third edition

Data Communications and Networking

Behrouz A. Forouzan

McGraw-Hill Forouzan Networking Series

Titles by Behrouz A. Forouzan:

Data Communications and Networking
TCP/IP Protocol Suite
Local Area Networks
Business Data Communications

DATA COMMUNICATIONS AND NETWORKING

Third Edition

Behrouz A. Forouzan

DeAnza College

with

Sophia Chung Fegan

McGraw Hill Higher Education

Boston Burr Ridge, IL Dubuque, IA Madison, WI New York San Francisco St. Louis
Bangkok Bogotá Caracas Kuala Lumpur Lisbon London Madrid Mexico City
Milan Montreal New Delhi Santiago Seoul Singapore Sydney Taipei Toronto

The McGraw-Hill Companies

Mc Graw Hill **Higher Education**

DATA COMMUNICATIONS AND NETWORKING, THIRD EDITION

Published by McGraw-Hill, a business unit of The McGraw-Hill Companies, Inc., 1221 Avenue of the Americas, New York, NY 10020. Copyright © 2004, 2001, 1998 by The McGraw-Hill Companies, Inc. All rights reserved. No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written consent of The McGraw-Hill Companies, Inc., including, but not limited to, in any network or other electronic storage or transmission, or broadcast for distance learning.

Some ancillaries, including electronic and print components, may not be available to customers outside the United States.

This book is printed on acid-free paper.

2 3 4 5 6 7 8 9 0 DOC/DOC 0 9 8 7 6 5 4

ISBN 0–07–251584–8

Publisher: *Elizabeth A. Jones*
Senior developmental editor: *Emily J. Lupash*
Marketing manager: *Dawn R. Bercier*
Project manager: *Sheila M. Frank*
Production supervisor: *Kara Kudronowicz*
Senior media project manager: *Jodi K. Banowetz*
Senior media technology producer: *Phillip Meek*
Designer: *Rick D. Noel*
Cover designer: *Elise Lansdon*
Cover image: © *Stone 996673-005 (RM), Skydivers in free fall formation, low angle view, Ken Fisher*
Compositor: *Interactive Composition Corporation*
Typeface: *10/12 Times Roman*
Printer: *R. R. Donnelley Crawfordsville, IN*

Library of Congress Cataloging-in-Publication Data

Forouzan, Behrouz A.
Data communications and networking / Behrouz A. Forouzan. — 3rd ed.
p. cm.
Includes index.
ISBN 0–07–251584–8
1. Data transmission systems. 2. Computer networks. I. Title.

TK5105 .F6617 2004
004.6—dc21 2003011267
CIP

www.mhhe.com

To Faezeh with love.

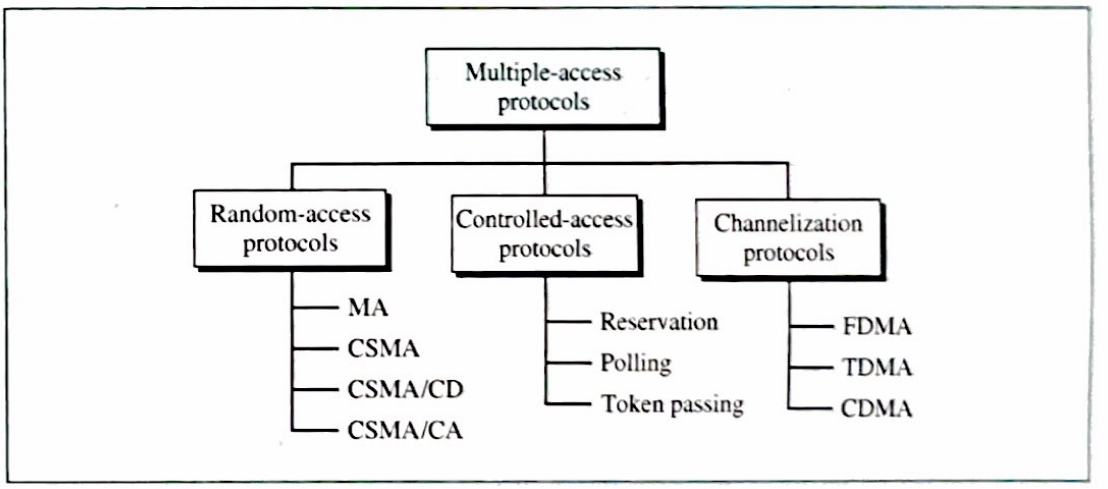
CHAPTER 13

Multiple Access

When nodes or stations are connected to or use a common link, called a multipoint or broadcast link, we need a multiple-access protocol to coordinate access to the link. The problem of controlling the access to the medium is similar to the rules of speaking in an assembly. Different procedures guarantee that the right to speak is upheld and ensure that two people do not speak at the same time, do not interrupt each other, do not monopolize the discussion, and so on.

The situation is the same with multipoint networks. Many formal protocols have been devised to handle access to the shared link. We categorize them into three groups. Protocols belonging to each group are shown in Figure 13.1.

Figure 13.1 *Multiple-access protocols*



13.1 RANDOM ACCESS

In a random access method, each station has the right to the medium without being controlled by any other station. However, if more than one station tries to send, there is an access conflict **(collision)** and the frames will be either destroyed or modified. To avoid access conflict or to resolve it when it happens, we need a procedure that answers the

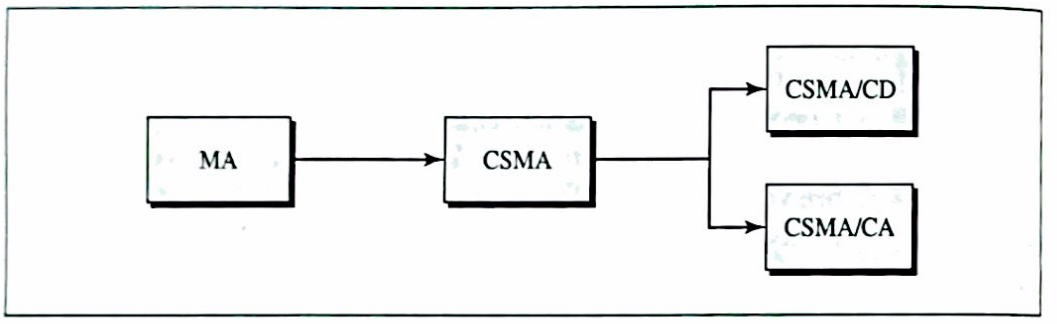
Ex.1027 / Page 7 of 37
TESLA, INC.

following questions:

- When can the station access the medium?
- What can the station do if the medium is busy?
- How can the station determine the success or failure of the transmission?
- What can the station do if there is an access conflict?

The random-access methods we study in this chapter have evolved as shown in Figure 13.2.

Figure 13.2 *Evolution of random-access methods*



The first method, known as ALOHA, used a very simple procedure called **multiple access (MA).** The method was improved with the addition of a procedure that forced the station to sense the medium before transmitting. This was called carrier sense multiple access (CSMA). This method later evolved into two parallel methods: CSMA/CD and CSMA/CA. CSMA/CD (CSMA with collision detection) defines procedures to be followed if a collision is detected while CSMA/CA (CSMA with collision avoidance) defines procedures to avoid the collision.

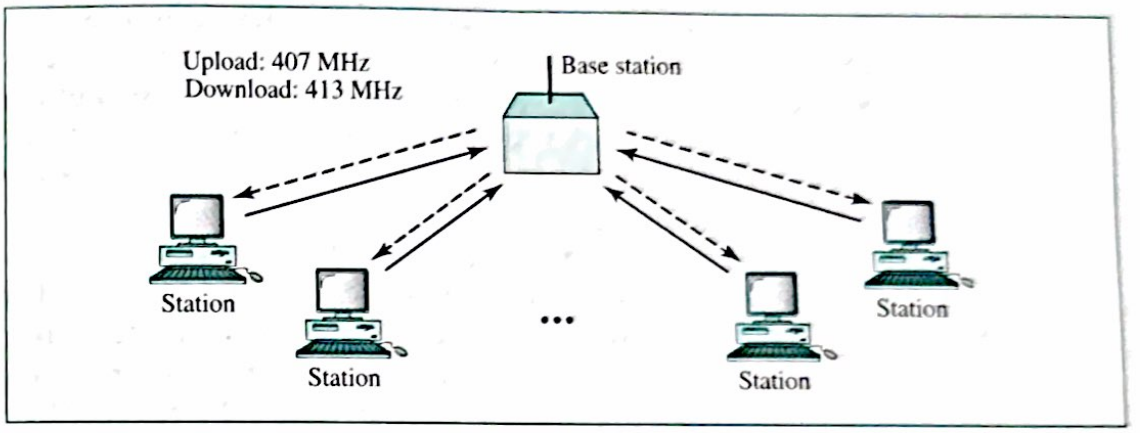
Multiple Access (MA)

ALOHA, the earliest random-access method, was developed at the University of Hawaii in the early 1970s. It was designed to be used on a radio (wireless) local area network (LAN) with a data rate of 9600 bps.

Figure 13.3 shows the basic idea behind an ALOHA network. A base station is the central controller. Every station that needs to send a frame to another station first sends it to the base station. The base station receives the frame and relays it to the intended destination. In other words, the base station acts as a hop. The uploading transmission (from a station to the base station) uses modulation with a carrier frequency of 407 MHz. The downloading transmission (from the base station to any station) uses a carrier frequency of 413 MHz.

