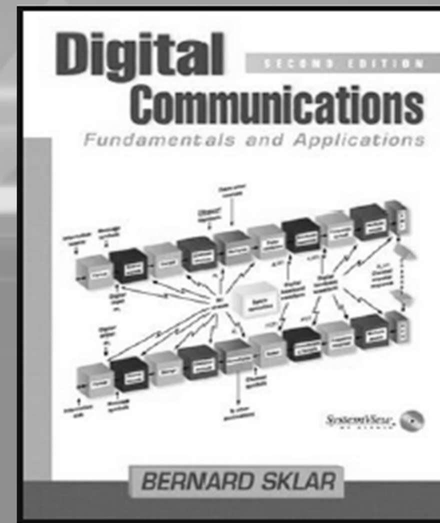


ENE 467
Digital Communications

TEACHING BY

ASST. PROF. SUWAT PATTARAMALAI, PH.D.

11. Multiplexing and Multiple Access

- Outcome
  - Understand the allocation of the communications resource
  - Can design multiplexing techniques (FDMA, TDMA, CDMA, and SDMA)
  - Know how to do multiple access communications system and architecture
  - Can design access algorithms (Aloha, Slotted aloha, Reservation aloha, and polling techniques)
  - Can design multiple access for LAN (CSMA, Ethernet, and token ring)

Multiplexing and Multiple Access

A communications resource (CR) represents the time and bandwidth that is available for communication signaling associated with a given system. It can be graphically envisioned with a plane, where the abscissa represents time, and the ordinate represents frequency. For the efficient development of a communication system, it is important to plan out the resource allocation among system users, so that no block of time/frequency is wasted, and so that the users can share the resource in an equitable manner.

The terms "multiplexing" and "multiple access" refers to the sharing of a CR. There is a subtle difference between multiplexing and multiple access. With *multiplexing,* users' requirements or plans for CR sharing are fixed, or at most, slowly changing. The resource allocation is assigned *a priori,* and the sharing is usually a process that takes place within the confines of a *local site* (e.g., a circuit board). *Multiple access,* however, usually involves the *remote sharing* of a resource, such as in the case of satellite communications. With a dynamically changing multiple access scheme, a system controller must become aware of each user's CR needs; the amount of time required for this information transfer constitutes an overhead and sets an upper limit on the efficiency of the utilization of the CR.

11.1 ALLOCATION OF THE COMMUNICATIONS RESOURCE

There are three basic ways to increase the throughput (total data rate) of a communications resource (CR). The first way is either to increase the transmitter's effective isotropic radiated power (EIRP) or to reduce system losses so that the received E_b/N_0 is increased. The second way is to provide more channel bandwidth. The third approach is to make the allocation of the CR more efficient. This third approach is the domain of communications multiple access. The problem, in the context of a satellite transponder, is to efficiently allocate portions of the transponder's fixed CR to a large number of users who seek to communicate digital information to each other at a variety of bit rates and duty cycles. The basic ways of distributing the communications resource, listed under the heading "multiplexing/multiple access" in Figure 11.1, are the following:

1. *Frequency division (FD).* Specified subbands of frequency are allocated.
2. *Time division (TD).* Periodically recurring time slots are identified. With some systems, users are provided a fixed assignment in time. With others, users may access the resource at random times.
3. *Code division (CD).* Specified members of a set of othogonal or nearly orthogonal spread spectrum codes (each using the full channel bandwidth) are allocated.
4. *Space division (SD) or multiple beam frequency reuse.* Spot beam antennas are used to separate radio signals by pointing in different directions. It allows for reuse of the same frequency band.
5. *Polarization division (PD) or dual polarization frequency reuse.* Orthogonal polarizations are used to separate signals, allowing for reuse of the same frequency band.

The key to *all* multiplexing and multiple access schemes is that various signals share a CR without creating unmanageable interference to each other in the detection process. The allowable limit of such interference is that signals on one CR channel should not significantly increase the probability of error in another channel. Orthogonal signals on separate channels will avoid interference between users. Signal waveforms $x_i(t)$, where $i = 1, 2, \ldots$, are defined to be orthogonal if they can be described in the time domain by

$$\int_{-\infty}^{\infty} x_i(t)x_j(t)\,dt = \begin{cases} K & \text{for } i = j \\ 0 & \text{otherwise} \end{cases} \tag{11.1}$$

where K is a nonzero constant. Similarly, the signals are orthogonal if they can be described in the frequency domain by

$$\int_{-\infty}^{\infty} X_i(f)X_j(f)\,df = \begin{cases} K & \text{for } i = j \\ 0 & \text{otherwise} \end{cases} \tag{11.2}$$

where the functions $X_i(f)$ are the Fourier transforms of the signal waveforms $x_i(t)$. Channelization, characterized by orthogonal waveforms, as shown in Equation (11.1), is called time-division multiplexing or time-division multiple access (TDM/TDMA), and that characterized by orthogonal spectra, as shown in Equation (11.2), is called frequency-division multiplexing or frequency-division multiple access (FDM/FDMA).

Multiplexing and Multiple Access

Multiplexing/Multiple Access

- Frequency division (FDM/FDMA)
- Time division (TDM/TDMA)
- Code division (CDM/CDMA)
- Space division (SDMA)
- Polarization division (PDMA)

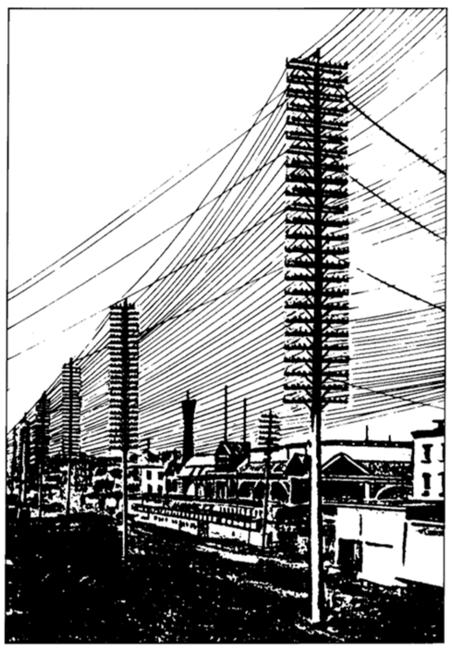


Figure 11.2 In the early days of telephony a pair of wires was needed for each trunk circuit.

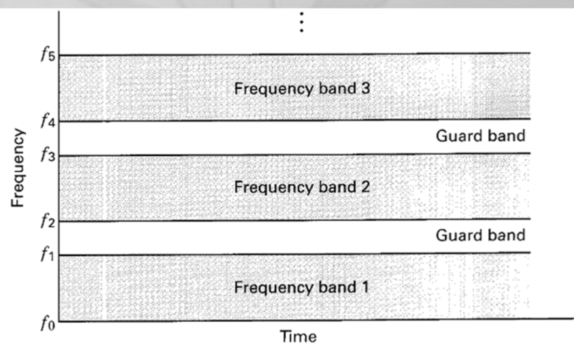
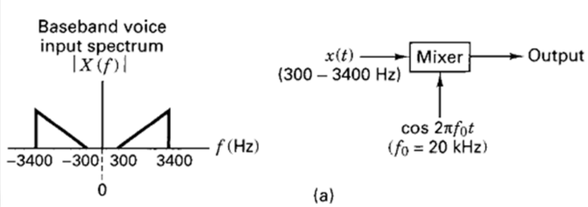


Figure 11.3 Frequency-division multiplexing.



(a)

11.1.1 Frequency-Division Multiplexing/Multiple Access

11.1.1.1 Frequency-Division Mulitplex Telephony

In the early days of telephony, a separate pair of wires was needed for each telephone trunk circuit (trunk circuits interconnect intercity switching centers). As illustrated in Figure 11.2, the skies of all the major cities in the world grew dark with overhead wires as the demand for telephone service grew. A major development in the early 1900s, frequency-division multiplex (FDM) telephony, made it possible to transmit several telephone signals simultaneously on a single wire, and thereby transformed the methods of telephone transmission.

The communications resource (CR) is illustrated in Figure 11.3 as the frequency-time plane. The channelized spectrum shown here is an example of FDM or FDMA. The assignment of a signal or user to a frequency band is *long term* or *permanent;* the CR can simultaneously contain several spectrally separate signals. The first frequency band contains signals that operate between frequencies f_0 and f_1, the second between frequencies f_2 and f_3, and so on. The spectral regions between assignments, called *guard bands,* act as buffer zones to reduce interference between adjacent frequency channels. We might ask: How does one transform a baseband signal so that it occupies a higher frequency band? The answer: *heterodyning* or *mixing,* also called *modulating,* the signal with a fixed frequency from a sine-wave oscillator.

If two input signals to a mixer are sinusoids with frequencies f_A and f_B, the mixing or multiplication will yield new sum and difference frequencies at f_{A+B} and f_{A-B}. The trigonometric identity

$$\cos A \cos B = \tfrac{1}{2}\left[\cos(A+B) + \cos(A-B)\right] \qquad (11.3)$$

describes the effect of the mixer. Figure 11.4a illustrates the mixing of a typical voice-grade telephone signal $x(t)$ (baseband frequency range is 300 to 3400 Hz) with a sinusoid from a 20-kHz oscillator. The baseband two-sided magnitude spectrum, $|X(f)|$, is shown in Figure 11.4a. Can the mixer be a linear device? No. The output signal of a linear device will only consist of the *same* component frequencies as the input signal, differing only in amplitude and/or phase.

Figure 11.4b illustrates the one-sided magnitude spectrum $|X(f - f_0)|$ at the mixer output. As a result of the mixing described by Equation 11.3), the output spectrum is a frequency-upshifted version of the baseband spectrum, centered at the oscillator frequency of 20 kHz. This spectrum is called a *double-sideband* (DSB) *spectrum* because the information appears in two different bands of the positive frequency domain. Figure 11.4c shows the lower sideband (LSB), whose frequency range is 16,600 to 19,700 Hz, the result of filtering the DSB spectrum. This sideband is sometimes referred to as the *inverted sideband* because the order of low-to-high frequency components is the reverse of that of the baseband components. Filtering can similarly be used to separate the upper sideband (USB), whose frequency range is 20,300 to 23,400 Hz, as shown in Figure 11.4d. This sideband is sometimes referred to as the *erect sideband* because the order of the low-to-high frequency components corresponds to that of the baseband components. Each sideband of the DSB spectrum contains the same information. Thus, only one sideband, either the USB or the LSB, is needed in order to retrieve the original baseband data.

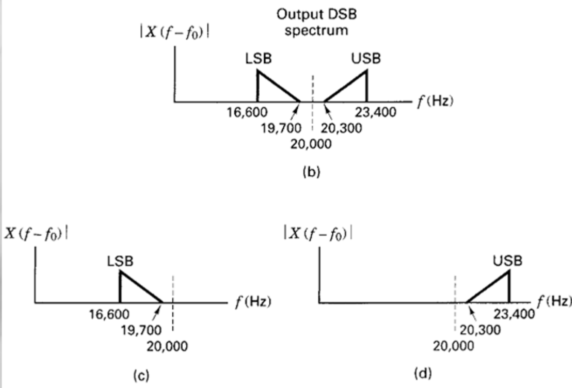


Figure 11.4 Heterodyning (mixing). (a) Mixing operation. (b) Mixer output spectrum. (c) Lower sideband. (d) Upper sideband.

A simple FDM example with three translated voice channels is seen in Figure 11.5. In channel 1, the 300- to 3400-Hz voice signal is mixed with a 20-kHz oscillator. In channels 2 and 3, a similar type of voice signal is mixed with a 16-kHz and 12-kHz oscillator, respectively. Only the lower sidebands are retained; the result of the mixing and filtering (to remove the upper sidebands) yields the frequency-shifted voice channels shown in Figure 11.5. The composite output waveform is just the sum of the three signals, having a total bandwidth in the range 8.6 to 19.7 kHz.

Figure 11.6 illustrates the two lowest levels of the FDM multiplex hierarchy for telephone channels. The first level consists of a *group* of 12 channels modulated onto subcarriers shown in the range 60 to 108 kHz. The second level is made up of five groups (60 channels) called a *supergroup* modulated onto the subcarriers shown in the range 312 go 552 kHz. The multiplexed channels are now treated as a composite signal that can be transmitted over cables or can be further modulated onto a carrier wave for radio transmission.

