the end of the text.

Books

Several books cover materials discussed in this chapter, including [Sch 03], [Gas 02]. [For 03], [Sta 04], [Sta 02], [Kei 02], [Jam 03], [AZ 03], [Tan 03], [Cou 01], [Com 06], [GW 04], and [PD 03].

16.4.2 Key Terms

adaptive antenna system (AAS)
Advanced Mobile Phone System (AMPS)
cellular telephony
digital AMPS (D-AMPS)
footprint
geostationary Earth orbit (GEO)
Global Positioning System (GPS)
Global System for Mobile Communication (GSM)
Globalstar
handoff
Interim Standard 95 (IS-95)
interleaved FDMA (IFDMA)
Internet Mobile Communication 2000 (IMT-2000)
Iridium
low-Earth-orbit (LEO)
medium-Earth-orbit (MEO)
mobile switching center (MSC)
multicarrier code division multiple access (MC-CDMA)
multiple-input multiple-output (MIMO)
multiuser MIMO (MU-MIMO)
orthogonal FDMA (OFDMA)
reuse factor
roaming
Software Defined Radio (SDR)
Teledesic
triangulation
trilateration
Universal Mobile Telecommunications System (UMTS)
Worldwide Interoperability for Microwave Access (WiMAX)

16.4.3 Summary

WiMAX is a wireless version of the wired access networks we discussed in Chapter 14. Two services are provided by WiMAX, fixed and mobile. WiMAX is based on the IEEE Project 802.16, which defines a wireless connection-oriented protocol. In WiMAX, the data-link layer is divided into three sublayers, and the physical layer is divided into two sublayers. WiMAX uses the reservation access method we discussed in Chapter 12.

Cellular telephony provides communication between two devices. One or both may be mobile. A cellular service area is divided into cells. Advanced Mobile Phone System (AMPS) is a first-generation cellular phone system. Digital AMPS (D-AMPS) is a second-generation cellular phone system that is a digital version of AMPS. Global System for Mobile Communication (GSM) is a second-generation cellular phone system used in Europe. Interim Standard 95 (IS-95) is a second-generation cellular phone system based on CDMA and DSSS. The third-generation cellular phone system provides universal personal communication. The fourth generation is the new generation of cellular phones that are becoming popular.

A satellite network uses satellites to provide communication between any points on Earth. A geostationary Earth orbit (GEO) is at the equatorial plane and revolves in phase with Earth's rotation. Global Positioning System (GPS) satellites are medium-Earth-orbit (MEO) satellites that provide time and location information for vehicles and ships. Iridium satellites are low-Earth-orbit (LEO) satellites that provide direct universal voice and data communications for handheld terminals. Teledesic satellites are low-Earth-orbit satellites that will provide universal broadband Internet access.

16.5 PRACTICE SET

16.5.1 Quizzes

A set of interactive quizzes for this chapter can be found on the book website. It is strongly recommended that the student take the quizzes to check his/her understanding of the materials before continuing with the practice set.

16.5.2 Questions

Q16-1. Explain the differences between a fixed WiMAX and a mobile WiMAX.

Q16-2. When we make a wireless Internet connection from our desktop at home, do we use fixed or mobile WiMAX?

Q16-3. When we make a wireless Internet connection from our cellular phone, do we use fixed or mobile WiMAX?

Q16-4. What is the relationship between a base station and a mobile switching center?

Q16-5. What are the functions of a mobile switching center?

Q16-6. Which is better, a low reuse factor or a high reuse factor? Explain your answer.

Q16-7. What is the difference between a hard handoff and a soft handoff?

Q16-8. What is AMPS?

Q16-9. What is the relationship between D-AMPS and AMPS?

Q16-10. What is GSM?

Q16-11. What is the function of the CDMA in IS-95?

Q16-12. What are the three types of orbits?

Q16-13. Which type of orbit does a GEO satellite have? Explain your answer.

Q16-14. What is a footprint?

Q16-15. What is the relationship between the Van Allen belts and satellites?

Q16-16. Compare an uplink with a downlink.

Q16-17. What is the purpose of GPS?

Q16-18. What is the main difference between Iridium and Globalstar?

Q16-19. To which generation does each of the following cellular telephony systems belong?

a. AMPS **b.** D-AMPS **c.** IS-95

16.5.3 Problems

P16-1. Explain how bidirectional communication can be achieved using a frame in Figure 16.5.

P16-2. In Figure 16.5, do we mean that downstream and upstream data are transmitting at the same time? Explain.

P16-3. Draw a cell pattern with a frequency-reuse factor of 5.

P16-4. Draw a cell pattern with a frequency-reuse factor of 3.

P16-5. What is the maximum number of callers in each cell in AMPS?

P16-6. What is the maximum number of simultaneous calls in each cell in an IS-136 (D-AMPS) system, assuming no analog control channels?

P16-7. What is the maximum number of simultaneous calls in each cell in a GSM, assuming no analog control channels?

P16-8. What is the maximum number of callers in each cell in an IS-95 system?

P16-9. Find the efficiency of AMPS in terms of simultaneous calls per megahertz of bandwidth. In other words, find the number of calls that can be used in 1-MHz bandwidth allocation.

P16-10. Repeat Problem P16-9 for D-AMPS.

P16-11. Repeat Problem P16-9 for GSM.

P16-12. Repeat Problem P16-9 for IS-95.

P16-13. Guess the relationship between a 3-kHz voice channel and a 30-kHz modulated channel in a system using AMPS.

P16-14. How many slots are sent each second in a channel using D-AMPS? How many slots are sent by each user in 1 s?

P16-15. Use Kepler's law to check the accuracy of a given period and altitude for a GPS satellite.

P16-16. Use Kepler's law to check the accuracy of a given period and altitude for an Iridium satellite.

P16-17. Use Kepler's law to check the accuracy of a given period and altitude for a Globalstar satellite.

P16-18. Find the efficiency of the AMPS protocol in terms of simultaneous calls per megahertz of bandwidth. In other words, find the number of calls that can be made in 1-MHz bandwidth allocation.

P16-19. Find the efficiency of the D-AMPS protocol in terms of simultaneous calls per megahertz of bandwidth. In other words, find the number of calls that can be made in 1-MHz bandwidth allocation.

P16-20. Find the efficiency of the GSM protocol in terms of simultaneous calls per megahertz of bandwidth. In other words, find the number of calls that can be made in 1-MHz bandwidth allocation.

P16-21. Find the efficiency of the IS-95 protocol in terms of simultaneous calls per megahertz of bandwidth. In other words, find the number of calls that can be made in 1-MHz bandwidth allocation.