It is obvious that there are potential collisions in this arrangement. The medium (air) is shared between the stations. When a station sends data at frequency 407 MHz to the base station, another station may attempt to do so at the same time. The data from the two stations collide and become garbled. The ALOHA protocol is very simple. It is

Figure 13.3 *ALOHA network*

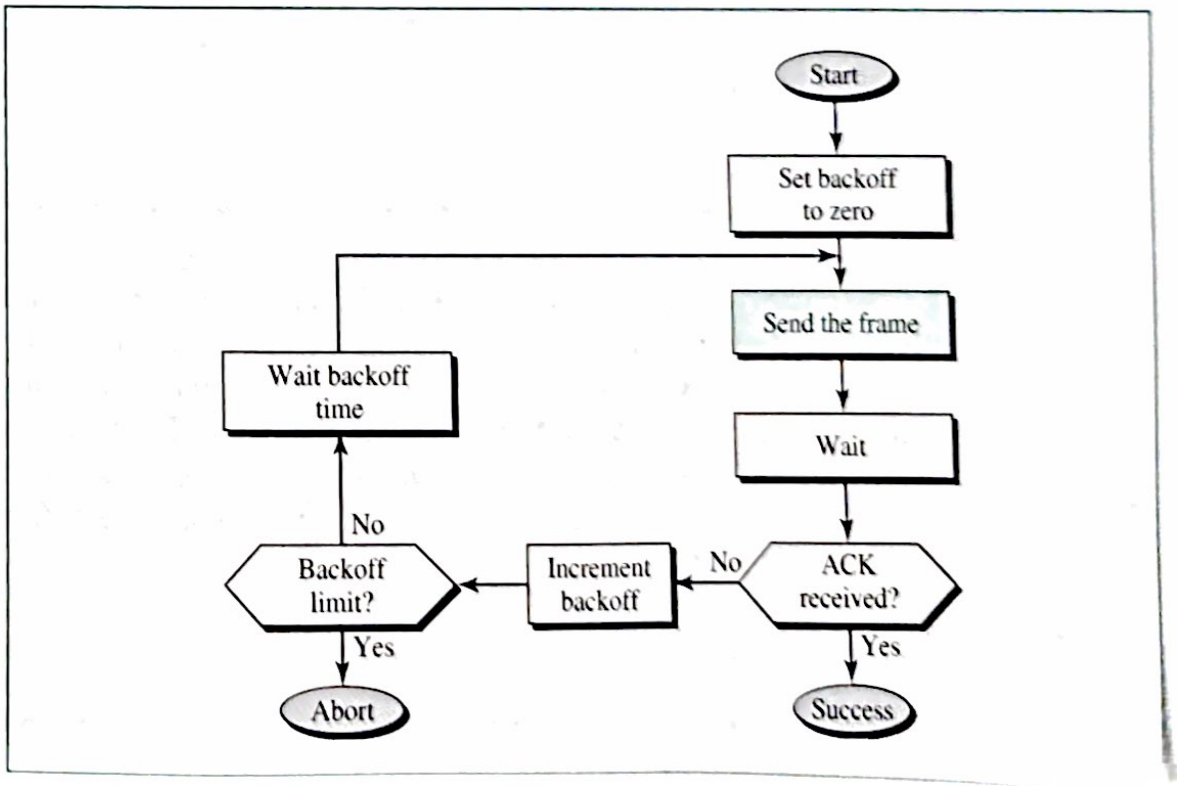


based on the following rules:

- **Multiple access.** Any station sends a frame when it has a frame to send.
- **Acknowledgment.** After sending the frame, the station waits for an acknowledgment (explicit or implicit). If it does not receive an acknowledgment during the allotted time, which is 2 times the maximum propagation delay (the time it takes for the first bit of the frame to reach every station), it assumes that the frame is lost; it tries sending again after a random amount of time.

The protocol flowchart is shown in Figure 13.4. A station that has a frame to send sends it. It then waits for a period of time, which is 2 times the maximum propagation delay. If it receives an acknowledgment, the transmission is successful. If there is no acknowledgment during this period, the station uses a backoff strategy (explained later) and sends the packet again. After several tries, if there is no acknowledgment, the station gives up.

Figure 13.4 *Procedure for ALOHA protocol*

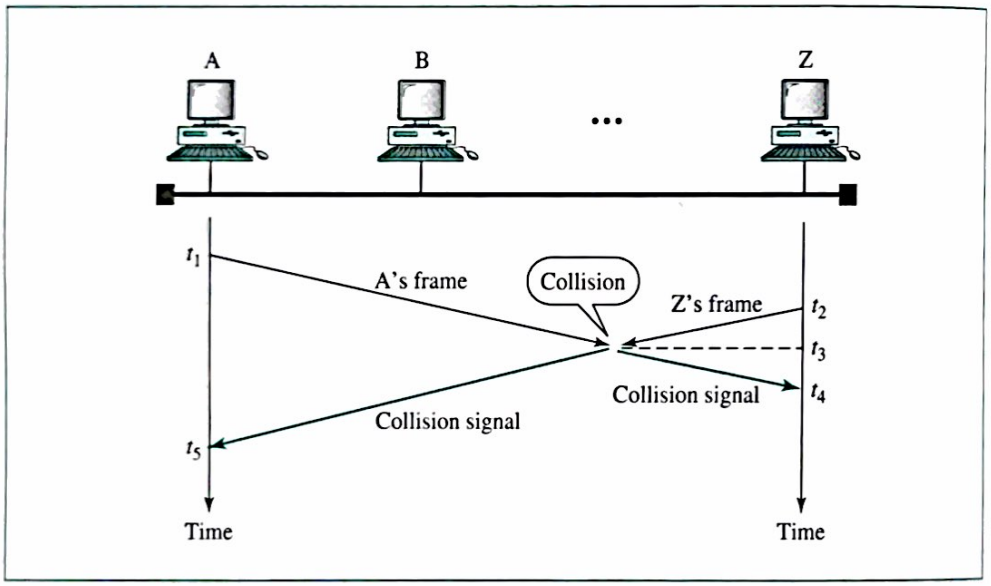


Carrier Sense Multiple Access (CSMA)

To minimize the chance of collision and, therefore, increase the performance, the CSMA method was developed. The chance of collision can be reduced if a station senses the medium before trying to use it. **Carrier sense multiple access (CSMA)** requires that each station first listen to the medium (or check the state of the medium) before sending. In other words, the CSMA is based on the principle "sense before transmit" or "listen before talk."

CSMA can reduce the possibility of collision, but it cannot eliminate it. One might ask why there may be a collision if each station listens to the medium before transmitting a frame. The possibility of collision still exists because of the propagation delay; when a station sends a frame, it takes a while (although very short) for the first bit to reach every station and for every station to sense it. In other words, a station may sense the medium and find it idle, only because propagation by another station has not yet reached this station. Figure 13.5 shows how a collision may happen.

Figure 13.5 *Collision in CSMA*

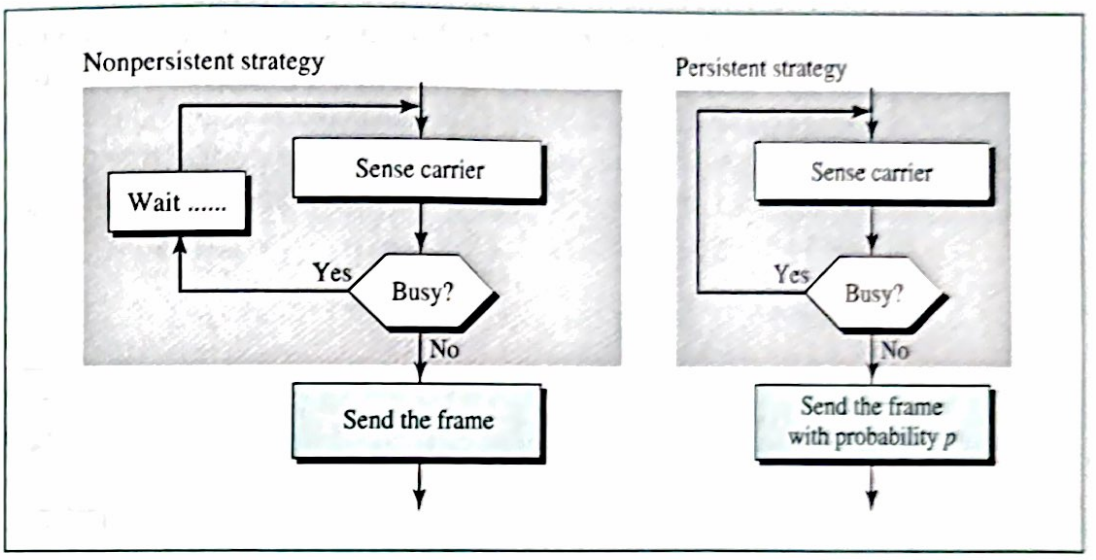


At time t_1, station A at the left end of the medium senses the medium. The medium is idle, so it sends a frame. At time t_2 ($t_2 > t_1$), station Z at the right end of the medium senses the medium and finds it idle because, at this time, propagation from station A has not reached station Z. Station Z also sends a frame. The two signals collide at time t_3 ($t_3 > t_2 > t_1$). Note that the result of the collision, which is a garbled signal, will also propagate now in both directions. It reaches station Z at time t_4 ($t_4 > t_3 > t_2 > t_1$) and station A at time t_5 ($t_5 > t_4 > t_3 > t_2 > t_1$).

Persistence Strategy

The **persistence strategy** defines the procedures for a station that senses a busy medium. Two substrategies have been devised: nonpersistent and persistent. (see Fig. 13.6).

Figure 13.6 *Persistence strategies*



Nonpersistent In a **nonpersistent strategy,** a station that has a frame to send senses the line. If the line is idle, the station sends immediately. If the line is not idle, the station waits a random period of time and then senses the line again. The nonpersistent approach reduces the chance of collision because it is unlikely that two or more stations wait the same amount of time and retry again simultaneously. However, this method reduces the efficiency of the network if the medium is idle when there are stations that have frames to send.

Persistent In a **persistent strategy,** a station senses the line. If the line is idle, the station sends a frame. This method has two variations: **1-persistent** and ***p*-persistent.**

In the 1-persistent method, if the station finds the line idle, the station sends its frame immediately (with a probability of 1). This method increases the chance of collision because two or more stations may send their frames after finding the line idle.

In the *p*-persistent method, if the station finds the line idle, the station may or may not send. It sends with probability p and refrains from sending with probability $1 - p$. For example, if p is 0.2, it means that each station, after sensing an idle line, sends with a probability of 0.2 (20 percent of the time) and refrains from sending with a probability of 0.8 (80 percent of the time). The station generates a random number between 1 and 100. If the random number is less than 20, the station will send; otherwise the station refrains from sending. The *p*-persistent strategy combines the advantages of the other two strategies. It reduces the chance of collision and improves the efficiency.

CSMA/CD

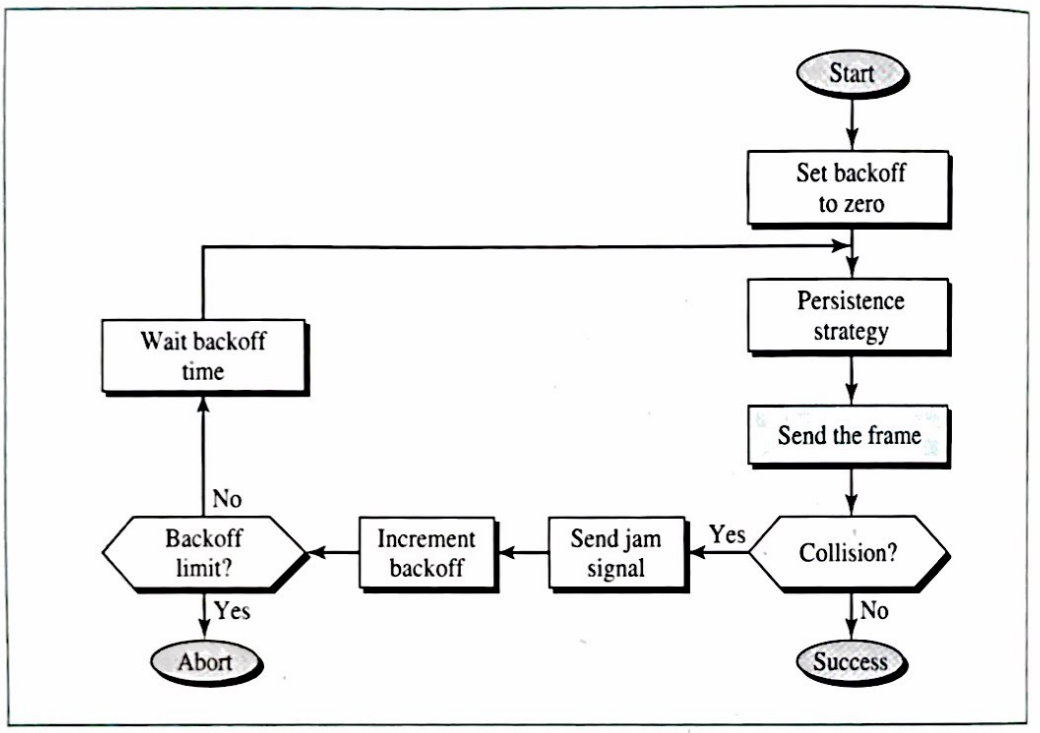
The CSMA method does not define the procedure for a collision. That is the reason CSMA was never implemented. **Carrier sense multiple access with collision detection (CSMA/CD)** adds a procedure to handle a collision.