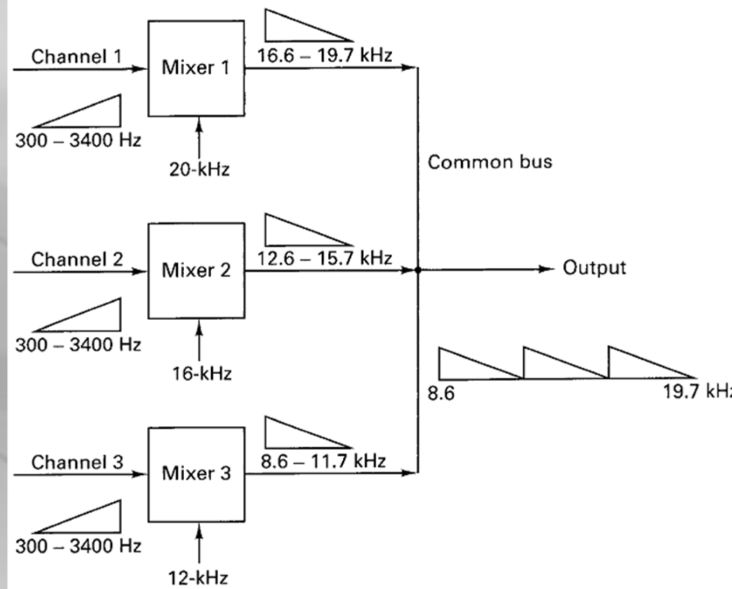


Figure 11.5 Simple FDM example. Three frequency-shifted voice channels.

11.1.1.2 Frequency-Division Multiple Access of Satellite Systems

Most of the world's communication satellites are positioned in a *geostationary* or *geosynchronous* orbit. This means that the satellite is in a circular orbit, in the same plane as the earth's equatorial plane, and at such an altitude (approximately 19,330 nautical miles) that the orbital period is identical with the earth's rotational period. Since such satellites appear stationary when viewed from the earth, three of them spaced 120° apart can provide worldwide coverage (except for the polar regions). Most communication satellite systems are made up of nonregenerative repeaters or transponders. *Nonregenerative* means that the uplink (earth-to-satellite) transmissions are simply amplified, frequency shifted, and retransmitted on the downlink (satellite-to-earth) without any demodulation/remodulation or signal processing. The most popular frequency band for commercial satellite communications, called *C-band*, uses a 6-GHz carrier for the uplink and a 4-GHz carrier for the downlink. For C-band satellite systems, *each satellite* is permitted, by international agreement, to use a 500-MHz-wide spectral assignment. Typically, each satellite has 12 transponders with a bandwidth of 36 MHz each. The most common 36-MHz transponders operate in an FDM/FM/FDMA (frequency-division multiplex, frequency-modulated, frequency-division multiple access) multidestination mode. Let us consider each component of this name:

1. *FDM*. Signals such as telephone signals, each one having a single-sideband 4-kHz spectrum (including guard bands) are FDM'd to form a multichannel composite signal.
2. *FM*. The composite signal is frequency-modulated (FM) onto a carrier and transmitted to the satellite.
3. *FDMA*. Subdivisions of the 36-MHz transponder bandwidth may be assigned to different users. Each user receives a specific bandwidth allocation whereby he or she can access the transponder.

Thus, composite FDM channels are FM modulated and transmitted to the satellite within the bandwidth allocation of an FDMA plan. The major advantage of FDMA (compared with TDMA) is its simplicity. The FDMA channels require no synchronization or central timing; each channel is almost independent of all other channels. Later we discuss some advantages of TDMA compared to FDMA.

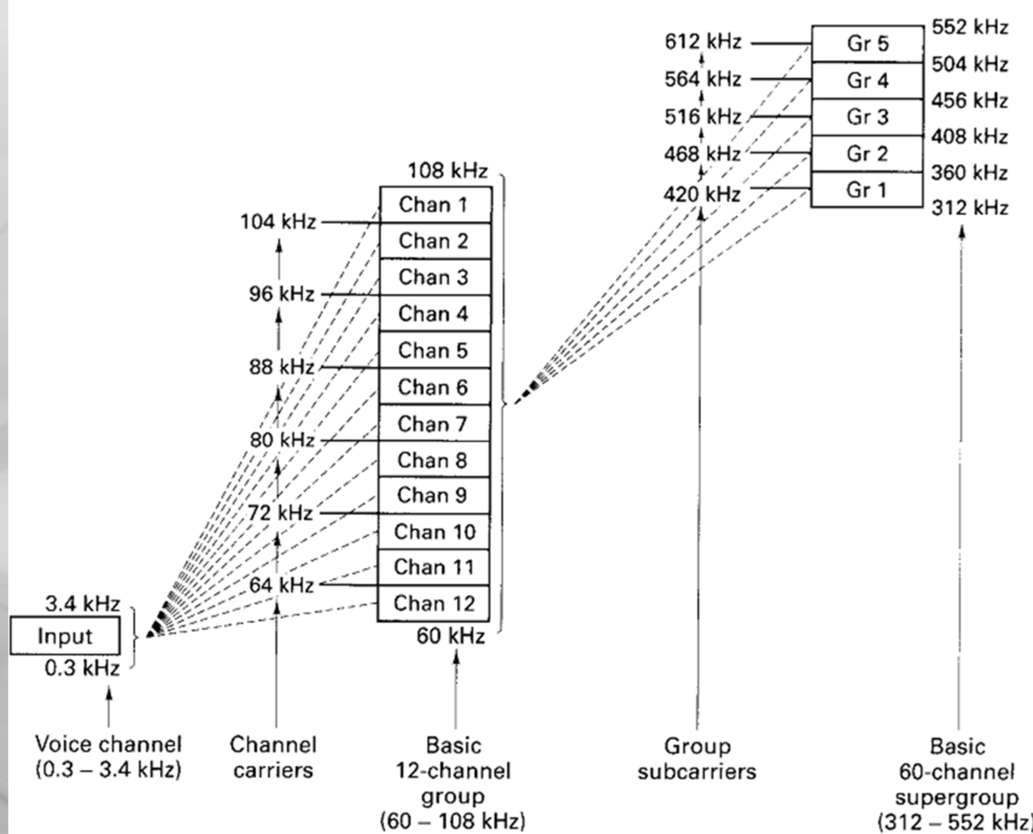


Figure 11.6 Modulation plan of a typical frequency-division multiplex system.

Multiplexing and Multiple Access

11.1.2 Time-Division Multiplexing/Multiple Access

In Figure 11.3, sharing of the communications resource (CR) is accomplished by allocating frequency bands. In Figure 11.7, the same CR is shared by assigning each of M signals or users the full spectral occupancy of the system for a short duration of time called a *time slot*. The unused time regions between slot assignments, called *guard times*, allow for some time uncertainty between signals in adjacent time slots, and thus act as buffer zones to reduce interference. Figure 11.8 is an illustration of a typical TDMA satellite application. Time is segmented into intervals called frames. Each frame is further partitioned into assignable user time slots. The frame structure repeats, so that a fixed TDMA assignment constitutes one or more slots that periodically appear during each frame time. Each earth station transmits its data in bursts, timed so as to arrive at the satellite coincident with its designated time slot(s). When the bursts are received by the satellite transponder, they are retransmitted on the downlink, together with the bursts from other stations. A receiving station detects and demultiplexes the appropriate bursts and feeds the information to the intended user.

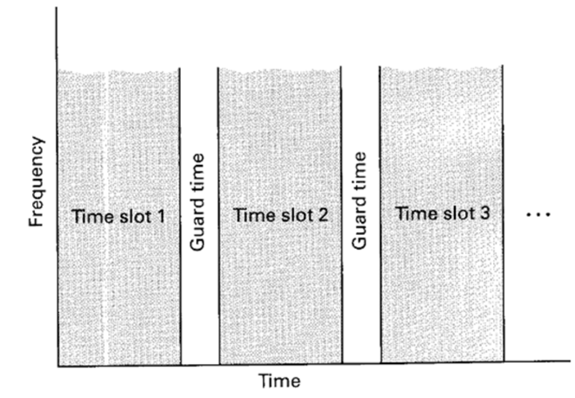


Figure 11.7 Time-division multiplexing.

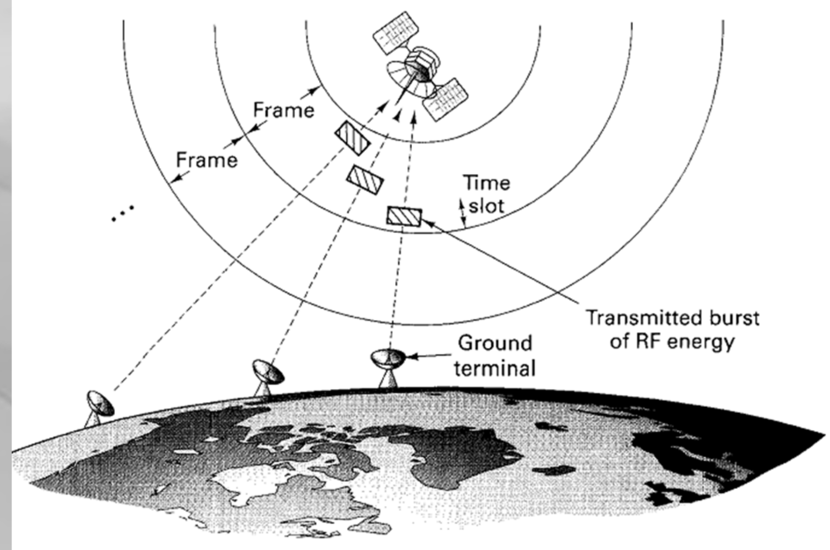


Figure 11.8 Typical TDMA configuration.

Multiplexing and Multiple Access

11.1.2.1 Fixed-Assignment TDM/TDMA

The simplest TDM/TDMA scheme, called *fixed-assignment TDM/TDMA*, is so named because the M time slots that make up each frame are preassigned to signal sources, long term. Figure 11.9 illustrates, in block diagram form, the operation of such a system. The multiplexing operation consists of providing each source with an opportunity to occupy one or more slots. The demultiplexing operation consists of deslotting the information and delivering the data to the intended sink. The two commutating switches in Figure 11.9 have to be synchronized so that the message corresponding to source 1, for example, appears on the channel 1 output, and so on. The message itself generally comprises a preamble portion and a data portion. The preamble portion usually contains synchronization, addressing, and error-control sequences.

A fixed-assignment TDM/TDMA scheme is extremely efficient when the source requirements are predictable, and the traffic is heavy (the time slots are most always filled). However, for bursty or sporadic traffic, the fixed-assignment scheme is wasteful. Consider the simple example shown in Figure 11.10. In this example there are four time slots per frame; each slot is preassigned to users A, B, C, and D, respectively. In Figure 11.10a we see a typical activity profile of the four users. During the first frame time, user C has no data to transmit; during the second frame time, user B has none, and during the third frame time, user A has none. In a fixed-assignment TDMA scheme, all of the slots within a frame are preassigned. If the "owner" of a slot has *no* data to send during a particular frame, that slot is wasted. The data stream, shown in Figure 11.10b, illustrates the wasted time slots in this example. When source requirements are unpredictable, as in this example, there can be more efficient schemes, involving the dynamic assignment of the slots rather than a fixed assignment. Such schemes are variously known as packet-switched systems, statistical multiplexers, or concentrators; the effect, shown in Figure 11.10c is to use all the slots in a frame in such a way that capacity is conserved.

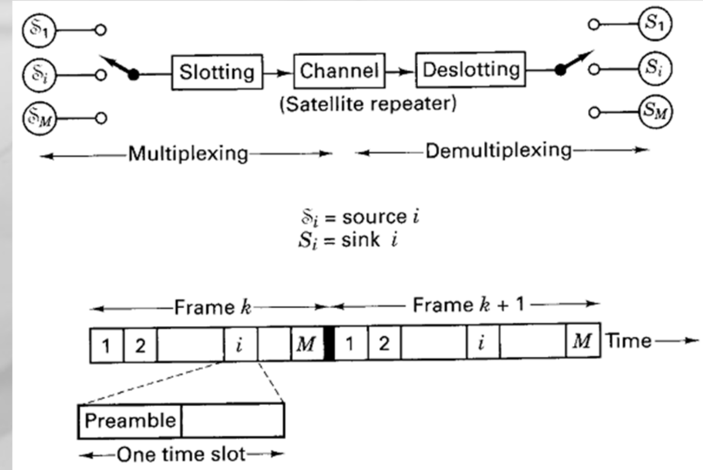


Figure 11.9 Fixed-assignment TDM.

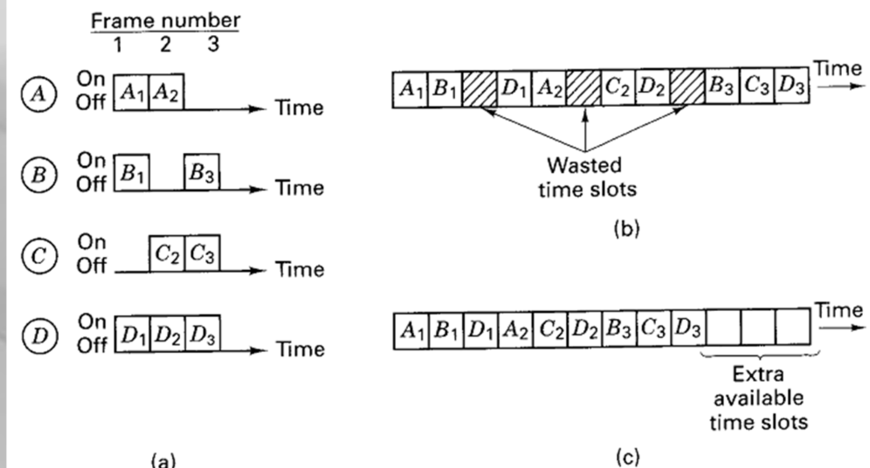


Figure 11.10 Fixed-assignment TDM versus packet switching. (a) Data source activity profiles. (b) Fixed-assignment time-division multiplexing. (c) Time-division packet switching (concentration).

Multiplexing and Multiple Access

11.1.3 Communications Resource Channelization

In Figure 11.3 we considered that the CR is partitioned into spectral bands, and in Figure 11.7 we viewed the same CR as being partitioned into time slots. Figure 11.11 represents a more general organization of the CR allowing for the assignment of a frequency band for a prescribed period of time. Such a multiple access scheme is referred to as *combined FDMA/TDMA*. For the assignments of frequency bands, let us assume an equal apportionment of the total bandwidth W, among M user groups or classes, so that M disjoint frequency bands of width W/M hertz are continuously available to their assigned group. Similarly, for the assignment of time slots, the time axis is partitioned into time frames, each of duration T, and the frames are partitioned into N slot times, each of duration T/N. We assume that the users are time synchronized and that the assigned slots are located periodically within the frames. Each user in each frequency band is permitted to transmit during each periodic appearance of the user's assigned slot, and is permitted to use the assigned channel bandwidth for the slot duration. A slot is uniquely determined as the mth slot within the nth frame. Referring to Figure 11.11, we can describe the time of a particular slot (n, m) with reference to time zero as follows:

$$\text{time of slot } (n, m) = nT + \frac{(m-1)T}{N} \leq t \leq nT + \frac{mT}{N}$$

$$n = 0, 1, \ldots; m = 1, 2, \ldots, N. \tag{11.4}$$

The nth frame time, T, is denoted by the time interval $[nT, (n+1)T]$. As can be seen in Figure 11.11, the domain of the unit signal is the intersection of the time slot (n, m) and the frequency band (j). Assume that a modulation/coding system is chosen so that the full bandwidth W of the CR can support R bits/s. In any frequency band having a bandwidth of W/M hertz, the associated bit rate will be R/M bits/s. FDMA alone would provide M bands each with a bandwidth of $1/M$ of the full bandwidth of the CR. TDMA alone would provide the full system bandwidth for each of the N slots, where the duration of each slot is $1/N$ of the frame time.

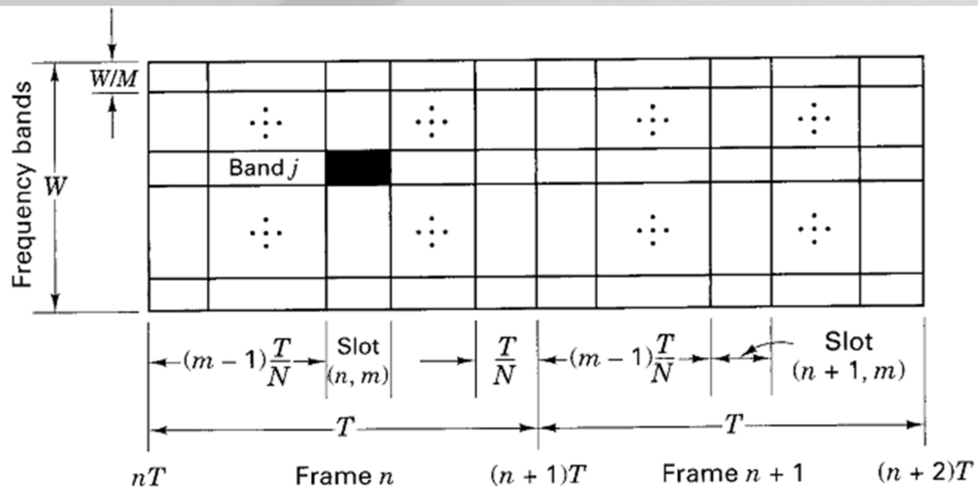


Figure 11.11 Communications resource: time/frequency channelization.

Multiplexing and Multiple Access

11.1.4 Performance Comparison of FDMA and TDMA

11.1.4.1 Bit Rate Equivalence of FDMA and TDMA

Figure 11.12 highlights the basic differences between an FDMA and TDMA system in a communications resource capable of supporting a total of R bits/s. In Figure 11.12a the system bandwidth is divided into M orthogonal frequency bands.

Hence each of the M sources, $\mathcal{S}_m (1 \leq m \leq M)$ can simultaneously transmit at a bit rate of R/M bits/s. In Figure 11.12b the frame is divided into M orthogonal time slots. Hence each of the M sources bursts its transmission at R bits/s, M times faster than the equivalent FDMA user for $(1/M)$th the time. In both cases, the source $\mathcal{S}_m$ transmits information at an average rate of R/M bits/s.

Let the information generated by each of the sources in Figure 11.12 be organized into b-bit groups, or *packets*. In the case of FDMA, the b-bit packets are transmitted in T seconds over each of the M disjoint channels. Therefore, the total bit rate required is

$$R_{\mathrm{FD}} = M \frac{b}{T} \quad \text{bits/s} \tag{11.5}$$

In the case of TDMA, the b bits are transmitted in T/M seconds from each source. Therefore, the bit rate required is

$$R_{\mathrm{TD}} = \frac{b}{T/M} \quad \text{bits/s} \tag{11.6}$$

Since Equations (11.5) and (11.6) yield identical results, we can conclude that

$$R_{\mathrm{FD}} = R_{\mathrm{TD}} = R = \frac{Mb}{T} \quad \text{bits/s} \tag{11.7}$$

Thus, both systems require the same full CR data rate, R bits/s.

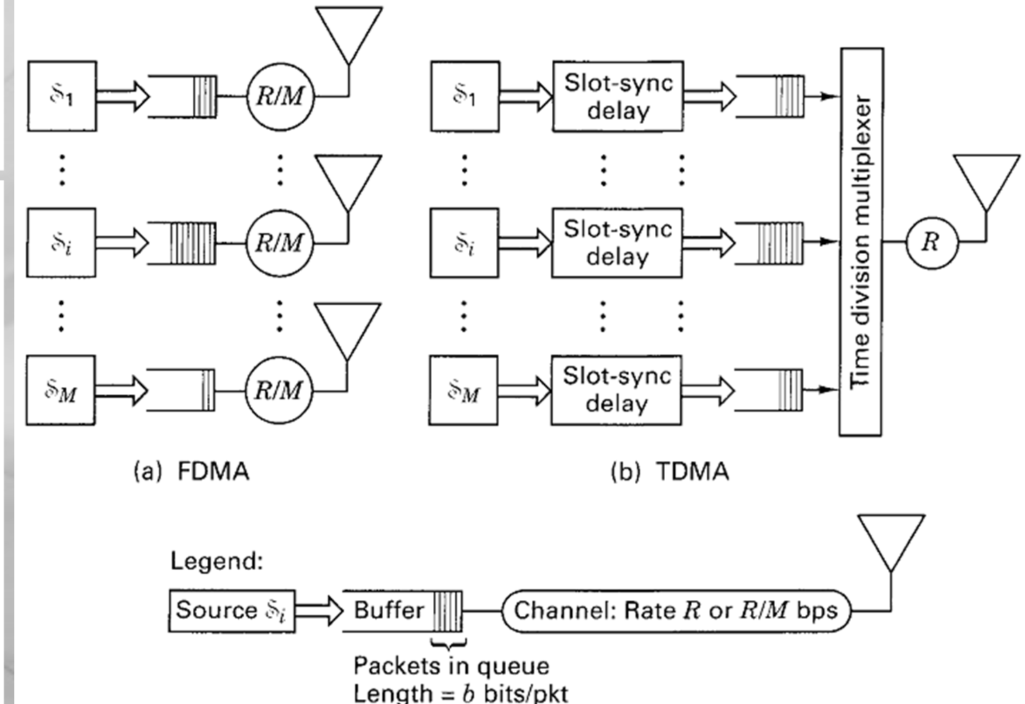


Figure 11.12 (a) FDMA: frequency divided into M orthogonal frequency bands. (b) TDMA: time divided into M orthogonal time slots (one packet per time slot).

11.1.4.2 Message Delays in FDMA and TDMA

From the previous sections it might appear that the duality between FDMA and TDMA will result in equivalent performance. This is not the case when the metric of performance is the average packet *delay*. It can be shown [1, 2] that TDMA is inherently superior to FDMA in the sense that the average packet delay using TDMA is less than the delay using FDMA.

As before, we assume that in the case of FDMA the system bandwidth is divided into M orthogonal frequency bands, and in the case of TDMA the frame is divided into M orthogonal time slots. For the analysis of message delay, the simplest case is that of deterministic data sources. It is assumed that the CR is 100% utilized, so that all frequency bands in the case of FDMA, and all time slots in the case of TDMA, are filled with data packets. For simplicity, it is also assumed that there are *no* overhead costs such as guard bands or guards times. The message delay can be defined as

$$D = w + \tau \tag{11.8}$$

where w is the average packet waiting time (prior to transmission) and τ is the packet transmission time. In the FDMA case, each packet is sent over a T-second interval, so the packet transmission time for FDMA is simply

$$\tau_{FD} = T \tag{11.9}$$

In the TDMA case, each packet is sent in slots of T/M seconds. We can thus write the TDMA packet transmission time with the use of Equation (11.7) as

$$\tau_{TD} = \frac{T}{M} = \frac{b}{R} \tag{11.10}$$

Since the FDMA channel is continuously available and packets are sent as soon as they are generated, the waiting time, w_{FD}, for FDMA is

$$w_{FD} = 0 \tag{11.11}$$

FDMA and TDMA bit streams are compared in Figure 11.13. For TDMA, Figure 11.13a illustrates that each user's slot begins at a different point in the T-second frame; that is, packet S_{mk} will start at $(m - 1)T/M$ seconds $(1 \le m \le M)$ after the packet generation instant. Therefore, the average waiting time that a TDMA packet sustains before transmission begins is

$$\begin{aligned} w_{TD} &= \frac{1}{M}\sum_{m=1}^{M}(m-1)\frac{T}{M} = \frac{T}{M^2}\sum_{n=0}^{M-1} n = \frac{T}{M^2}\frac{(M-1)(M)}{2} \\ &= \frac{T}{2}\left(1 - \frac{1}{M}\right) \end{aligned} \tag{11.12}$$

The maximum waiting time before transmission of a packet is $(M - 1)T/M$ seconds, and on the average a packet will wait $\frac{1}{2}(M - 1)(T/M) = (T/2)(1 - 1/M)$ seconds, as given by Equation (11.12).

To compare the average delay times, D_{FD} and D_{TD}, for FDMA and TDMA, respectively, we combine Equations (11.9) and (11.11) into Equation (11.8), and similarly combine Equations (11.10) and (11.12) into Equation (11.8), yielding

$$D_{FD} = T \tag{11.13}$$

$$D_{TD} = \frac{T}{2}\left(1 - \frac{1}{M}\right) + \frac{T}{M} = D_{FD} - \frac{T}{2}\left(1 - \frac{1}{M}\right) \tag{11.14}$$

Using Equation (11.7), Equation (11.14) can be written as

$$D_{TD} = D_{FD} - \frac{b}{2R}(M - 1) \tag{11.15}$$

The result indicates that TDMA is inherently superior to FDMA, from a message-delay point of view. Although Equation (11.15) assumed that the data source is deterministic, the smaller average message delays for TDMA schemes hold up for any independent message arrival process [1, 2].

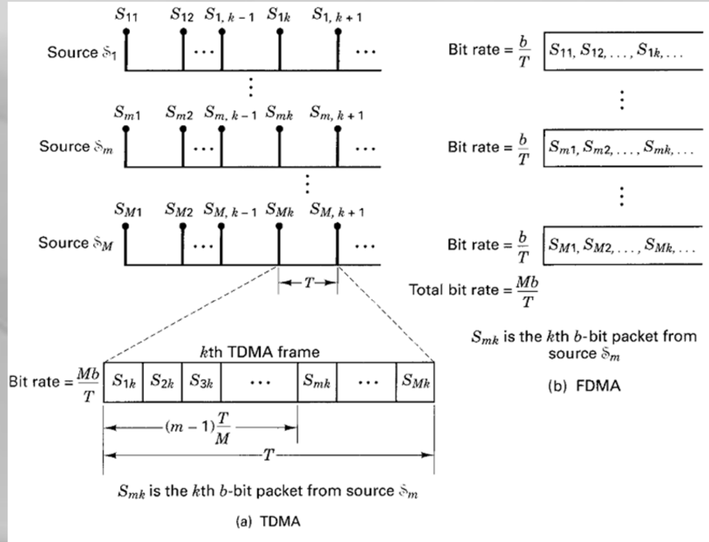


Figure 11.13 (a) TDMA and (b) FDMA channelization.