In this method, any station can send a frame. The station then monitors the medium to see if transmission was successful. If so, the station is finished. If, however, there was a collision, the frame needs to be sent again. To reduce the probability of collision the second time, the station waits—it needs to **back off.** The question is,

How much? It is reasonable that the station waits a little the first time, more if a collision occurs again, much more if it happens a third time, and so on.

In the exponential backoff method, the station waits an amount of time between 0 and $2^N \times$ maximum_propagation_time (the time needed for a bit to reach the end of the network), where N is the number of attempted transmissions. In other words, it waits between 0 and $2 \times$ (maximum_propagation_time) for the first time, between 0 and $2^2 \times$ (maximum_propagation_time) for the second time, and so on. Figure 13.7 shows the procedure for CSMA/CD.

Figure 13.7 *CSMA/CD procedure*

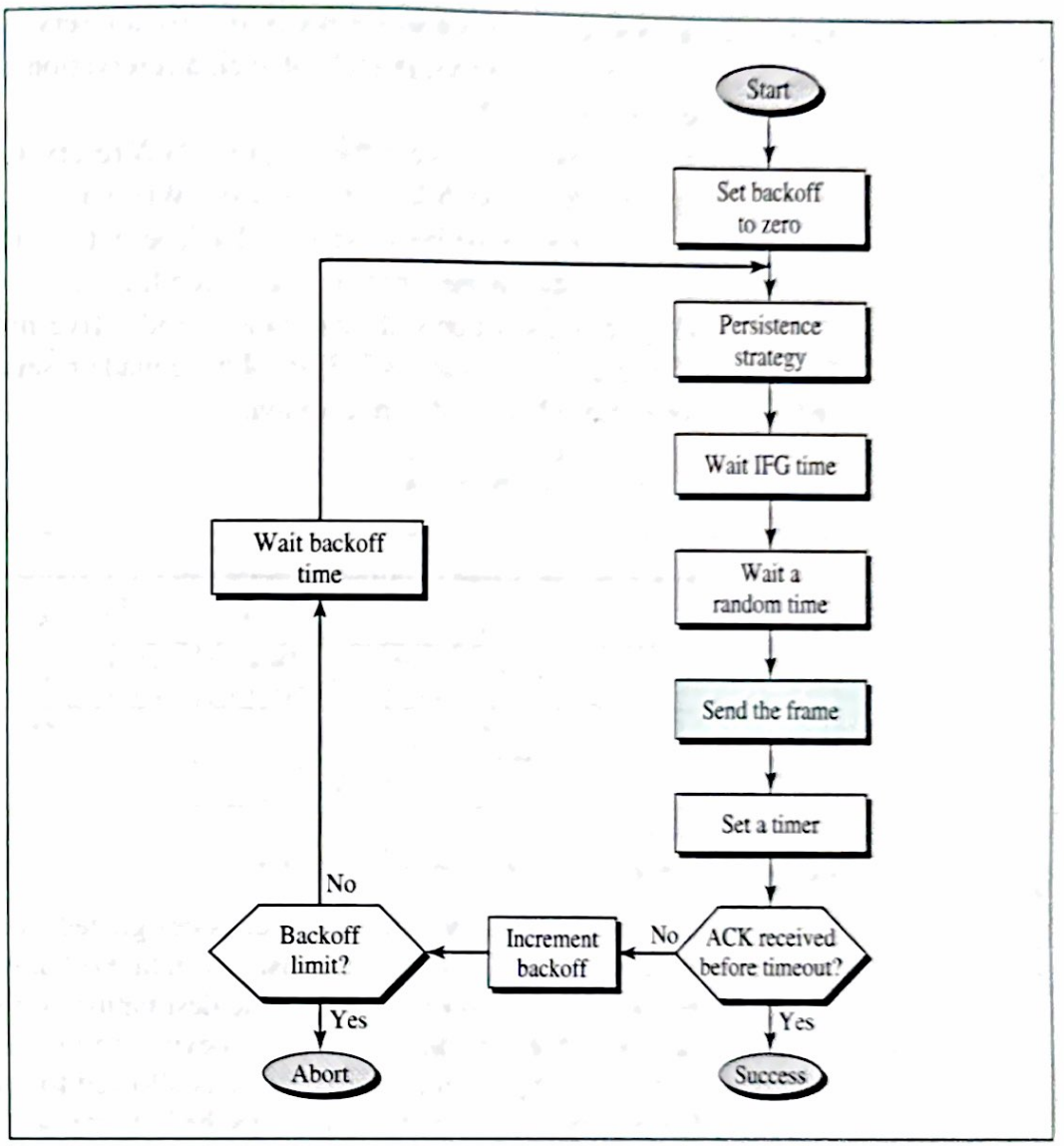


The station that has a frame to send sets the backoff parameter N to zero. It then senses the line using one of the persistence strategies. After sending the frame, if it does not hear a collision until the whole frame has been sent, the transmission is successful. However, if the station hears a collision, it sends a jam signal to the line to inform other stations of the situation and to alert them that a collision has occurred; all stations discard the part of the frame received. The station then increments the value of the backoff parameter by 1. It checks to see if the value of the backoff parameter exceeds the limit (usually 15). If this value exceeds the limit, it means that the station has tried enough and should give up the attempt; the station should abort the procedure. If the value has not exceeded the limit, the station waits a random backoff time based on the current value of the backoff parameter and senses the line again. CSMA/CD is used in traditional Ethernet (discussed in Chapter 14).

CSMA/CA

The CSMA/CA procedure differs from the previous procedures in that there is no collision. The procedure avoids collision (see Fig. 13.8). The station uses one of the

Figure 13.8 *CSMA/CA procedure*



persistence strategies. After it finds the line idle, the station waits an IFG (interframe gap) amount of time. It then waits another random amount of time. After that, it sends the frame and sets a timer. The station waits for an acknowledgment from the receiver. If it receives the acknowledgment before the timer expires, the transmission is successful. If the station does not receive an acknowledgment, it knows that something is wrong (the frame is lost or the acknowledgment is lost). The station increments the value of the backoff parameter, waits for a backoff amount of time, and resenses the line. CSMA/CA is used in wireless LANs (see Chapter 15).

13.2 CONTROLLED ACCESS

In **controlled access,** the stations consult one another to find which station has the right to send. A station cannot send unless it has been authorized by other stations. We discuss three popular controlled-access methods.

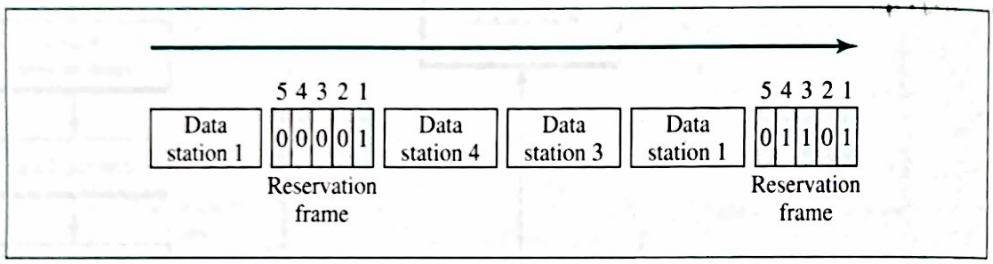
Reservation

In the reservation access method, a station needs to make a reservation before sending data. Time is divided into intervals. In each interval, a reservation frame precedes the data frames sent in that interval.

If there are N stations in the system, there are exactly N reservation minislots in the reservation frame. Each minislot belongs to a station. When a station needs to send a data frame, it makes a reservation in its own minislot. The stations that have made reservations can send their data frames after the reservation frame.

Figure 13.9 shows a situation with five stations and a five-minislot reservation frame. In the first interval, only stations 1, 3, and 4 have made reservations. In the second interval, only station 1 has made a reservation.

Figure 13.9 *Reservation access method*



Polling

Polling works with topologies in which one device is designated as a **primary station** and the other devices are **secondary stations.** All data exchanges must be made through the primary device even when the ultimate destination is a secondary device. The primary device controls the link; the secondary devices follow its instructions. It is up to the primary device to determine which device is allowed to use the channel at a given time. The primary device, therefore, is always the initiator of a session. If the primary wants to receive data, it asks the secondaries if they have anything to send; this function is called **polling.** If the primary device wants to send data, it tells the secondary target to get ready to receive; this function is called selecting.

Select

The **select mode** is used whenever the primary device has something to send. Remember that the primary controls the link. If the primary is neither sending nor receiving data, it knows the link is available. If it has something to send, the primary device sends it. What it does not know, however, is whether the target device is prepared to receive. So the primary must alert the secondary to the upcoming transmission and wait for an acknowledgment of the secondary's ready status. Before sending data, the primary creates and transmits a select (SEL) frame, one field of which includes the address of the intended secondary. See Figure 13.10.

Poll

The polling function is used by the primary device to solicit transmissions from the secondary devices. Figure 13.11 shows the situation.