11.1.5 Code-Division Multiple Access

In Figure 11.3 the CR plane was illustrated as being shared by slicing it horizontally to form FDMA frequency bands, and in Figure 11.7 the same CR plane was illustrated as being shared by slicing it vertically to form TDMA time slots. These two techniques are the most common choices for multiple access applications. Figure 11.14 illustrates the CR being partitioned by the use of a hybrid combination of FDMA and TDMA known as *code-division multiple access* (CDMA). CDMA is an application of spread-spectrum (SS) techniques. Spread-spectrum techniques can be classified into two major categories: *direct-sequence* SS and *frequency hopping* SS. We introduce frequence hopping CDMA (FH-CDMA) in this chapter, and we treat direct-sequence CDMA together with the overall subject of spread-spectrum techniques in Chapter 12.

It is easiest to visualize *frequency hopping* CDMA, illustrated in Figure 11.14, as the short-term assignment of a frequency band to various signal sources. At each successive time slot, whose duration is usually brief, the frequency band assignments are reordered. In Figure 11.14, during time slot 1, signal 1 occupies band 1, signal 2 occupies band 2, and the signal 3 occupies band 3. During time slot 2, signal 1 hops to band 3, signal 2 hops to band 1, and signal 3 hops to band 2, and so on. The CR can thus be fully utilized, but the participants, having their frequency bands reassigned at each time slot, appear to be playing "musical chairs." Each user employs a pseudonoise (PN) code, orthogonal (or nearly orthogonal) to all the other user codes, that dictates the frequency hopping band assignments. Details of PN code sequences are treated in Section 12.2. Figure 11.14 is an oversimplified view of the way the CR is shared in frequency hopping CDMA, since the symmetry implies that each frequency hopping signal is in time synchronism with each of the other signals. This is *not the case.* In fact, one of the attractions of CDMA compared to TDMA is that there is no need for synchronization among user groups (only between a transmitter and a receiver within a group).

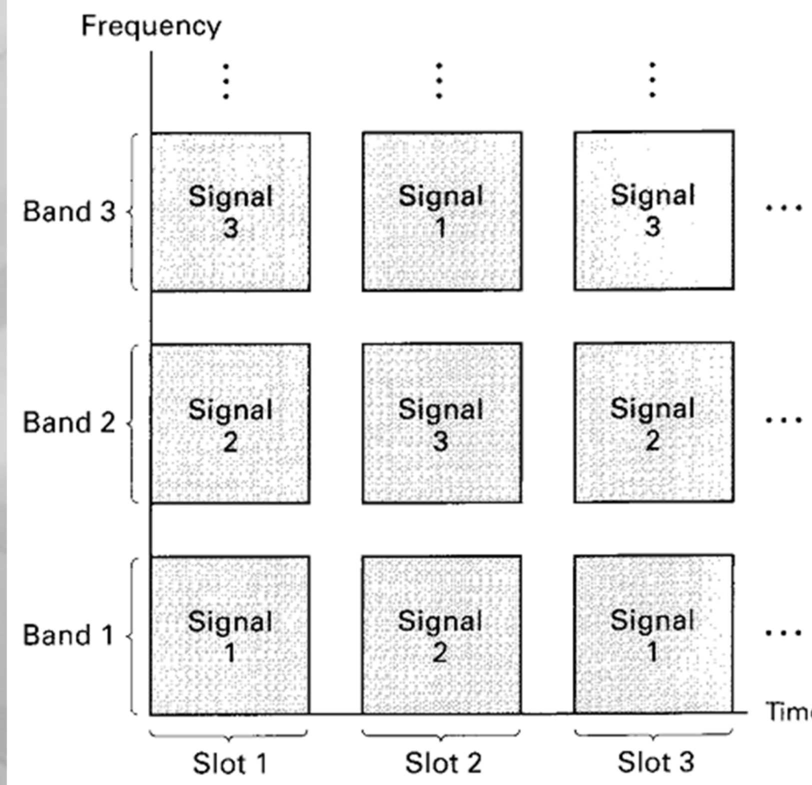


Figure 11.14 Code-division multiplexing.

The block diagram in Figure 11.15 illustrates the frequency hopping modulation process. At each frequency hop time the PN generator feeds a code sequence to a device called a *frequency hopper*. The frequency hopper synthesizes one of the allowable hop frequencies. Assume that the data modulation has an M-ary frequency shift keying (MFSK) format. The essential difference between a conventional MFSK system and a frequency hopping (FH) MFSK system is that in the conventional system, a data symbol modulates a carrier wave that is *fixed* in frequency, but in the hopping system, the data symbol modulates a carrier wave that *hops* across the total CR bandwidth. The FH modulation in Figure 11.15 can be thought of as a two-step process—data modulation and frequency hopping modulation—even though it can be implemented in a single step, where the modulator produces a transmission tone based on the simultaneous dictates of the PN code and the data. Frequency hopping systems are covered in detail in Section 12.4.

One might ask: Don't the FDMA and TDMA options provide sufficient multiple access flexibility? FDMA and TDMA methods can surely be relied on to apportion the communications resource equitably. Of what use is this hybrid technique? CDMA offers some unique advantages, as follows:

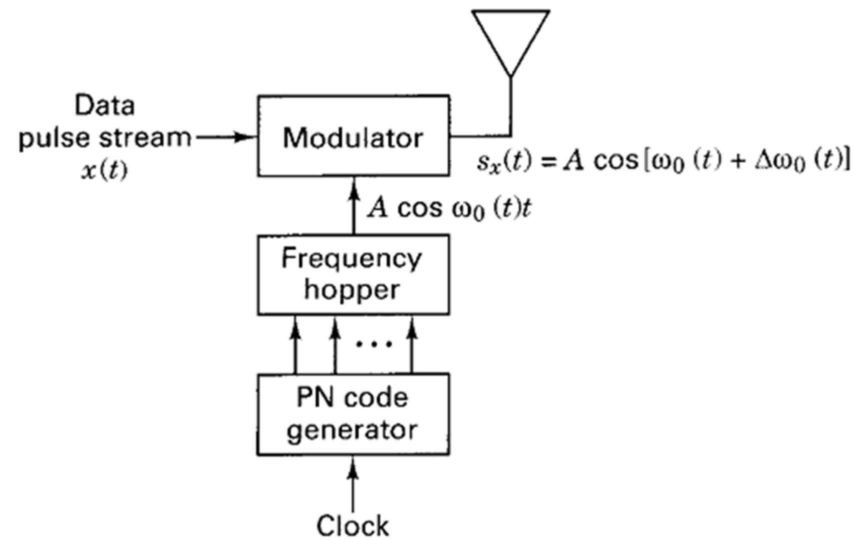


Figure 11.15 CDMA frequency hopping modulation process.

1. *Privacy.* When the code for a particular user group is only distributed among authorized users, the CDMA process provides communications privacy, since the transmissions cannot easily be intercepted by unauthorized users without the code.
2. *Fading channels.* If a particular portion of the spectrum is characterized by fading, signals in that frequency range are attenuated. In an FDMA scheme, a user who was unfortunate enough to be assigned to the fading position of the spectrum might experience highly degraded communications for as long as the fading persists. However, in a FH-CDMA scheme, only during the time a user hops into the affected portion of the spectrum will the user experience degradation. Therefore, with CDMA, such degradation is shared among all the users.
3. *Jam resistance.* During a given CDMA hop, the signal bandwidth is identical to the bandwidth of conventional MFSK, which is typically equal to the minimum bandwidth necessary to transmit the MFSK symbol. However, over a duration of many time slots, the system will hop over a frequency band which is much wider than the data bandwidth. We refer to this utilization of bandwidth as spread spectrum. In Chapter 12 we develop, in detail, the resistance to jamming that spread spectrum affords a user.
4. *Flexibility.* The most important advantage of CDMA schemes, compared to TDMA, is that there need be no precise time coordination among the various simultaneous transmitters. The orthogonality between user transmissions on different codes is not affected by transmission-time variations. This will become clear upon closer examination of the autocorrelation and cross-correlation properties of the codes, considered in Chapter 12.

11.1.6 Space-Division and Polarization-Division Multiple Access

Figure 11.16a depicts the INTELSAT IVA application of space-division multiple access (SDMA), also called *multiple-beam frequency reuse.* INTELSAT IVA used a dual-beam receive antenna feeding two receivers to allow simultaneous access of the satellite from two different regions of the earth. The frequency band allocated to each receive beam was identical because the uplink signals were spatially separated. In such cases, the frequency band is said to be *reused.*

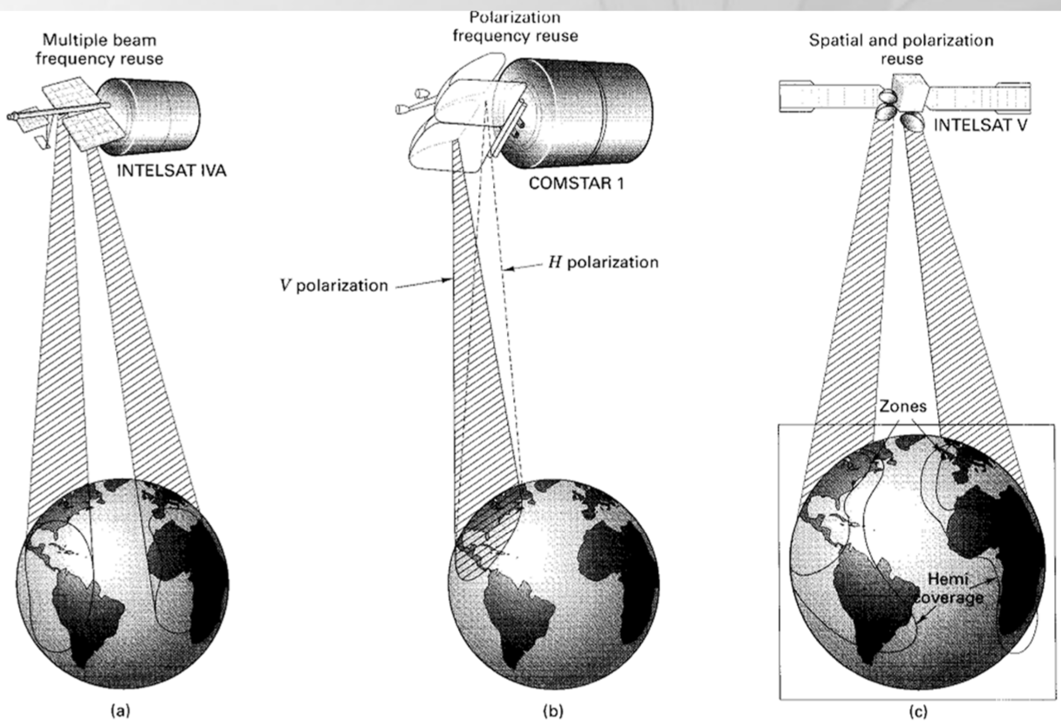


Figure 11.16 SDMA and PDMA. (a) INTELSAT IVA. (b) COMSTAR 1. (c) INTELSAT V (Atlantic coverage).

Figure 11.16b depicts an application of polarization-division multiple access (PDMA), also called *dual-polarization frequency reuse,* from COMSTAR 1. Here separate antennas were used, each with different polarization and followed by separate receivers, allowing simultaneous access of the satellite from the same region of the earth. Each corresponding earth station antenna needs to be polarized in the same way as its counterpart in the satellite. (This is generally accomplished by providing each participating earth station with an antenna that has dual polarization.) The frequency band allocated to each antenna beam could be identical because the uplink signals were orthogonal in polarization. As with SDMA, the frequency band in PDMA is said to be reused. Figure 11.16c depicts an application of the simultaneous use of SDMA and PDMA in INTELSAT V. There are two separate hemispheric coverages, west and east. There are also two smaller zone beams; each zone beam overlaps a portion of one of the hemispheric beams and is separated from it by orthogonal polarization. Thus, there is a fourfold reuse of the spectrum.

Multiplexing and Multiple Access

11.2 MULTIPLE ACCESS COMMUNICATIONS SYSTEM AND ARCHITECTURE

A *multiple access protocol* or *multiple access algorithm* (MAA) is that rule by which a user knows how to use time, frequency, and code functions to communicate through a satellite to other users. A multiple access system is a combination of hardware and software that supports the MAA. The general goal of a multiple access system is to provide communications service in a timely, orderly, and efficient way.