Figure 13.10 *Select*

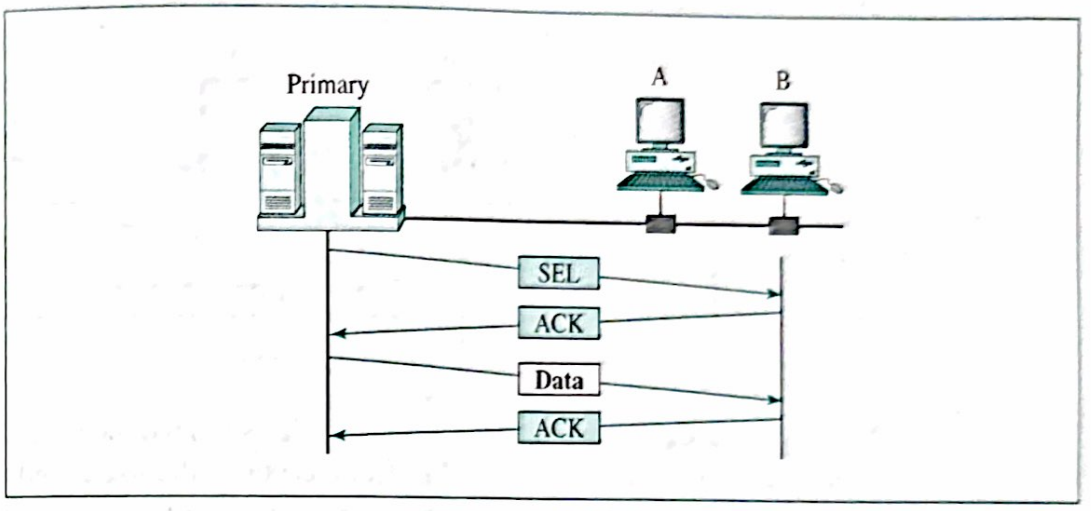
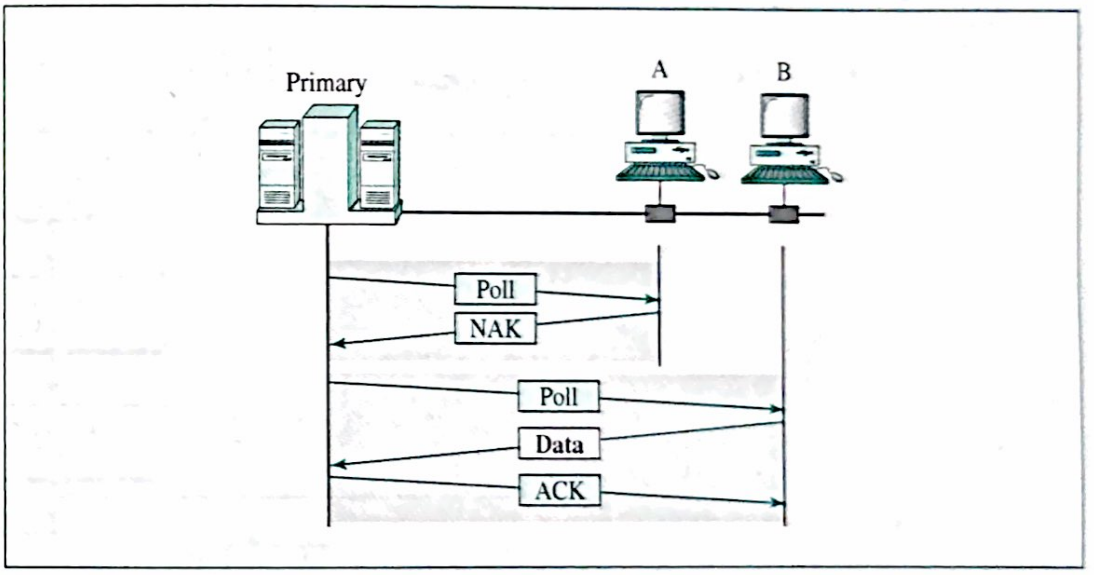


Figure 13.11 *Poll*

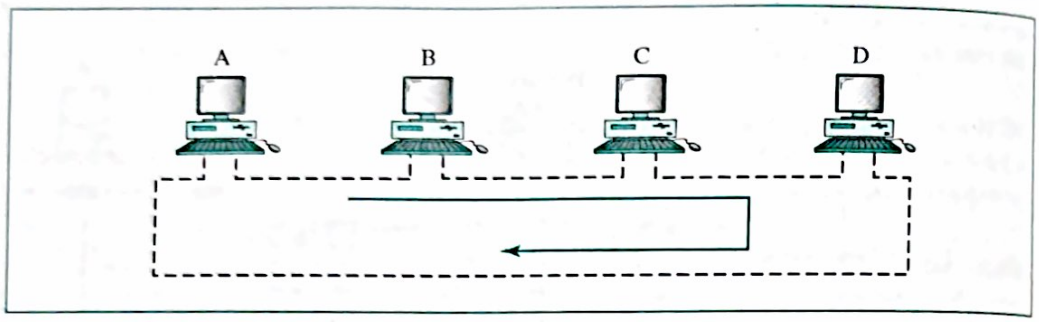


When the primary is ready to receive data, it must ask (poll) each device in turn if it has anything to send. When the first secondary is approached, it responds either with a NAK frame if it has nothing to send or with data (in the form of a data frame) if it does. If the response is negative (a NAK frame), the primary then polls the next secondary in the same manner until it finds one with data to send. When the response is positive (a data frame), the primary reads the frame and returns an acknowledgment (ACK frame), verifying its receipt.

Token Passing

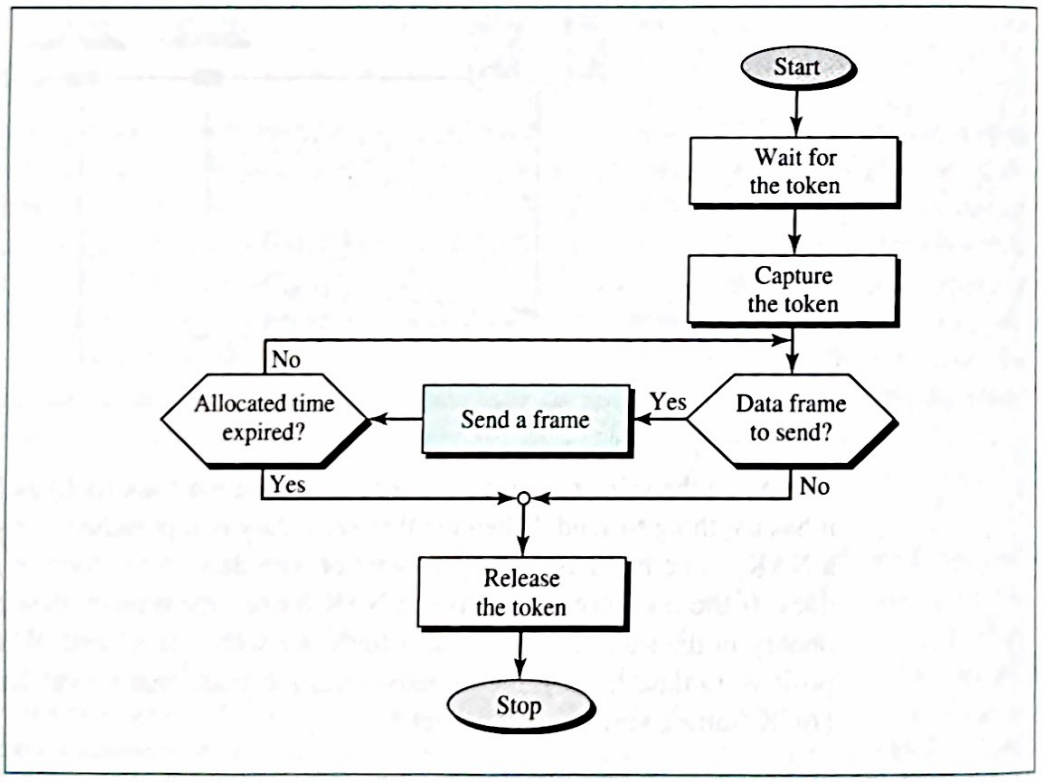
In the **token-passing** method, a station is authorized to send data when it receives a special frame called a token. In this method, stations are arranged around a ring. Each station has a predecessor and a successor. Frames are coming from the predecessor and going to the successor. Figure 13.12 shows the idea.

Figure 13.12 *Token-passing network*



When no data are being sent, a token circulates around the ring. If a station needs to send data, it waits for the token. The station captures the token and sends one or more frames (as long as it has frames to send or the allocated time has not expired), and finally it releases the token to be used by the successor station (the next station on the physical or logical ring). Figure 13.13 shows a very simplified procedure for token passing. In reality, other features (such as priority and reservation) are added to the process.

Figure 13.13 *Token-passing procedure*



13.3 CHANNELIZATION

Channelization is a multiple-access method in which the available bandwidth of a link is shared in time, frequency, or through code, between different stations. In this section, we discuss three channelization protocols: FDMA, TDMA, and CDMA. The first two

are related to procedures previously discussed in the physical layer. CDMA is a data link multiple-access protocol.

FDMA

In **frequency-division multiple access (FDMA),** the available bandwidth is shared by all stations. Each station uses its allocated band to send its data. Each band is reserved for a specific station. The band belongs to the station all the time. FDMA is a data link layer protocol that uses FDM at the physical layer (see Chapter 6). We will see how FDMA is used in cellular telephone and satellite networks in Chapter 17.

In FDMA, the bandwidth is divided into channels.

TDMA

In **time-division multiple access (TDMA),** the entire bandwidth is just one channel. The stations share the capacity of the channel in time. Each station is allocated a time slot during which it can send data. TDMA is a data link layer protocol that uses TDM at the physical layer (see Chapter 6). We will see how TDMA is used in the cellular telephone network in Chapter 17.

In TDMA, the bandwidth is just one channel that is timeshared.

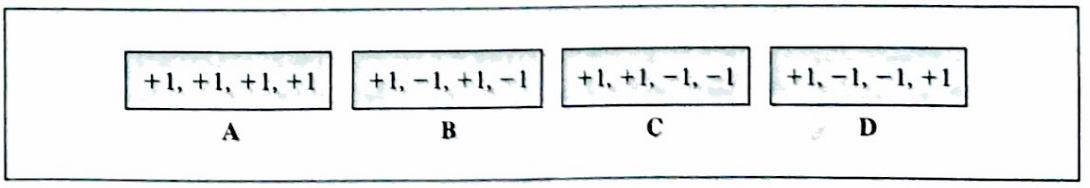
CDMA

Code-division multiple access (CDMA) was conceived several decades ago. Recent advances in electronic technology have finally made its implementation possible. CDMA differs from FDMA because only one channel occupies the entire bandwidth of the link. It differs from TDMA because all stations can send data simultaneously; there is no timesharing.

In CDMA, one channel carries all transmissions simultaneously.

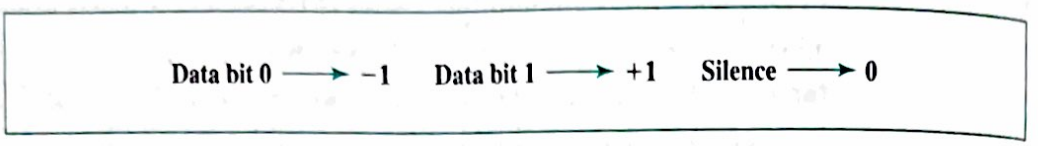
CDMA is based on coding theory. Each station is assigned a code, which is a sequence of numbers called **chips.** Suppose we have four stations; each has a sequence of chips which we designate as A, B, C, and D (see Fig. 13.14). Later in this chapter we show how we chose these sequences.

Figure 13.14 *Chip sequences*



We follow these rules for encoding: If a station needs to send a 0 bit, it sends a −1; if it needs to send a 1 bit, it sends a +1. When a station is idle, it sends no signal, which is represented by a 0. These are shown in Figure 13.15.

Figure 13.15 *Encoding rules*

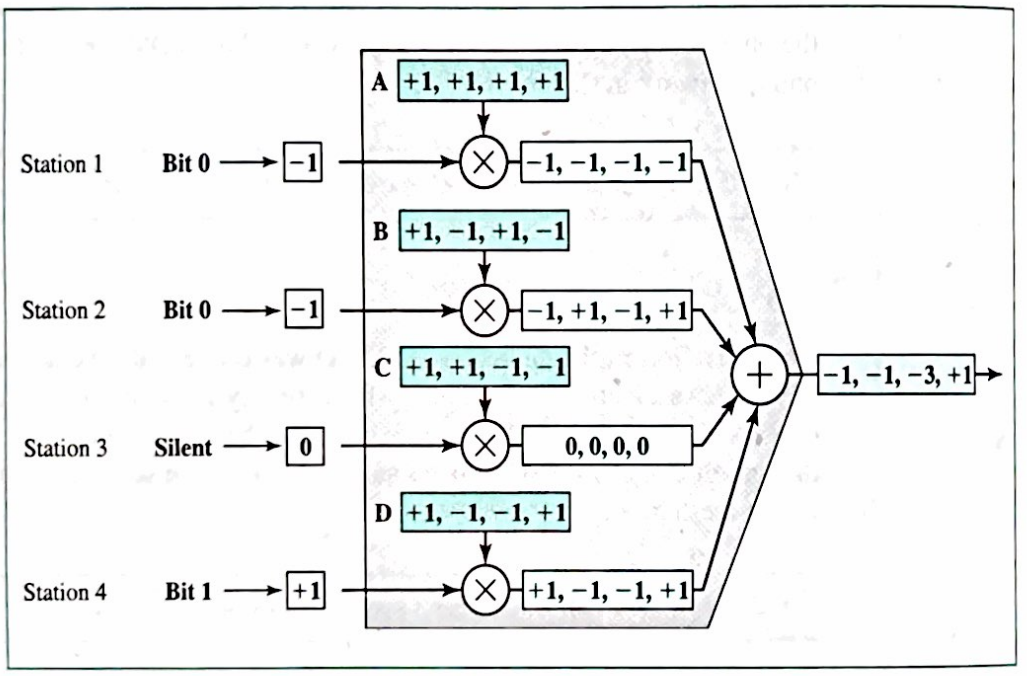


As a simple example, we show how four stations share the link during 1-bit interval. The procedure can easily be repeated for additional intervals. We assume that stations 1 and 2 are sending a 0 bit and channel 4 is sending a 1 bit. Station 3 is silent.

Multiplexer

Figure 13.16 shows the situation at the multiplexer. The steps are as follows:

Figure 13.16 *CDMA multiplexer*

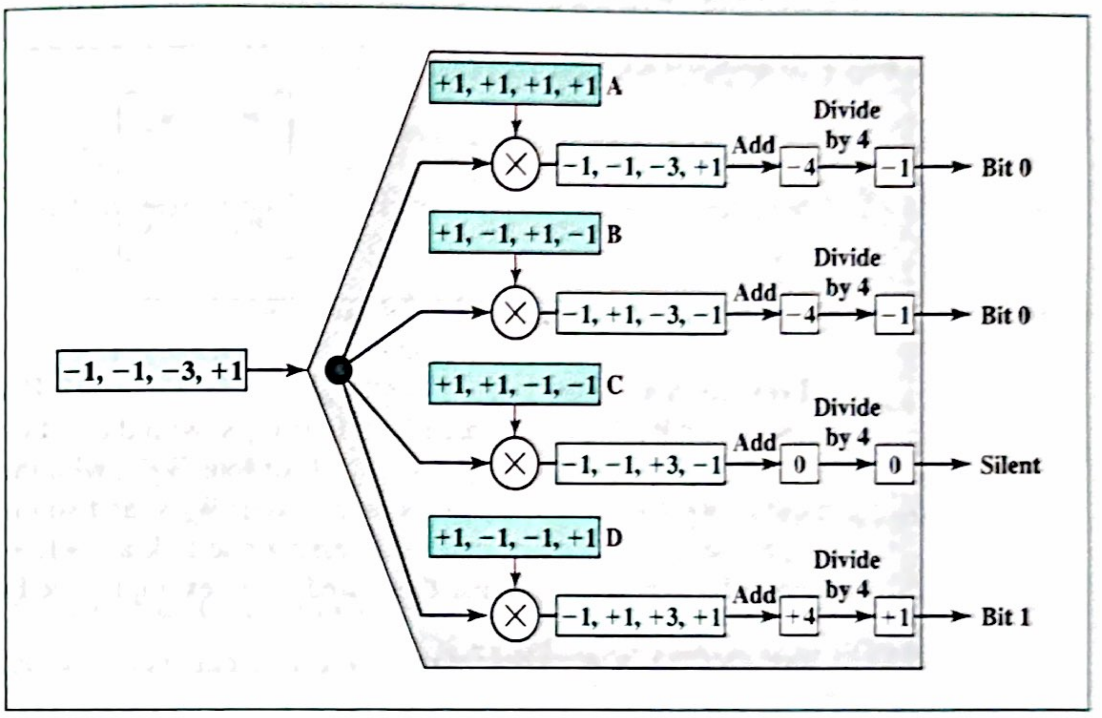


1. The multiplexer receives one encoded number from each station (−1, −1, 0, and +1).
2. The encoded number sent by station 1 is multiplied by each chip in sequence A. A new sequence is the result (−1, −1, −1, −1). Likewise, the encoded number sent by station 2 is multiplied by each chip in sequence B. The same is true for the remaining two encoded numbers. The result is four new sequences.
3. All first chips are added, as are all second, third, and fourth chips. The result is one new sequence.
4. The sequence is transmitted through the link.

Demultiplexer

Figure 13.17 shows the situation at the demultiplexer. The steps are as follows:

Figure 13.17 *CDMA demultiplexer*



1. The demultiplexer receives the sequence sent across the link.
2. It multiplies the sequence by the code for each receiver. The multiplication is done chip by chip.
3. The chips in each sequence are added. The result is always +4, −4, or 0.
4. The result of step 3 is divided by 4 to get −1, +1, or 0.
5. The number in step 4 is decoded to 0, 1, or silence by the receiver.

Observation

We see that each station receives what is sent by the sender. Note that the third receiver does not receive data because the sender was idle. There is only one sequence flowing through the channel, the sum of the sequences. However, each receiver detects its own data from the sum.

Orthogonal Sequences Let us return to the chip sequences. We did not choose the sequences randomly; they were carefully selected. The sequences in our example are called **orthogonal sequences.** We show how to generate orthogonal sequences, and then we discuss their properties.

Sequence Generation To generate sequences, we use a **Walsh table,** a two-dimensional table with an equal number of rows and columns. Each row is a sequence of chips. The Walsh table W_1 for a one-chip sequence has one row and one column. We can choose −1 or +1 for the chip for this trivial table (we chose +1). According to Walsh, if we

know the table for N sequences W_N, we can create the table for $2N$ sequences W_{2N}, as shown in Figure 13.18. The $\overline{W_N}$ with the overhead bar stands for the complement of W_N, where each +1 is changed to −1 and vice versa.

Figure 13.18 *W_1 and W_{2N}*

$$W_1 = \begin{bmatrix} +1 \end{bmatrix} \qquad W_{2N} = \begin{bmatrix} W_N & W_N \\ W_N & \overline{W_N} \end{bmatrix}$$

Let us see how we can create W_2 and W_4 from W_1. Figure 13.19 shows the process. After we select W_1, W_2 can be made from four W_1's, with the last one the complement of W_1. After W_2 is generated, W_4 can be made of four W_2's, with the last one the complement of W_2. Of course, W_8 is composed of four W_4's, and so on. According to the tables, the sequences for two stations accessing one link are +1, +1 and +1, −1. The sequences for four stations are like those used in our example (see Fig. 13.16).

Figure 13.19 *Sequence generation*

$$W_1 = \begin{bmatrix} +1 \end{bmatrix} \qquad W_2 = \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix} \qquad W_4 = \begin{bmatrix} +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 \end{bmatrix}$$

Properties of Orthogonal Sequences Orthogonal sequences have properties that are suitable for CDMA. They are as follows:

1. If we multiply a sequence by −1, every element in the sequence is complemented (+1 becomes −1 and −1 becomes +1). We can see that when a station is sending −1 (bit 0), it is sending its complement.
2. If we multiply two sequences, element by element, and add the results, we get a number called the **inner product.** If the two sequences are the same, we get N, where N is the number of sequences; if they are different, we get 0. The inner product uses a dot as the operator. So $A \cdot A$ is N, but $A \cdot B$ is 0.
3. The inner product of a sequence by its complement is $-N$. So $A \cdot (-A)$ is $-N$.

Example 1

Check to see if the second property about orthogonal codes holds for our CDMA example.

Solution

The inner product of each code by itself is N. This is shown for code C; you can prove for yourself that it holds true for the other codes.

$$C \cdot C = [+1, +1, -1, -1] \cdot [+1, +1, -1, -1] = 1 + 1 + 1 + 1 = 4$$

If two sequences are different, the inner product is 0.

$$B \cdot C = [+1, -1, +1, -1] \cdot [+1, +1, -1, -1] = 1 - 1 - 1 + 1 = 0$$

Example 2

Check to see if the third property about orthogonal codes holds for our CDMA example.

Solution

The inner product of each code by its complement is $-N$. This is shown for code C; you can prove for yourself that it holds true for the other codes.

$$C \cdot (-C) = [+1, +1, -1, -1] \cdot [-1, -1, +1, +1] = -1 - 1 - 1 - 1 = -4$$

The inner product of a code with the complement of another code is 0.

$$B \cdot (-C) = [+1, -1, +1, -1] \cdot [-1, -1, +1, +1] = -1 + 1 + 1 - 1 = 0$$

Orthogonal Codes in CDMA

Now we can understand why our trivial example of CDMA works.