Figure 11.17 illustrates some basic choices for the architecture of a satellite multiple access system. The legend indicates the symbols used for an earth station with and without an MAA controller, and a satellite with and without an MAA controller. Figure 11.17a illustrates the case where one earth station is designated as the master, or the controller. This earth station possesses an MAA computer and responds to the service requests of all other users. Notice that a user's request entails a transmission through the satellite and back down to the controller. The controller's response entails another transmission through the satellite; hence there are two up- and downlink transmissions required for each service assignment. Figure 11.17b illustrates the case where the MAA control is distributed among all the earth stations; there is no single controller. Each earth station uses the same algorithm and they each have identical knowledge regarding access requests and assignments; therefore, only one round trip is required for each service assignment. Figure 11.17c illustrates the case where the MAA controller is in the satellite. A service request goes from user to satellite, and the response from the satellite can follow immediately; therefore, only one round trip is required for each service assignment.

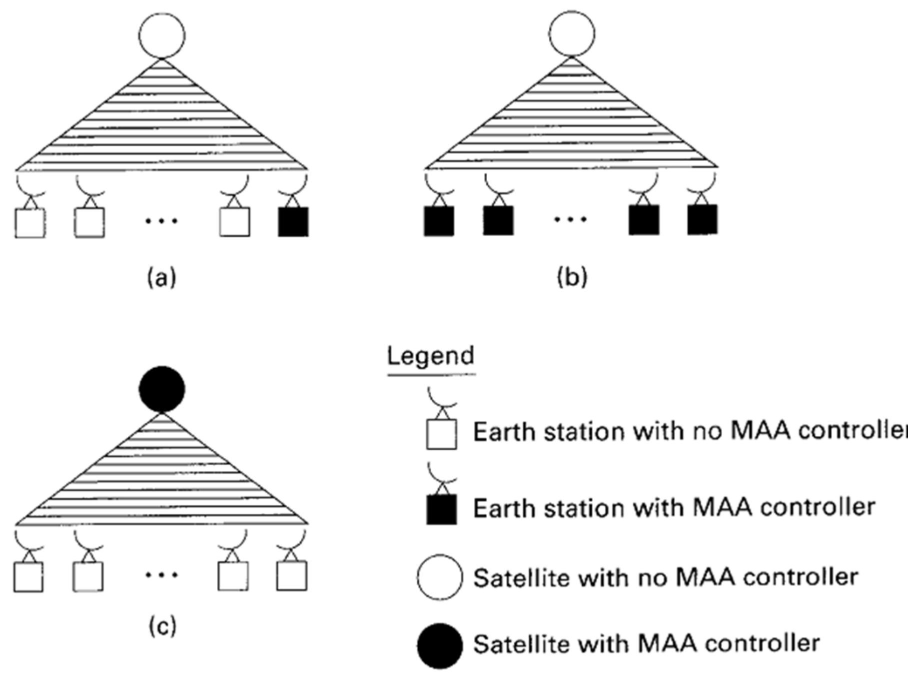


Figure 11.17 Satellite multiple access architecture. (a) Single earth station control. (b) Distributed earth station control. (c) Satellite control.

Multiplexing and Multiple Access

11.2.1 Multiple Access Information Flow

Figure 11.18 is a flow diagram describing the basic flow of information between the multiple access algorithm (MAA) or controller and an earth station; the numbers below correspond to those on the figure. Recall from the preceding section that the control may be lodged in the satellite, in a master station, or distributed among all the earth stations. The flow proceeds as follows:

1. *Channelization.* This term refers to the most general allocation information [e.g., channels 1 to N may be allocated for the Army and channels $(N+1)$ to M for the Navy]. This information seldom changes, and may be distributed to the earth stations by the use of a newsletter rather than via the communication system.
2. *Network state (NS).* This term refers to the state of the CR. A station is advised regarding the availability of the communications resource and where in the resource (e.g., time, frequency, code position) to transmit its service request(s).
3. *Service request.* Then the station makes its request(s) for service (e.g., allocation for m message slots).
4. Upon receipt of the service request(s), the controller sends the station a schedule regarding where and when to position its data in the CR.
5. The station transmits its data according to its assigned schedule.

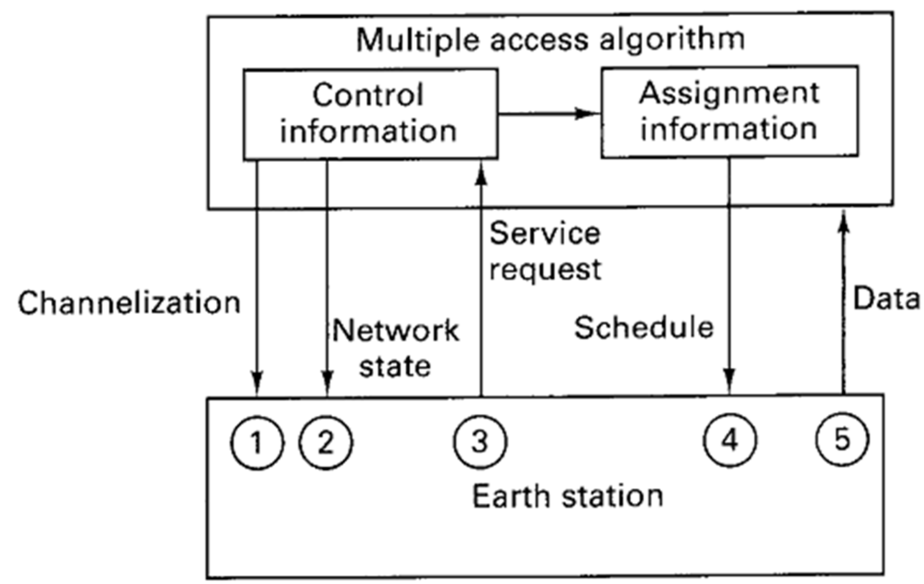


Figure 11.18 Multiple access information flow.

Multiplexing and Multiple Access

11.2.2 Demand-Assignment Multiple Access

Multiple access schemes are termed *fixed assignment* when a station has periodic access to the channel independent of its actual need. By comparison, dynamic assignment schemes, sometimes called *demand-assignment multiple access* (DAMA), give the station access to the channel only when it requests access. If the traffic from a station tends to be burst-like or intermittent, DAMA procedures can be much more efficient than fixed-assignment procedures. A DAMA scheme capitalizes on the fact that actual demand *rarely* equals the peak demand. If a system's capacity is equal to the total peak demand and if the traffic is bursty, the system will be underutilized most of the time. However, by using buffers and DAMA, a system with reduced average capacity can handle bursty traffic, at the cost of some queueing delay. Figure 11.19 summarizes the difference between a fixed system, whose capacity is equal to the sum of the user requirements, and a dynamic system, whose capacity is equal to the average of the user requirements.

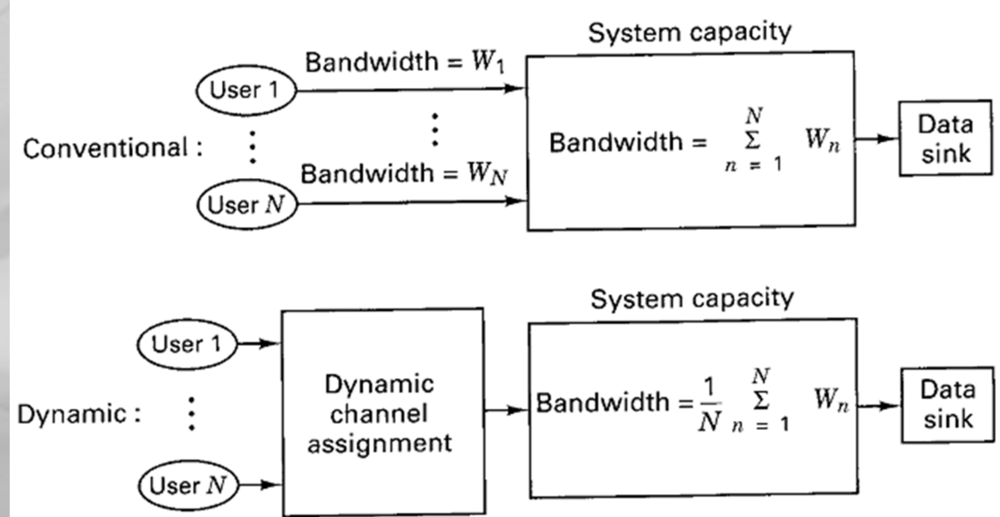


Figure 11.19 Bandwidth reduction for systems using dynamic channel assignment.

11.3 ACCESS ALGORITHMS

11.3.1 Aloha

In 1971, the University of Hawaii began operation of its ALOHA system. A communication satellite was used to interconnect the several university computers by use of a random access protocol [3–7]. The system concept was extremely simple, consisting of the following modes:

1. *Transmission mode.* Users transmit at any time they desire, encoding their transmissions with an error detection code.
2. *Listening mode.* After a message transmission, a user listens for an acknowledgment (ACK) from the receiver. Transmissions from different users will sometimes overlap in time, causing reception errors in the data in each of the contending messages. We say that the messages have *collided.* In such cases, the errors are detected, and the users receive a negative acknowledgment (NAK).
3. *Retransmission mode.* When a NAK is received, the messages are simply retransmitted. Of course, if the colliding users were to retransmit immediately, they would collide again. Therefore, the users retransmit after a *random* delay.
4. *Timeout mode.* If, after a transmission, the user does not receive either an ACK or NAK within a specified time, the user retransmits the message.

11.3.1.1 Message Arrival Statistics

Assume that the total system demand requires an average message or packet arrival rate of λ successful or accepted messages per second. Because of the presence of collisions, some of the messages will be unsuccessful or rejected. Therefore, we define the total traffic arrival rate λ_t, as the acceptance rate λ, plus the rejection rate λ_r, as follows:

$$\lambda_t = \lambda + \lambda_r \tag{11.16}$$

Let us denote the length of each message or packet as b bits. Then we can define the average amount of successful traffic or *throughput,* ρ', on the channel in units of bits per second, as

$$\rho' = b\lambda \tag{11.17}$$

We can also define the *total traffic,* G', on the channel, in units of bits per second, as

$$G' = b\lambda_t \tag{11.18}$$

With the channel capacity (maximum bit rate) designated as R bits per second, let us further define a *normalized throughput*

$$\rho = \frac{b\lambda}{R} \tag{11.19}$$

and a *normalized total traffic*

$$G = \frac{b\lambda_t}{R} \tag{11.20}$$

Normalized throughput, ρ, expresses throughput as a fraction $(0 \le \rho \le 1)$ of channel capacity. Normalized total traffic, G, expresses total traffic as a fraction $(0 \le G \le \infty)$ of the channel capacity. Notice that G can take on values greater than unity.

We can also define the transmission time of each packet as

$$\tau = \frac{b}{R} \quad \text{seconds/packet} \tag{11.21}$$

By substituting Equation (11.21) into Equations (11.19) and (11.20), we can write

$$\rho = \lambda\tau \tag{11.22}$$

and

$$G = \lambda_t \tau \tag{11.23}$$

A user can successfully transmit a message as long as no other user began one within the previous τ seconds or starts one within the next τ seconds. If another user began a message within the previous τ seconds, its tail end will collide with the current message. If another user begins a message within the next τ seconds, it will collide with the tail end of the current message. Thus a space of 2τ seconds is needed for each message.

The message arrival statistics for unrelated users of a communication system is often modeled as a Poisson process. The probability of having K new messages arrive during a time interval of τ seconds is given by the Poisson distribution [8] as

$$P(K) = \frac{(\lambda\tau)^K e^{-\lambda\tau}}{K!} \qquad K \geq 0 \tag{11.24}$$

where λ is the average message arrival rate. Because the users transmit without regard for each other in the ALOHA system, this expression is useful for calculating the probability that exactly $K = 0$ other messages are transmitted during a time interval 2τ. This is the probability, P_s, that a user's message transmission was successful (experienced no collisions). To compute P_s, assuming that all traffic is Poisson, we use λ_t and 2τ in Equation 11.24. Thus,

$$P_s = P(K=0) = \frac{(2\tau\lambda_t)^0 e^{-2\tau\lambda_t}}{0!} = e^{-2\tau\lambda_t} \tag{11.25}$$

In Equation (11.16) we defined the total traffic arrival rate λ_t in terms of the successful portion λ and the repetition or unsuccessful portion λ_r; then, by definition, the probability of a successful packet can be expressed as

$$P_s = \frac{\lambda}{\lambda_t} \tag{11.26}$$

By combining Equations (11.25) and (11.26), we have

$$\lambda = \lambda_t e^{-2\tau\lambda_t} \tag{11.27}$$

By combining Equation (11.27) with Equations (11.22) and (11.23), we can write

$$\rho = Ge^{-2G} \tag{11.28}$$

Equation (11.28) relates the normalized throughput, ρ, to the normalized total traffic, G, on the channel for the ALOHA system. A plot of this relationship labeled "pure ALOHA" is shown in Figure 11.20; as G increases, ρ increases until a point is reached where further traffic increases create a large enough collision rate to cause a reduction in the throughput. The maximum ρ, equal to $1/2e = 0.18$, occurs at a value of $G = 0.5$. Therefore, for a pure ALOHA channel, only 18% of the CR can be utilized. Simplicity of control is achieved at the expense of channel capacity [7, 9].