In the Multiplexer In the multiplexer, each station is sending the appropriate sequence:

- Station 1 is sending $-A$ (-1 was multiplied by A), station 2 is sending $-B$, station 3 is sending an empty sequence (all zeros), and station 4 is sending D.
- The sequence that comes out of the multiplexer is the sum of all sequences.

$$S = -A - B + D$$

In the Demultiplexer In the demultiplexer, all stations receive S:

- Station 1 finds the inner product of S and A.

$$S \cdot A = (-A - B + D) \cdot A = -A \cdot A - B \cdot A + D \cdot A = -4 + 0 + 0 = -4$$

  The result is then divided by 4, which is -1. This is interpreted as a 0 bit.
- Station 2 finds the inner product of S and B.

$$S \cdot B = (-A - B + D) \cdot B = -A \cdot B - B \cdot B + D \cdot B = 0 - 4 + 0 = -4$$

  The result is then divided by 4, which is -1. This is interpreted as a 0 bit.
- Station 3 finds the inner product of S and C.

$$S \cdot C = (-A - B + D) \cdot C = -A \cdot C - B \cdot C + D \cdot C = 0 + 0 + 0 = 0$$

  This is interpreted as a silent station.
- Station 4 finds the inner product of S and D.

$$S \cdot D = (-A - B + D) \cdot D = -A \cdot D - B \cdot D + D \cdot D = 0 + 0 + 4 = 4$$

  The result is then divided by 4, which is $+1$. This is interpreted as a 1 bit.

13.4 KEY TERMS

1-persistent strategy
ALOHA
back off
carrier sense multiple access (CSMA)
carrier sense multiple access/collision avoidance (CSMA/CA)
carrier sense multiple access/collision detection (CSMA/CD)
channelization
chip
code-division multiple access (CDMA)
collision
controlled access
frequency-division multiple access (FDMA)
inner product
multiple access (MA)
nonpersistent strategy
orthogonal sequence
persistence strategy
poll mode
polling
p-persistent strategy
primary station
random access
reservation
secondary station
select mode
time-division multiple access (TDMA)
token passing
Walsh table

13.5 SUMMARY

- Medium access methods can be categorized as random, controlled, or channelized.
- In the carrier sense multiple-access (CSMA) method, a station must listen to the medium prior to sending data onto the line.
- A persistence strategy defines the procedure to follow when a station senses an occupied medium.
- Carrier sense multiple access with collision detection (CSMA/CD) is CSMA with a postcollision procedure.
- Carrier sense multiple access with collision avoidance (CSMA/CA) is CSMA with procedures that avoid a collision.
- Reservation, polling, and token passing are controlled-access methods.
- In the reservation access method, a station reserves a slot for data by setting its flag in a reservation frame.
- In the polling access method, a primary station controls transmissions to and from secondary stations.
- In the token-passing access method, a station that has control of a frame called a token can send data.
- Channelization is a multiple-access method in which the available bandwidth of a link is shared in time, frequency, or through code, between stations on a network.
- FDMA, TDMA, and CDMA are channelization methods.

- In FDMA, the bandwidth is divided into bands; each band is reserved for the use of a specific station.
- In TDMA, the bandwidth is not divided into bands; instead the bandwidth is timeshared.
- In CDMA, the bandwidth is not divided into bands, yet data from all inputs are transmitted simultaneously.
- CDMA is based on coding theory and uses sequences of numbers called chips. The sequences are generated using Walsh tables.

13.6 PRACTICE SET

Review Questions

1. What is the advantage of controlled access over random access?
2. List in order the protocols that evolved from MA.
3. When do we increment the backoff in a network that uses ALOHA?
4. How do the two persistent strategies differ?
5. What is the purpose of the jam signal in CSMA/CD?
6. How does CSMA/CD differ from CSMA/CA?
7. Name three popular controlled-access methods.
8. Is the reservation access method suitable for a very large network in which many stations are idle? Why or why not?
9. Discuss the difference between polling and selecting.
10. Why is token passing a controlled-access procedure?
11. Name three channelization protocols.
12. Discuss the number of bands per medium bandwidth for FDMA, TDMA, and CDMA.
13. How is CDMA superior to FDMA? How is CDMA superior to TDMA?
14. What is a collision?
15. What is an inner product?

Multiple-Choice Questions

16. The most primitive random access method is _____.
    a. ALOHA
    b. CSMA
    c. Channelization
    d. Token passing
17. In the _____ random-access method there is no collision.
    a. ALOHA
    b. CSMA/CD
    c. CSMA/CA
    d. Token-passing

CHAPTER 17

Cellular Telephone and Satellite Networks

We discussed wireless LANs in Chapter 15. Wireless technology is also used in cellular telephony and satellite networks. We discuss the former in this chapter as well as examples of channelization access methods (see Chapter 13). We also briefly discuss satellite networks, a technology that eventually will be linked to cellular telephony to access the Internet directly.

17.1 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 small office, 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 Fig. 17.1).

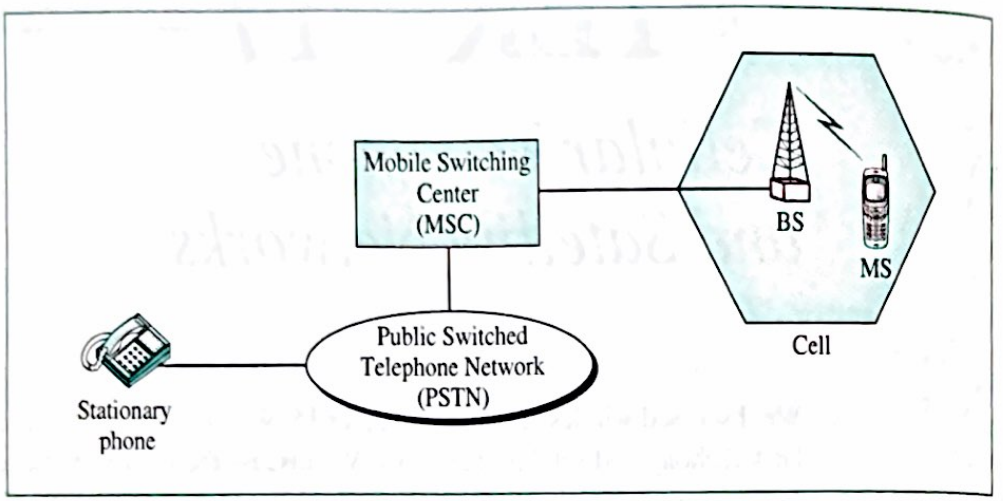
Cell size is not fixed and can be increased or decreased depending on the population of the area. The typical radius of a cell is 1 to 12 miles. High-density areas require more geographically smaller cells to meet traffic demands than do lower-density areas. Once determined, cell size is optimized to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

Frequency-Reuse Principle

In general, neighboring cells cannot use the same set of frequencies for communication because it may create interference for the users located near the cell boundaries. However,

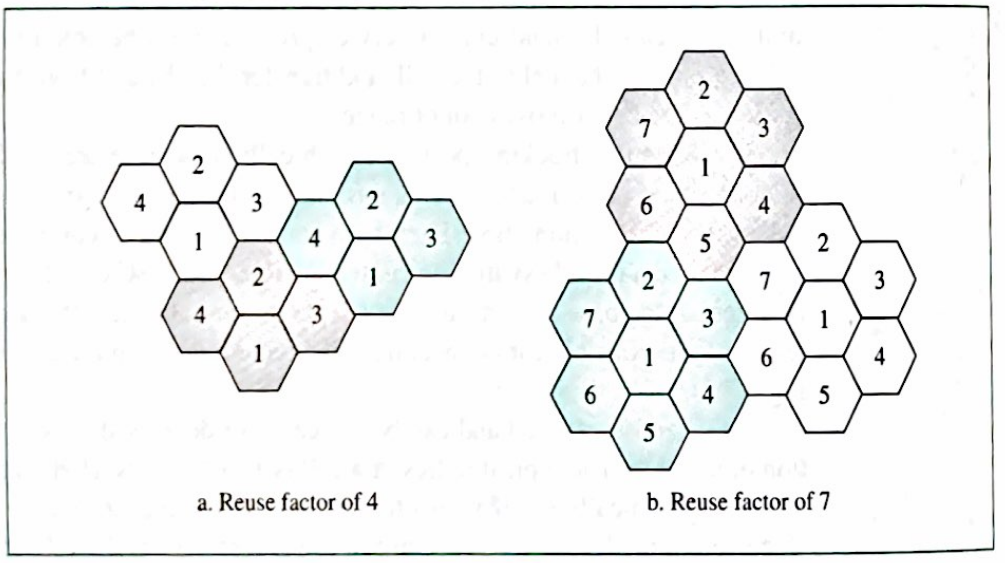
Ex.1027 / Page 25 of 37
TESLA, INC.

Figure 17.1 *Cellular system*



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 17.2 shows two of them.

Figure 17.2 *Frequency reuse patterns*



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

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 station may continue with the new base station before breaking 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.

First Generation

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

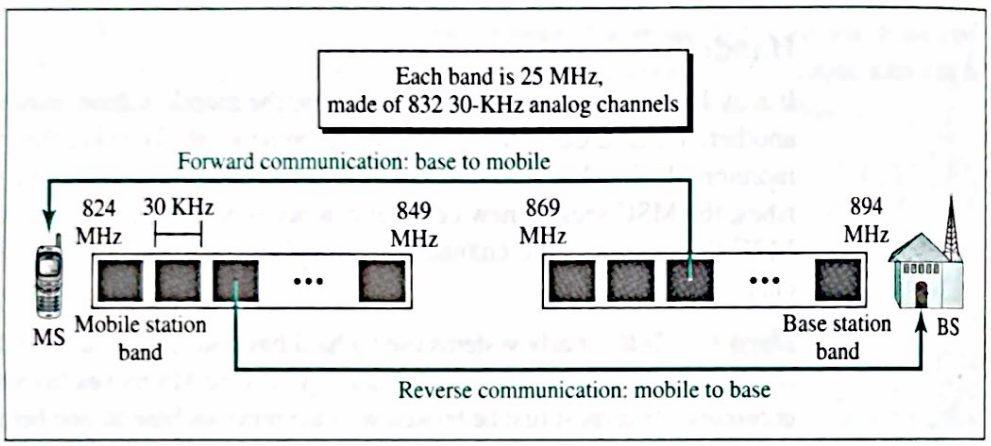
AMPS

Advanced Mobile Phone System (AMPS) is one of the leading analog cellular systems in North America. It uses FDMA 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 for forward (base station to mobile station) and 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 17.3.

Figure 17.3 *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 17.4 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.

Second Generation

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, as shown in Figure 17.5. We will discuss each system separately.

Figure 17.4 *AMPS reverse communication band*

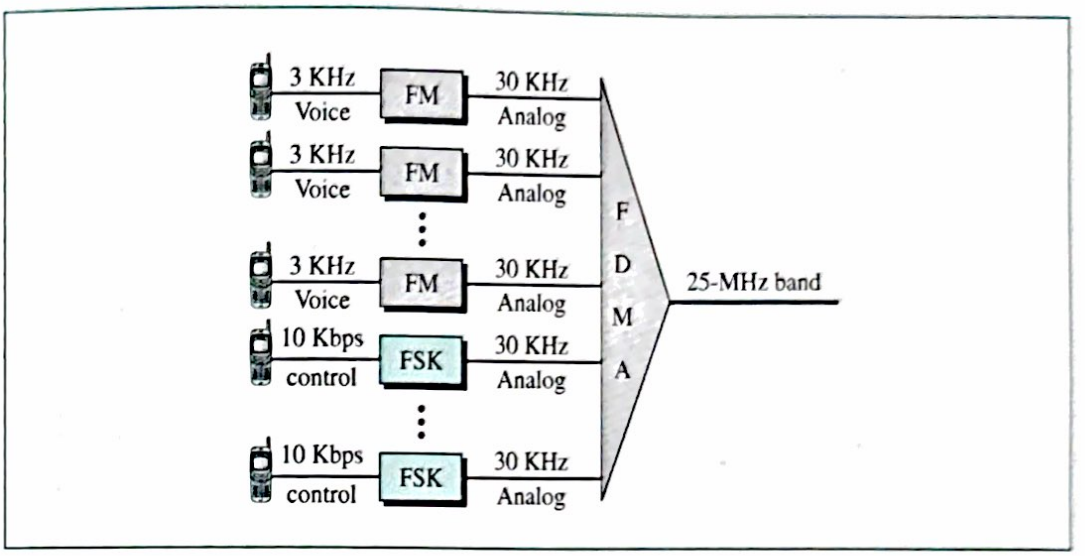
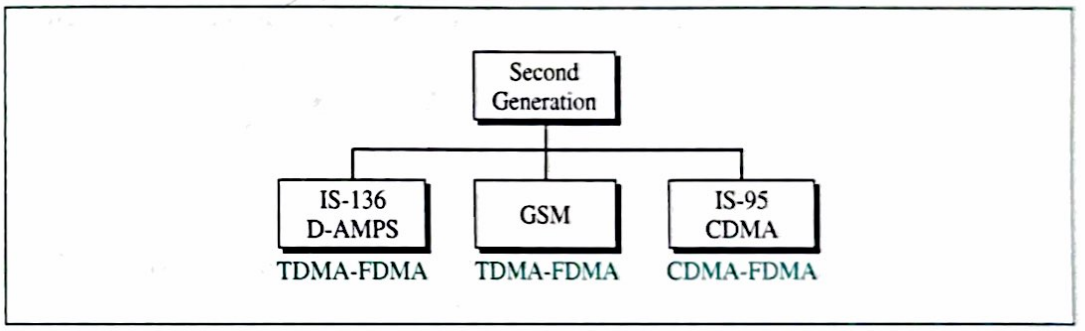


Figure 17.5 *Second-generation cellular phone systems*



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.

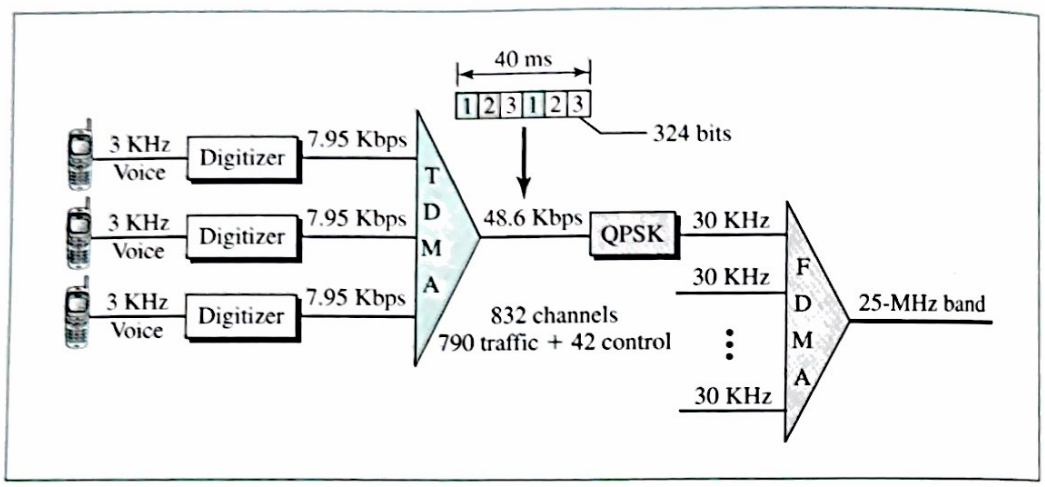
Band D-AMPS uses the same bands and channels as AMPS.

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 17.6 shows, the system sends 25 frames per second, with 1944 bits per 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 comes from the digitized voice; 64 bits for control, and 101 bits for error correction. In other words, each channel drops 159 bits of data in each of the two channels assigned to it. The system adds 64 control bits and 101 error-correcting bits.

Figure 17.6 *D-AMPS*



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 are frequency-multiplexed in the 25-MHz band. 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 of 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 17.7. Each band is divided into 124 channels of 200 KHz separated by guard bands.

Transmission Figure 17.8 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 are

Figure 17.7 *GSM bands*

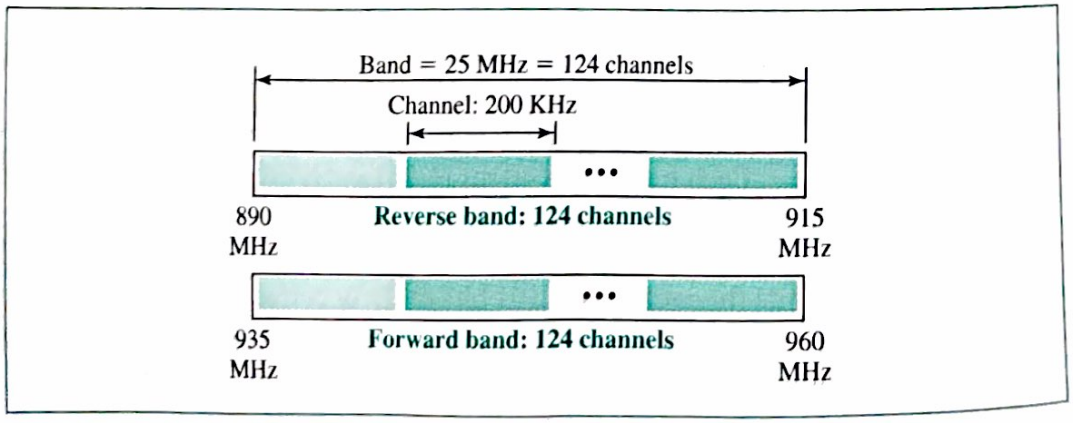
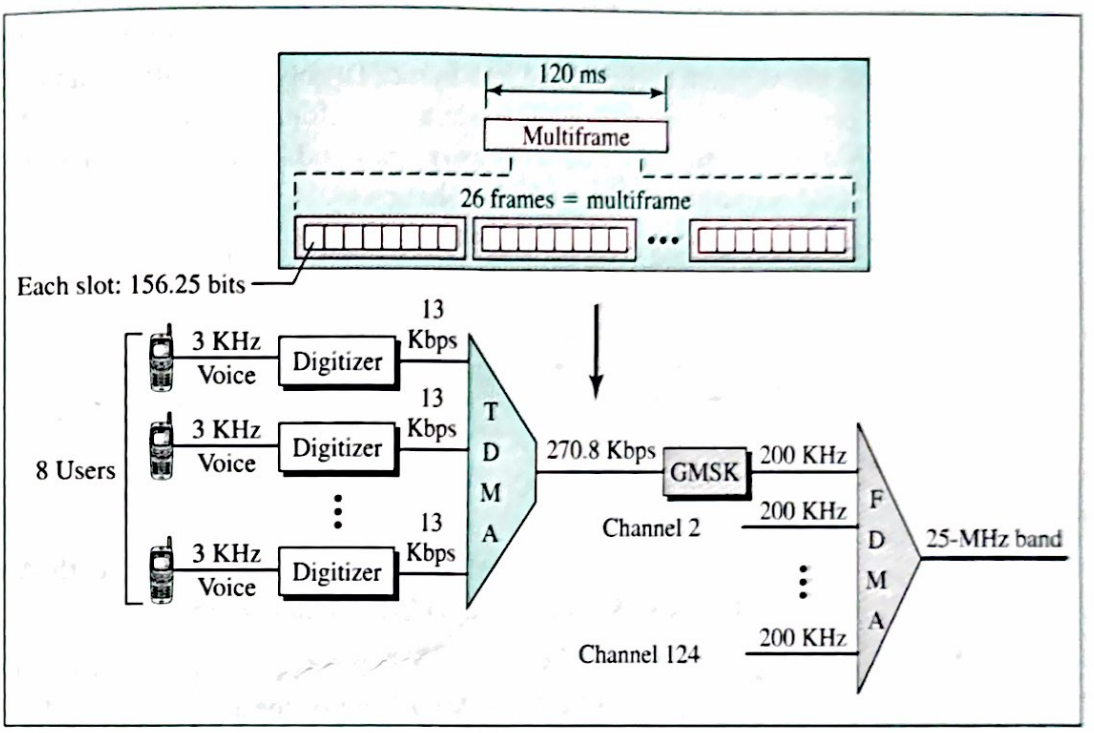


Figure 17.8 *GSM*



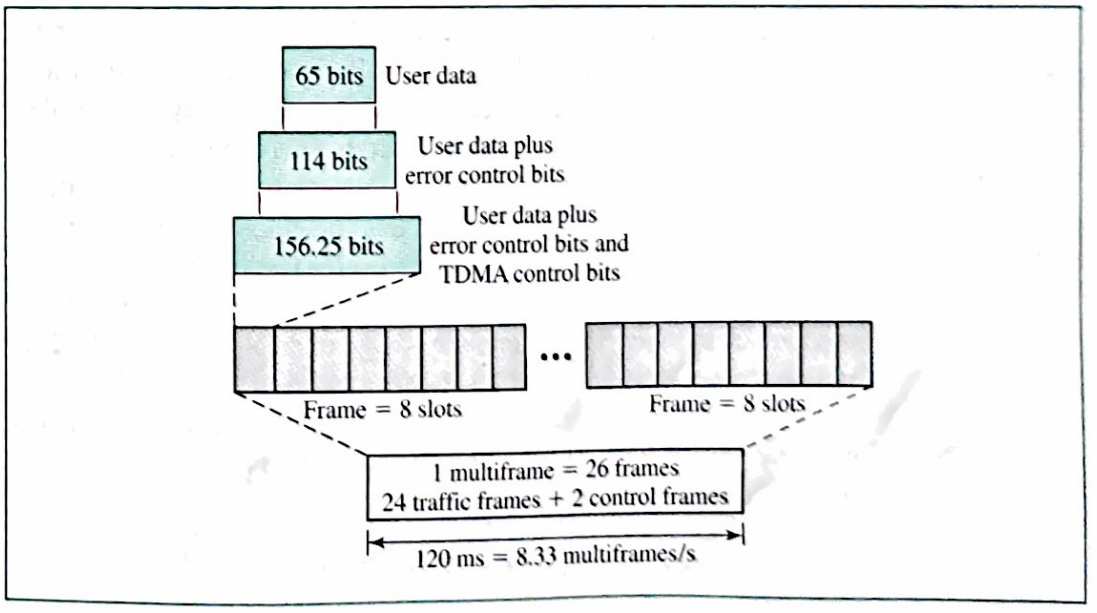
multiplexed together, creating a TDM frame. Twenty-six frames are combined to form a multiframe. We can calculate the bit rate of each channel as follows:

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

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 multiplexed together using FDMA. The result is a 25-MHz band.