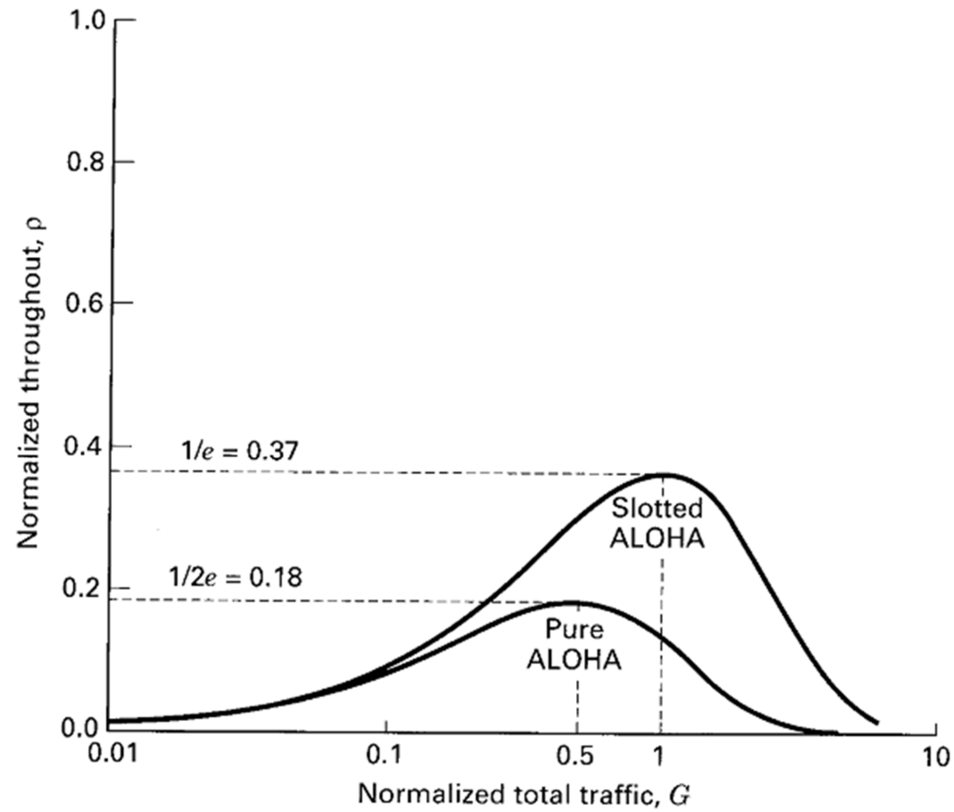


Figure 11.20 Throughput in ALOHA channels (successful transmissions versus total transmissions).

11.3.2 Slotted ALOHA

The pure ALOHA scheme can be improved by requiring a small amount of coordination among the stations. The slotted ALOHA (S-ALOHA) is such a system. A sequence of synchronization pulses is broadcast to all stations. As with pure ALOHA, packet lengths are constant. Messages are required to be sent in the slot time between synchronization pulses, and can be started only at the *beginning* of a time slot. This simple change reduces the rate of collisions by half, since only messages transmitted in the same slot can interfere with one another. It can be shown [9, 10] that for S-ALOHA, the reduction in the *collision window* from 2τ to τ results in the following relationship between normalized throughput ρ and normalized total traffic G:

$$\rho = Ge^{-G} \tag{11.29}$$

The plot of Equation (11.29) is shown in Figure 11.20 labeled "slotted ALOHA." Here the maximum value of ρ is $1/e = 0.37$, or an improvement of two times the pure ALOHA protocol.

The retransmission mode described for the pure ALOHA system was modified for S-ALOHA so that if a negative acknowledgment (NAK) occurs, the user retransmits after a *random* delay of an integer number of slot times. Figure 11.21 illustrates the S-ALOHA operation. A packet of data bits is shown transmitted by user k followed by the satellite acknowledgment (ACK). Also shown are users m and n simultaneously transmitting packets, which results in a collision; a NAK is returned. Each using station employs a random-number generator to select its retransmission time. The figure illustrates an example of the m and n retransmission at their respective randomly selected times. Of course, there is some probability that users m and n will recollide. However, in that case, they simply repeat the retransmission, using another random delay.

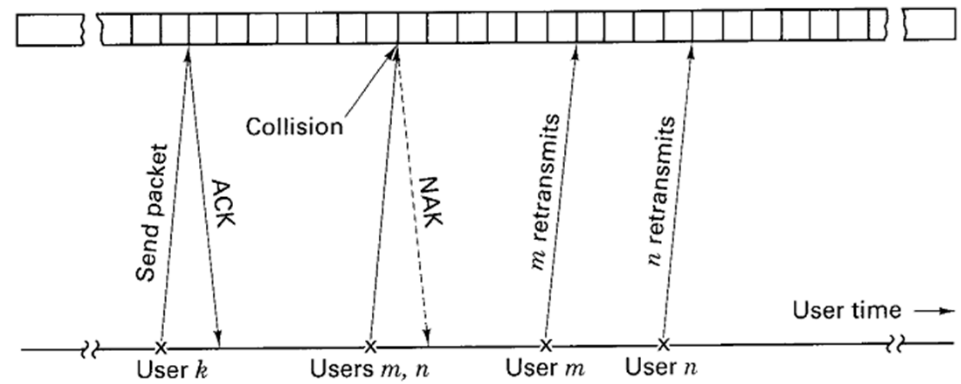


Figure 11.21 Random access scheme: slotted ALOHA operation.

Example 11.1 Poisson Process

Assuming that packet transmissions and retransmission can both be described as a Poisson process, calculate the *probability* that a data packet transmission in an S-ALOHA system will experience a collision with *one other user*. Assume that the total traffic rate $\lambda_t = 10$ packets/s and the packet duration $\tau = 10$ ms.

Solution

$$\begin{aligned} P(K=1) &= \left.\frac{(\tau\lambda_t)^K e^{-\tau\lambda_t}}{K!}\right|_{K=1} \\ &= (10 \times 0.01)^1 e^{-0.1} = 0.1e^{-0.1} \\ &= 0.09 \end{aligned}$$

11.3.3 Reservation-ALOHA

A significant improvement was made to the ALOHA system with the introduction of the reservation-ALOHA (R-ALOHA) [11] scheme. The R-ALOHA system has two basic modes: an unreserved mode and a reserved mode; each is described as follows:

Unreserved Mode (Quiescent State)

1. A time frame is established and divided into a number of small reservation subslots.
2. Users use these small subslots to reserve message slots.
3. After requesting a reservation, the user listens for an acknowledgment and a slot assignment.

Reserved Mode

1. The time frame is divided into $M + 1$ slots whenever a reservation is made.
2. The first M slots are used for message transmissions.
3. The last slot is subdivided into subslots to be used for reservation/requests.
4. Users send message packets only in their assigned portions of the M slots.

Consider the R-ALOHA example shown in Figure 11.22. In the quiescent state, with no reservations, time is partitioned into short subslots for making reservations. Once a reservation is made, the system is configured so that $M = 5$ message slots followed by $V = 6$ reservations subslots becomes the timing format. The figure illustrates a request and an acknowledgment in progress. In this example the station seeks to reserve three message slots. The reservation acknowledgment advises the using station where to locate its first data packet. Since the control is distributed so that all participants receive the downlink transmissions and are thus aware of the reservations and time format, the acknowledgment need not disclose any more than the location of the first slot. As shown in Figure 11.22, the station sends its second packet in the slot following the first packet. The user further knows that the next slot is comprised of six subslots for reservations, so *no* packets are transmitted during this time. The third and final packet is sent in the following slot. When there are no reservations taking place, the system reverts back to its quiescent format of subslots only. Since the control is distributed, all the participants are made aware of the quiescent format by receiving appropriate synchronizing pulses on the downlink. Other interesting reservation schemes are discussed in Reference [12, 13].

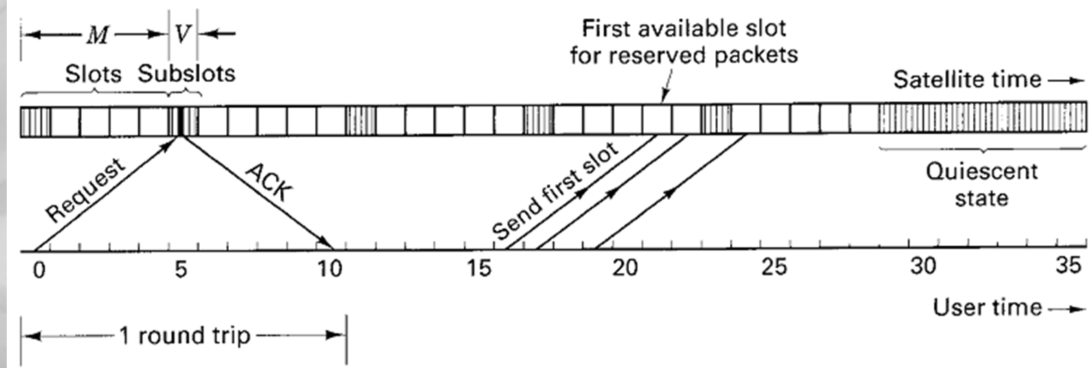


Figure 11.22 Example of reservation ALOHA. Station seeks to reserve three slots ($M = 5$ slots, $V = 6$ subslots).

11.3.4 Performance Comparison of S-ALOHA and R-ALOHA

From Chapters 3 and 4 the basic quality measure of a digital modulation scheme is its P_B versus E_b/N_0 curve. This measure is particularly useful because E_b/N_0 is a *normalized signal-to-noise ratio*; being normalized, the curves allow us to compare the performance of various modulation schemes. There is a similar performance measure for multiple access schemes. Here we are interested in the average delay versus normalized throughput. What would an *ideal delay–throughput curve* look like? Figure 11.23 illustrates such a curve. For normalized throughput values of, $0 \le \rho < 1$, the delay equals zero until $\rho = 1$; then the delay increases without bound. Figure 11.23 also shows a *typical* delay–throughput curve and the direction in which the curve will move as delay performance improves.

Figure 11.24 compares the delay–throughput performance of S-ALOHA with that of R-ALOHA (formatted with two message slots and six reservation subslots). Knowing the location of the *ideal* curve it is easy to compare the delay performance of these two systems. For a throughput of less than approximately 0.20, the S-ALOHA manifests less average delay than does R-ALOHA. But for values of ρ between 0.20 and 0.67, it is apparent that R-ALOHA is superior, since the average delay is less. Why does the S-ALOHA perform better at low traffic intensity? The S-ALOHA algorithm does not require the overhead of the reservation subslots as does R-ALOHA. Therefore, at low values of ρ, R-ALOHA pays the price of greater delay due to the greater overhead. For $\rho > 0.2$, the collisions and retransmissions inherent in the S-ALOHA system cause it to incur greater delay (unbounded at $\rho = 0.37$), more quickly than the R-ALOHA system. At higher throughput ($0.2 < \rho < 0.67$), the overhead structure of R-ALOHA ensures that its delay degradation grows in a more orderly manner than S-ALOHA. For R-ALOHA, an unbounded delay is not reached until $\rho = 0.67$.

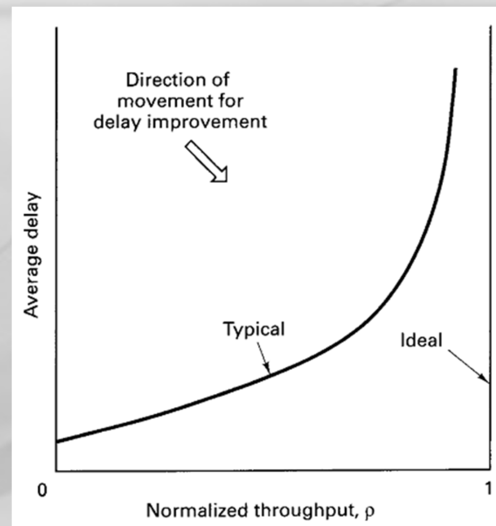


Figure 11.23 Delay–throughput characteristic.