Figure 17.9 shows the user data and overhead in a multiframe.

Figure 17.9 *Multiframe components*



The reader may have noticed the large amount of overhead in TDMA. The user data is 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 dominating 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.

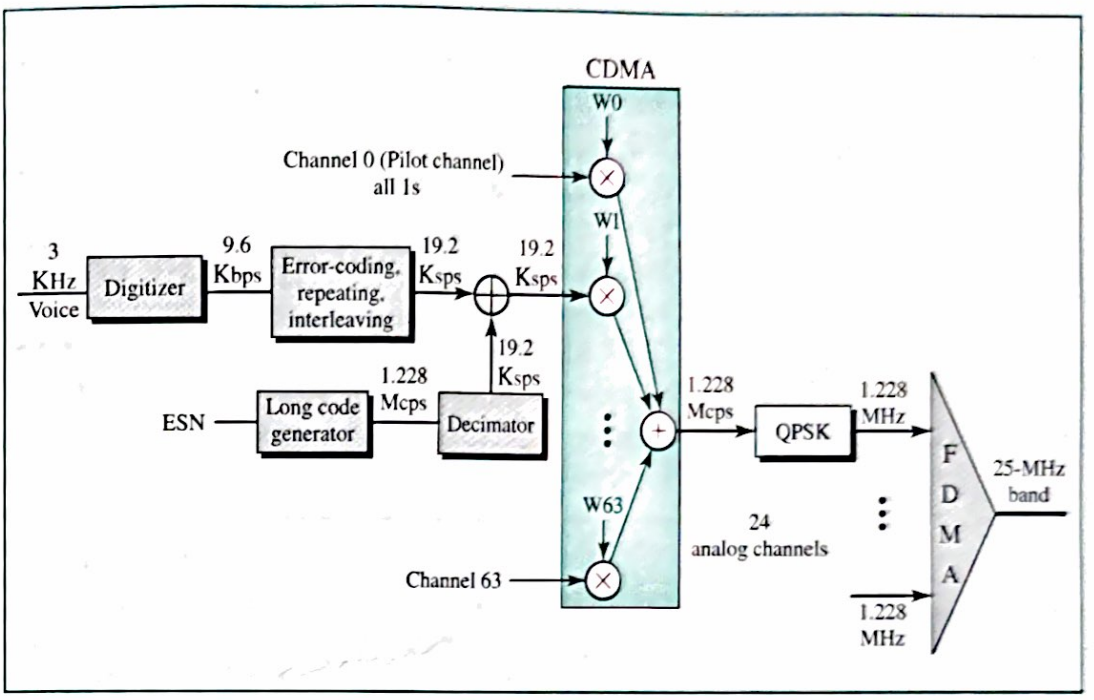
Forward Transmission IS-95 has two different transmission techniques: one for use in the forward (base to mobile) direction and another for use in the reverse (mobile to base) direction. In the forward direction, communications between the base and all mobiles are synchronized; the base sends synchronized data to all mobiles. Figure 17.10 shows a simplified diagram for the forward direction.

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 (kilosignals per seconds). 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 fed to the CDMA multiplexer. For each traffic channel, one Walsh 64 × 64 row chip is selected. The result is a signal of 1.288 Mcps (megachips per second).

$$19.2 \text{ Ksps} \times 64 \text{ cps} = 1.288 \text{ Mcps}$$

Figure 17.10 *IS-95 forward transmission*



The CDMA-multiplexed signal is fed into a QPSK modulator to produce a signal of 1.288 MHz. The resulting bandwidth is shifted appropriately, using FDMA.

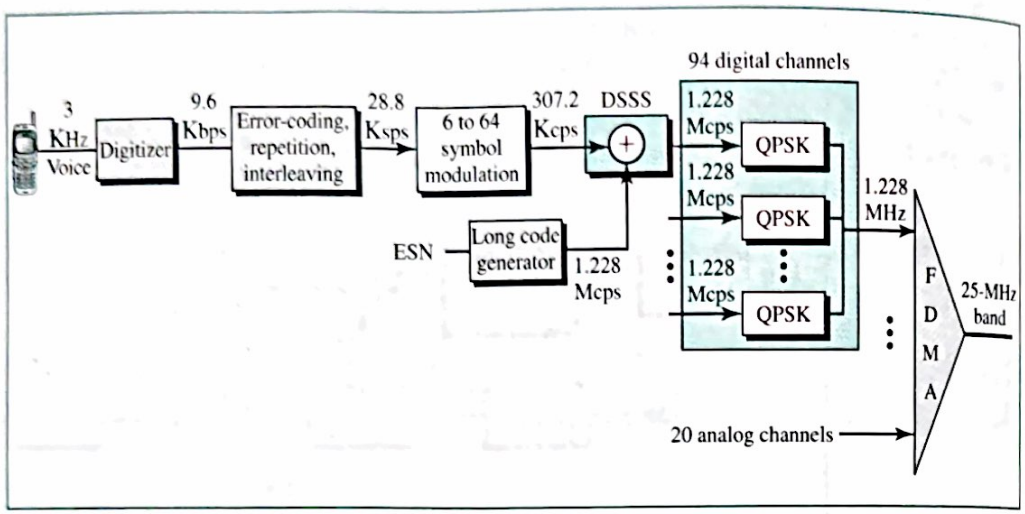
An analog channel creates 64 digital channels, of which 55 channels are traffic channels (carrying digitized voice). Nine channels are used for control and synchronization:

- 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.
- Channel 32 gives information about the system to the mobile station.
- Channels 1 to 7 are used for paging, to send messages to one or more mobile stations.
- 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 15. Figure 17.11 shows a simplified diagram for 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

Figure 17.11 *IS-95 reverse transmission*



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.

PCS

Before we leave the discussion on second-generation cellular telephones, let us explain a term generally heard in relation to this generation: PCS. **Personal Communications System (PCS)** does not refer to a single technology such as GSM, IS-136, or IS-95. It is a generic name for a commercial system that offers several kinds of communication services. Common features of these systems can be summarized:

1. They may use any second-generation technology (GSM, IS-136, or IS-95).
2. They use the 1900-MHz band, which means that a mobile station needs more power because higher frequencies have a shorter range than lower ones. However, since a station's power is limited by the FCC, the base station and the mobile station need to be close to each other (smaller cells).
3. They offer communication services such as short message service (SMS) and limited Internet access.

Third Generation

The third generation of cellular telephony refers to a combination of technologies that provide a variety of services. Ideally, when it matures, the third generation can provide both digital data and voice communication. Using a small portable device, a person should be 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, can download and listen to music, can surf the Internet or play games, can have a video conference, and can 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 **Internet Mobile Communication for year 2000 (IMT-2000).** The blueprint defines some criteria for 3G technology as outlined below:

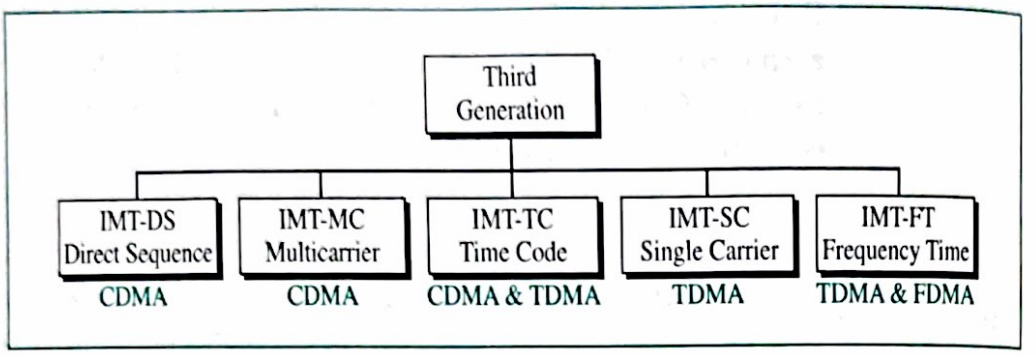
- Voice quality comparable to that of the existing public telephone network.
- Data rate of 144 Kbps for access in a moving vehicle (car), 384 Kbps for access as the user walks (pedestrians), and 2 Mbps for the stationary user (office or home).
- Support for packet-switched and circuit-switched data services.
- A band of 2 GHz.
- Bandwidths of 2 MHz.
- Interface to the Internet.

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

IMT-2000 Radio Interface

Figure 17.12 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 17.12 *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 only uses TDMA.

IMT-FT This standard uses a combination of FDMA and TDMA.

17.2 SATELLITE NETWORKS

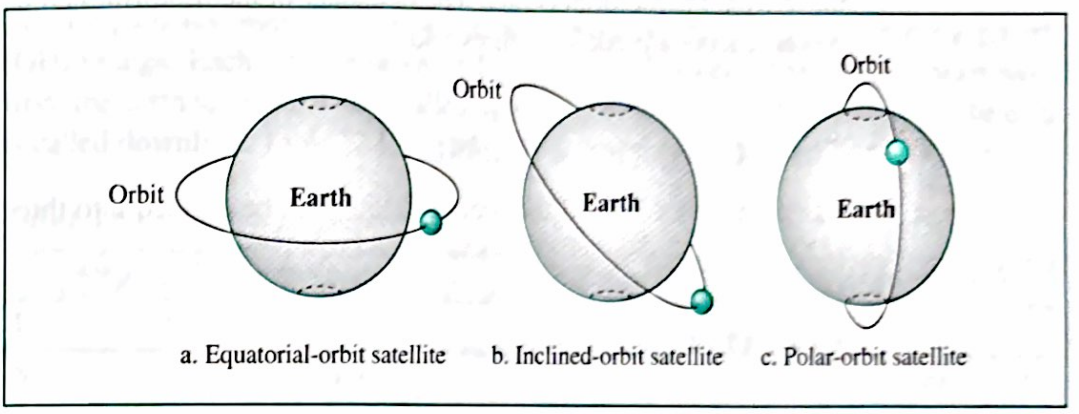
A **satellite network** is a combination of nodes that provides communication from one point on the earth to another. A node in the network can be a satellite, an earth station, or an end-user terminal or telephone. Although a real satellite, such as the moon, can be used as a relaying node in the network, 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 large 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.

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 17.13.

Figure 17.13 *Satellite orbits*



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.

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

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

Example 1

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

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

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

Example 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

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

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 geosynchronous 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