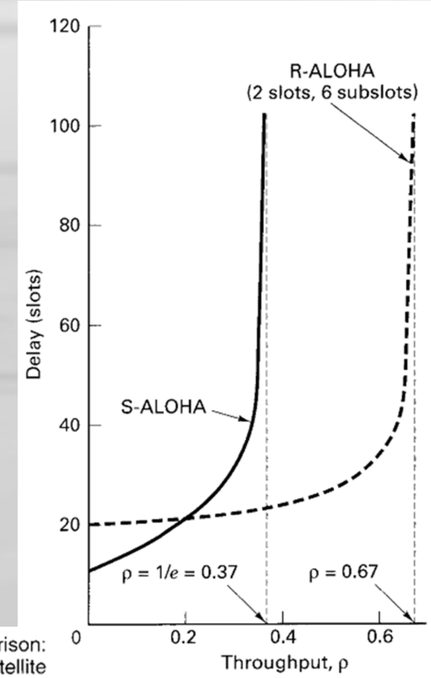


Figure 11.24 Delay–throughput comparison: S-ALOHA versus R-ALOHA on a satellite channel.

Example 11.2 Channel Utilization

(a) Normalized throughput, ρ, is a measure of channel utilization. It can be found by forming the ratio of the successfully transmitted message traffic, in bits per second to the total message traffic, including rejected messages, in bits per second. Calculate the normalized throughput of a channel that has a maximum data rate $R = 50$ kbits/s and operates with $M = 10$ ground stations, each station transmitting at the average rate of $\lambda = 2$ packets/second. The system format provides for $b = 1350$ bits/packet.

(b) Which of the three ALOHA schemes discussed—pure, slotted, and reservation—could be successfully used with this channel?

Solution

(a) Generalizing Equation (11.19) to allow for traffic from multiple stations, we have

$$\rho = \frac{Mb\lambda}{R}$$

$$= \frac{10(1350)\,(2)}{50{,}000}$$

$$= 0.54$$

(b) Only the R-ALOHA scheme could be used for this system, since with each of the other schemes, 54% of the resource cannot be utilized.

11.3.5 Polling Techniques

One way to impose order on a system with multiple users having random access requirements is to institute a controller that periodically polls the user population to determine their service requests. If the user population is large (e.g., thousands of terminals) and the traffic is bursty, the time required to poll the population can be an excessive overhead burden. One technique for rapidly polling a user population [4, 14] is called a *binary tree search*. Figure 11.25 illustrates a satellite example of such a tree search to resolve contention among users. In this example, assume that the total user population is eight terminals; let them be identified by the binary numbers 000 to 111 as shown in Figure 11.25. Assume that terminals 001, 100, and 110 are contending for the service of a single channel. The tree search operates by continually partitioning the population until there is just a single branch remaining. The terminal corresponding to that branch is the "winner" and hence the first terminal to access the channel. The operation is repeated and again yields a single terminal that may next use the channel. The algorithm proceeds according to the following steps (see Figure 11.25):

1. The satellite requests the transmission of the contending terminals' first (left-most) bit of their identification (ID) numbers.
2. Terminal 001 transmits a zero, and terminals 100 and 110 each transmit a one. The satellite, on the basis of received signal strength, selects one or zero as the bit it "heard." In this example the satellite chooses binary one and informs the users accordingly. Half the user population now knows that it has not been selected. The terminals in the "losing" half "bow out" of contention during this pass through the tree. In this example terminal 001 bows out.
3. The satellite requests the transmission of the second identifying bit from the remaining contending terminals.
4. Terminal 100 transmits a zero, and terminal 110 transmits a one.
5. Assume that the satellite selects the zero and notifies the contenders accordingly. Terminal 110 bows out. The process continues until it is clear that terminal 100 is free to access the satellite.
6. When the channel becomes available, steps 1–5 are repeated.

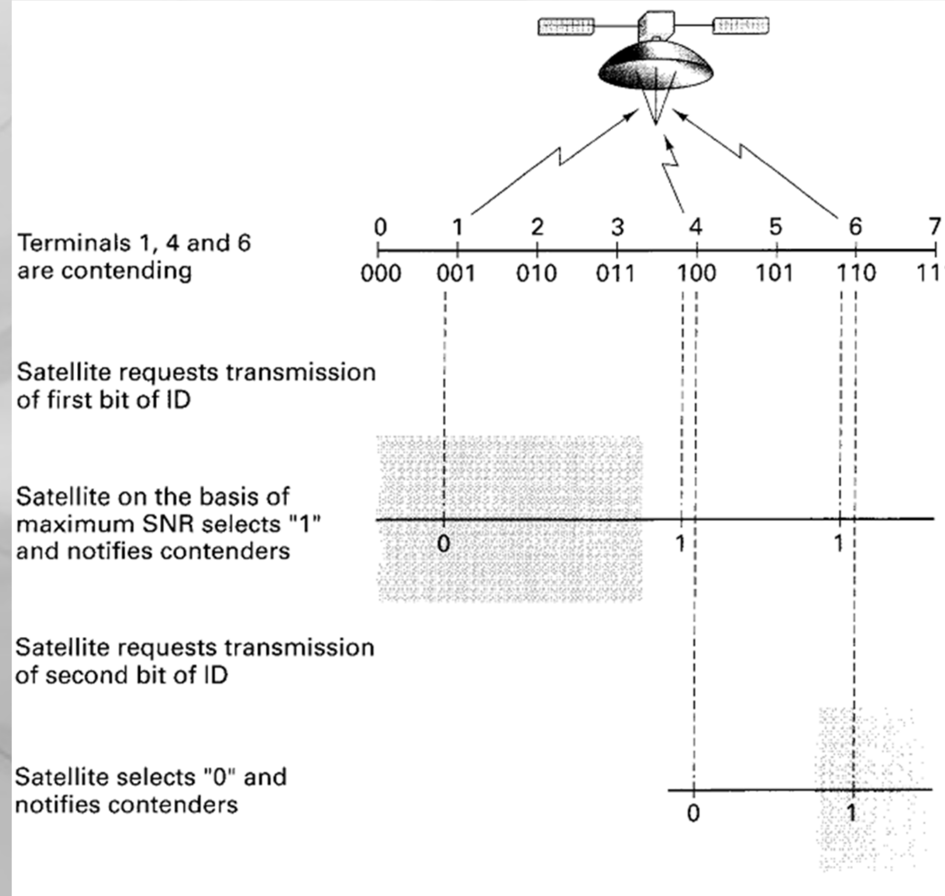


Figure 11.25 Tree search to resolve contention (eight terminal example).

Multiplexing and Multiple Access

Example 11.3 Comparison between Binary Tree Search and Straight Polling

(a) A binary tree search requires $n = \log_2 Q$ decisions for each pass through a population of Q terminals. A savings in time is possible with a tree search if the population is large and the average demand for service is small. Calculate the time needed for the straight polling of a population of 4096 terminals, to provide channel availability to 100 terminals requesting service. Compare the result with the time needed to perform a binary tree search 100 times, over the same population. Assume that the time required to poll one terminal and the time required for one decision of a binary tree search are each equal to 1 s.

(b) Develop an expression for Q', the largest number of terminals that results in the same (or less) time expended for binary tree searching as compared to straight polling.

(c) Compute Q' for part (a).

Solution

(a) Straight polling of 4096 terminals:

$$T = 4096 \times 1 \text{ s} = 4096 \text{ s}$$

Binary tree search for 100 terminals requires 100 passes through the binary tree:

$$T' = (100 \times \log_2 4096) \times 1 \text{ s} = 1200 \text{ s}$$

(b) Q' is the maximum number of terminals that will result in $T' \leq T$ in part (a). This will occur when

$$Q'' \log_2 Q \times 1 \text{ s/decision} = Q \times 1 \text{ s/poll}$$

$$Q' = \lfloor Q'' \rfloor = \left\lfloor \frac{Q}{\log_2 Q} \right\rfloor \tag{11.30}$$

where $\lfloor x \rfloor$ is the largest integer no greater than x.

(c) Q' for part (a)

$$Q' = \left\lfloor \frac{4096}{\log_2 4096} \right\rfloor = 341 \text{ terminals}$$

A binary tree search for 341 terminals entails a search time of 4092 s.

Multiplexing and Multiple Access

11.5 MULTIPLE ACCESS TECHNIQUES FOR LOCAL AREA NETWORKS

A local area network (LAN) can be used to interconnect computers, terminals, printers, and so on, located within a building or a small set of buildings. While long-haul networks use the public telephone network for economic reasons, LAN designers usually lay their own high-bandwidth cables. Bandwidth is not as scarce as it is in the long-haul cases. Not being forced to optimize bandwidth, a LAN can use simple access algorithms [6, 25–27].

11.5.1 Carrier-Sense Multiple Access Networks

Ethernet is a LAN access scheme developed by the Xerox Corporation. The Ethernet scheme is based on the assumption that each local machine can sense the state of a common broadcast channel before attempting to use it. The technique is known as *carrier-sense multiple access with collision detection* (CSMA/CD). The word "carrier," here, means *any* electrical activity on the cable. Figure 11.41a illustrates the bit field format for the Ethernet specification; the details are listed as follows:

1. The maximum packet size is 1526 bytes, where a byte is 8 bits. The packet breakdown is 8-byte preamble + 14-byte header + 1500-byte data + 4-byte parity.
2. The minimum packet size is 72 bytes, consisting of an 8-byte preamble + 14-byte header + 46-byte data + 4-byte parity.
3. The minimum spacing between packets is 9.6 μs.
4. The preamble contains a 64-bit synchronization pattern of alternating ones and zeros, ending with two consecutive ones: (1 0 1 0 1 0 . . . 1 0 1 0 1 1).
5. The receiving station examines a destination address field in the header to see if it should accept a particular packet. The first bit indicates the type of address (0 = single address, 1 = group address); an entire field of ones means an all-station broadcast.
6. The source address is the unique address of the transmitting machine.
7. The type field determines how the data field is to be interpreted. For example, bits in the type field can be used to describe such things as data encoding, encryption, message priority, and so on.
8. The data field is an integer number of bytes from a minimum of 46 to a maximum of 1500.
9. The parity check field houses the parity bits which are generated by the following generating polynomial (see Section 6.7):

$$X^{32} + X^{26} + X^{23} + X^{22} + X^{16} + X^{12} + X^{11} + X^{10} + X^{8} + X^{7} + X^{5} + X^{4} + X^{2} + X + 1$$

Multiplexing and Multiple Access

The Ethernet multiple access algorithm defines the following user action or response:

1. *Defer.* The user must not transmit when the carrier is present or within the minimum packet spacing time.
2. *Transmit.* The user may transmit if not deferring until the end of the packet or until a collision is detected.
3. *Abort.* If a collision is detected, the user must terminate packet transmission and transmit a short jamming signal to ensure that all collision participants are aware of the collision.
4. *Retransmit.* The user must wait a random delay time (similar to the ALOHA system) and then attempt retransmission.
5. *Backoff.* The delay before the nth attempt is a uniformly distributed random number from 0 to $2^n - 1$, for $(0 < n \leq 10)$. For $n > 10$, the interval remains 0 to 1023. The unit of time for the retransmission delay is 512 bits (51.2 μs).

Figure 11.41b illustrates a 10-Mbits/s data stream with Manchester PCM formatting from the Ethernet specification. Notice that with such formatting, each bit cell or bit position contains a transition. A binary one is characterized by transitioning from a low level to a high level, while a binary zero has the opposite transition. Therefore, the presence of data transitions denotes to all "listeners" that the carrier is present. If a transition is not seen between 0.75 and 1.25 bit times since the last transition, the carrier has been lost, indicating the end of a packet.

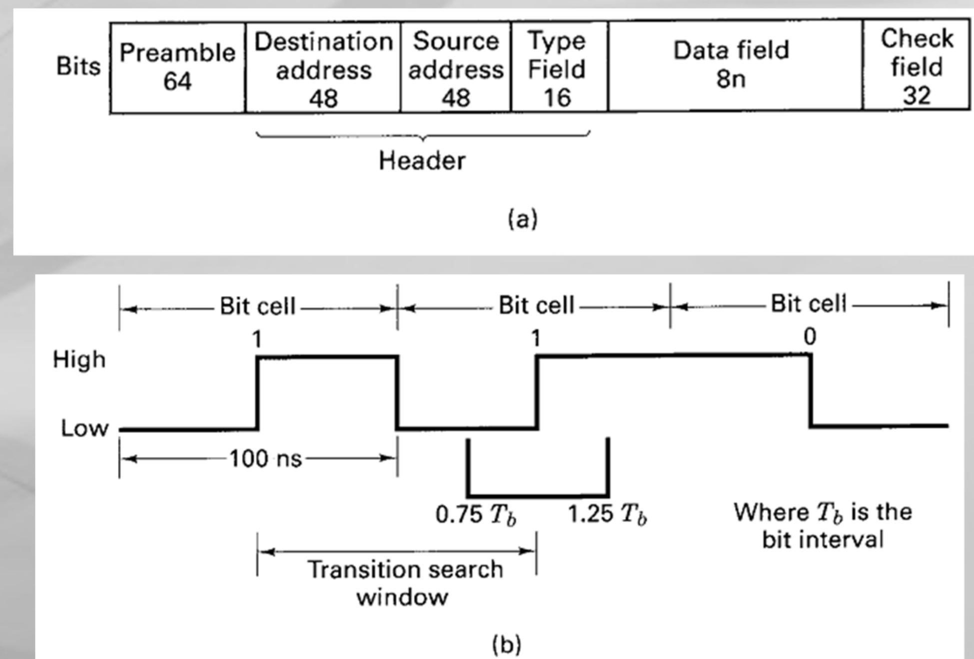


Figure 11.41 Ethernet bit field and PCM format. (a) Ethernet specification. (b) Manchester PCM format.

11.5.2 Token-Ring Networks

A carrier-sense network consists of a cable onto which all stations are passively connected. A *ring network*, by comparison, consists of a series of point-to-point cables between consecutive stations. The interfaces between the ring and the stations are active rather than passive. Figure 11.42a illustrates a typical unidirectional ring with interface connections to several stations. Figure 11.42b illustrates the state of the interface for the listen mode and the transmit mode. In the *listen mode* the input bits are copied to the output with a delay of one bit time. In the *transmit mode*, the connection is broken so that the station can enter its own data onto the ring. The token is defined as a special bit pattern (e.g., 1 1 1 1 1 1 1 1) which circulates on the ring whenever all stations are idle. How does the system ensure that message data do not contain a tokenlike sequence? *Bit stuffing* is used to prevent this pattern from occurring in the data. For the 8-bit token example shown, a bit-stuffing algorithm would insert a zero into the data stream after each sequence of seven consecutive ones. The data receiver would use a similar algorithm to dispose of the inserted bit following any sequence of seven consecutive ones. The token-ring access scheme works as follows:

1. A station wanting to send a message monitors the token appearing at the interface. When the last bit of the token appears, the station inverts it (e.g., 1 1 1 1 1 1 1 0). The station then breaks the interface connection and enters its own data onto the ring.
2. As bits come back around the ring, they are removed by the sender. There is no limit on the size of the packets, because the entire packet never appears on the ring at one instant.
3. After transmitting the last bit of its message, the station must regenerate the token. After the last data bit has circled the ring and has been removed, the interface is switched back to the listen mode.
4. Contention is not possible with a token-ring system. During heavy traffic, as soon as a token is regenerated, the next downstream station requiring service will see and remove the token. Thereby, permission to transmit rotates smoothly around the ring. Since there is only one token, there is no contention.

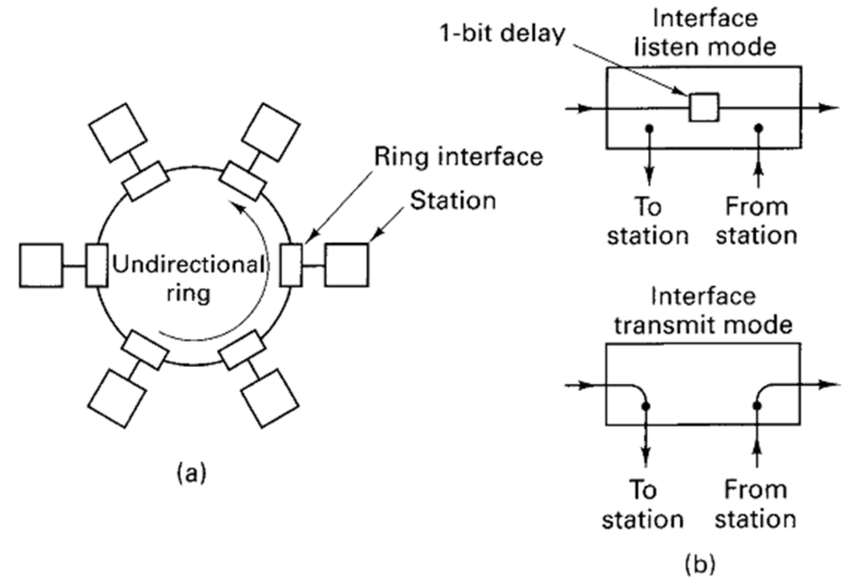


Figure 11.42 Token-ring network. (a) Network. (b) Listen and transmit modes.

The ring itself must have sufficient delay to enable a complete token to circulate when all stations are idle. A major issue in ring network design is the propagation distance or "length" of a bit. If the data rate is R Mbits/s, a bit is emitted every $(1/R)$ microseconds. Since the propagation rate along a typical coaxial cable is 200 m/μs, each bit occupies $200/R$ meters on the ring.

Multiplexing and Multiple Access

Example 11.4 Minimum Ring Size

If an 8-bit token is to be used on a 5-Mbits/s token-ring network, calculate the minimum *propagation distance,* d_p, needed for the ring circumference. Assume that the propagation velocity v_p is 200 m/µs.

Solution

$$R = 5 \text{ Mbits/s}$$

Time to emit one bit, t_b:

$$t_b = \frac{1}{5 \times 10^6} \text{ s}$$

Time to emit the 8-bit token, t_t:

$$t_t = \frac{8}{5 \times 10^6} \text{ s}$$

Propagation distance for the 8-bit token:

$$\begin{aligned} d_p &= t_t \times v_p \\ &= \tfrac{8}{5}\ \mu\text{s} \times 200\ \text{m}/\mu\text{s} \\ &= 320\ \text{m} \end{aligned}$$

Multiplexing and Multiple Access

11.5.3 Performance Comparison of CSMA/CD and Token-Ring Networks

Figure 11.43 compares the delay-throughput characteristics of a CSMA/CD network with a token-ring network. In each case, the cable length is 2 km, there are 50 stations on the network, the average packet length is 1000 bits, and the header size is 24 bits. Figure 11.43a, the case where the transmission rate is 1 Mbits/s, illustrates that under these assumptions, CSMA/CD and token ring perform almost equally well. In Figure 11.43b, only one parameter has been changed as compared to Figure 11.43a; the transmission rate was increased to 10 Mbits/s. The difference for CSMA/CD is considerable; for normalized throughput, $\rho < 0.22$, CSMA/CD performs better than token ring. However, for $\rho > 0.22$, token ring clearly manifests better delay-throughput characteristics. To understand the reason for the poor CSMA/CD performance in Figure 11.43b, let us review the definition of ρ, described in Equations (11.17) and (11.19) and shown as

$$\rho = \frac{b\lambda}{R} = \frac{\rho'}{R}$$

where $\rho' = b\lambda$ is the channel throughput in bits per second and R is the channel capacity (maximum transmission bit rate). As R increases, channel throughput must increase accordingly for a given value of ρ. At higher channel throughput rates, a significant portion of the CSMA/CD transmission attempts ends in collision [26].

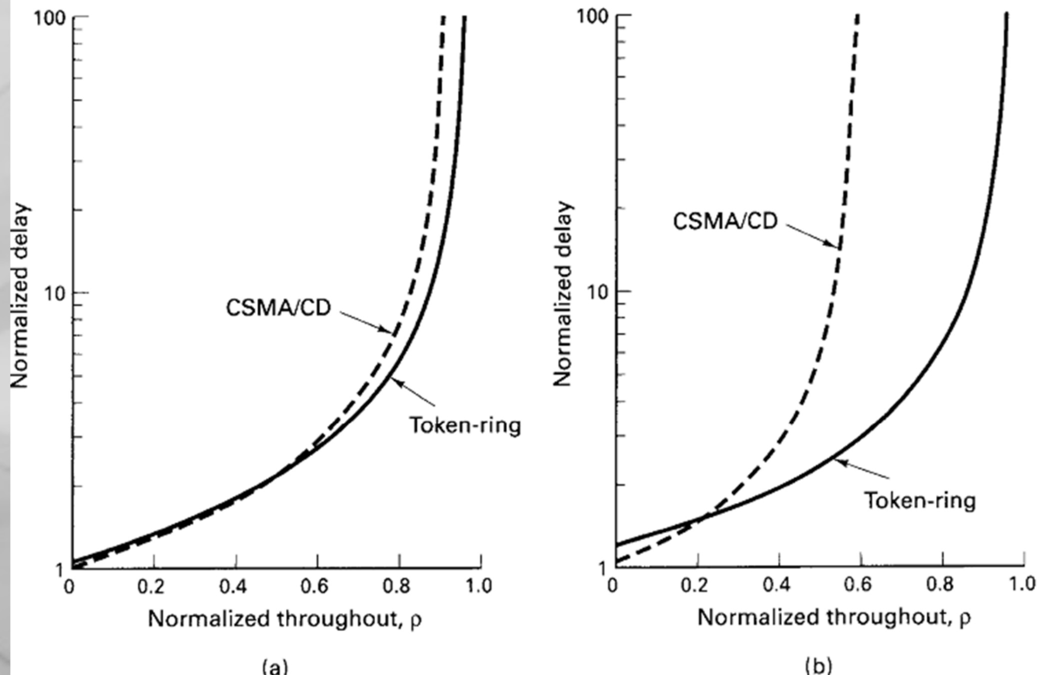


Figure 11.43 Delay versus throughput performance for CSMA/CD and token-ring networks. (a) Transmission rate = 1 Mbits/s. (b) Transmission rate = 10 Mbits/s. (Reprinted with permission from W. Bux, "Local-Area Subnetworks: A Performance Comparison," *IEEE Trans. Commun.*, vol. COM29, no. 10, Oct. 1981, pp. 1465–1473. © 1981 IEEE.)

Multiplexing and Multiple Access

CONCLUSION

In this chapter we have outlined the concepts of resource sharing. The classical approaches of FDM/FDMA and TDM/TDMA were discussed in some detail. We also described a hybrid multiple access technique called CDMA, and introduced some of the satellite multiple access techniques that became popular in the 1970s and 1980s, known as multiple-beam frequency reuse and dual-polarization frequency reuse.

We described the demand-assignment (DAMA) techniques in the context of several versions of the ALOHA algorithm, and we considered several of the multiple-access techniques employed with INTELSAT, such as FDM/FM, SPADE, TDMA, and SS/TDMA. Finally, we examined two popular algorithms used for local area networks: carrier-sense multiple access with collision detection (CSMA/CD) and a token-ring network. The goals of the chapter were to introduce an assortment of multiple access techniques rather than attempting a rigorous treatment of any of them.

Problems and Questions

- Problems

11.3. Equations (11.13) to (11.15) demonstrate that the average message delay time for TDMA is less than that for FDMA. Discuss the practical benefits of such reduced delay in TDMA, as a function of frame time, for a satellite link with a one-way range of 36,000 km. For what values of frame time can there be a significant advantage of TDMA over FDMA?

11.6. Verify that for a pure ALOHA access scheme, the normalized throughput is bounded by $1/2e$ and that this maximum occurs when the normalized total traffic is equal to 0.5.

11.12. A group of slotted-ALOHA stations generate a total of 120 requests per second, including both original and retransmissions. Each request is for a 12.5-ms duration slot.

(a) What is the normalized total traffic on the channel?

(b) What is the probability of a successful transmission on the first attempt?

(c) What is the probability of exactly two collisions before a successful transmission?

11.16. A TDMA system operates at 100 Mbits/s with a 2-ms frame time. Assume that all slots are of equal length and that a guard time of 1 μs is required between slots.

(a) Compute the efficiency of the communications resource (CR) for the case of 1, 2, 5, 10, 20, 50, and 100 slots per frame.

(b) Repeat part (a) assuming that a 100-bit preamble is required at the start of each slot. Compute the efficiency of the CR in terms of the desired information transmission.

(c) Graph the results of parts (a) and (b).

11.18. (a) Consider a token-ring network operating at a transmission rate of 10 Mbits/s over a cable having a propagation velocity of 200 m/μs. How many meters of cable is equal to a delay of 1 bit at each ring interface?

(b) If the token is 10 bits long, and all but three stations are switched off during evening hours, what is the minimum cable length needed for the ring?

Problems and Questions

- Questions

11.1. What is typically meant by the terms *communications resource?* (See Chapter 11, introduction.)

11.2. What are the similarities and differences between the terms *multiplexing* and *multiple access?* (See Chapter 11, introduction.)

11.3. Why is it not possible to use a *linear device* as a mixer (See Section 11.1.1.1 and Appendix A.)

11.4. Is there any theoretical capacity advantage in providing users with FDMA service versus TDMA service? (See Section 11.1.4.1.)

11.5. What benefits might there be in using CDMA versus either FDMA or TDMA access schemes? (See Section 11.1.5.)