Packet Switching in Radio Channels: Part III—Polling and (Dynamic) Split-Channel Reservation Multiple Access

FOUAD A. TOBAGI AND LEONARD KLEINROCK, FELLOW, IEEE

Abstract—**Here we continue the analytic study of packet switching in radio channels which we reported upon in our two previous papers [1], [2]. Again we consider a population of terminals communicating with a central station over a packet-switched radio channel. The allocation of bandwidth among the contending terminals can be *fixed* [e.g., time-division multiple access (TDMA) or frequency-division multiple access (FDMA)], *random* [e.g., ALOHA or carrier sense multiple access (CSMA)] or *centrally controlled* (e.g., polling or reservation). In this paper we show that with a large population of bursty users, (as expected) random access is superior to both fixed assignment and polling. We also introduce and analyze a dynamic reservation technique which we call split-channel reservation multiple access (SRMA) which is interesting in that it is both simple and efficient over a large range of system parameters.**

I. INTRODUCTION

THE primary goal of a computer network is to permit users and processes at one research center to interactively access and use data, programs, and computing resources that exist at other research centers. An excellent example is the ARPANET, which currently consists of more than 60 nodes. However, it may be observed that the full potential of such resource-sharing networks will not be realized until new techniques are developed that provide a high-quality, flexible, and responsive *terminal access* capability.

Numerous papers have already appeared in the literature which discuss the advantages of using radio as an alternative to wire communication for terminal-to-computer communication [3]-[5]. The key property is that radio is a *multiaccess broadcast* medium; that is, any number of users may access the channel at the same time (however, signals on the same carrier frequency which overlap in time may result in mutual destruction), and the transmission of a signal by a user may be received over a wide area by any number of receivers operating at the same frequency. Broadcast radio communications provide us with easy access to central computer installations and computer networks, and the collection and dissemination of data over large distributed geographical areas is independent of the availability of preexisting (telephone) wire networks. Moreover, wireless connections are particularly suitable for communication with and among mobile users, a constantly growing area of interest and application. Accordingly, the Advanced Research Projects Agency of the Department of Defense undertook an effort whose goal is to develop new techniques for radio communication among geographically distributed, fixed or mobile, user terminals and to provide improved frequency management strategies to meet the critical shortage of the RF spectrum [5].

In the computer-to-computer data transmission case (e.g., file transmissions), one often sees a higher utilization of the communication channels than is the case with terminal traffic. The reason is simply that terminals, large in number and often geographically distributed, are basically *bursty* sources of data, i.e., they tend to generate demands at a very low duty cycle.

Let us consider an environment consisting of a population of M identical user terminals wishing to communicate with a central station over a radio channel of limited bandwidth, say W Hz.[1] The basic question here is how to allocate this bandwidth among the contending terminals such that the limiting communications resource is efficiently utilized and such that the terminals' delays are within an acceptable range. The various known alternatives fall into the three following categories.

A. Fixed Assignment

This technique consists of allocating the channel to users independent of their activity, by partitioning the time-bandwidth space into slots which are assigned to the user population in a static predetermined fashion. It takes two common forms: orthogonal, such as frequency-division multiple access (FDMA) and synchronous time-division multiple access (TDMA—commonly known as time-division multiplexing), and "quasi-orthogonal," such as code-division multiple access (CDMA). Assuming all users to be identical, FDMA consists of assigning to each user a fraction W/M of the bandwidth, along with buffering capabilities required to handle the statistical fluctuations due to the random message arrivals. TDMA consists of assigning fixed predetermined channel time slots to each user; it also results in assigning a fraction $1/M$ of the total channel capacity and also requires buffering capabilities. A number of disadvantages of FDMA exist when compared with TDMA: wasted bandwidth for adequate frequency separation, lack of flexibility in achieving dynamic allocation of bandwidth, lack of broadcast operation. The only major disadvantage in TDMA is the need to provide rapid burst synchronization and sufficient burst separation to avoid time overlap. However, in a satellite communication environment, INTELSAT's MAT-1 experimental TDMA system has shown that guard bands of less than 200 ns are achievable and new operational systems are moving towards the use of TDMA.

CDMA allows more than one user to share a common band in a nondestructive fashion.

[1] The bandwidth is assumed to be modulated at 1 bit/Hz·s.

Paper approved by the Editor for Computer Communication of the IEEE Communications Society for publication after presentation at the National Computer Conference, Anaheim, CA, 1975. Manuscript received September 12, 1975; revised February 17, 1976. This work was supported by the Advanced Research Projects Agency of the Department of Defense under Contract DAHC15-73-C0368.

The authors are with the Computer Science Department, University of California, Los Angeles, CA 90024.

B. Random Access (No Assignment)

In this category, the entire bandwidth is provided as a single high-speed channel to be shared dynamically by the users in some fashion. This resource-sharing is accomplished through *packet-switching,* a packet being merely a package of data prepared by one user for transmission to some other user in the system. The difficulty in controlling a channel which must carry its own control information gives rise to the so-called random-access techniques. The random-access techniques studied so far are ALOHA [3], [6], [7] and carrier sense multiple access (CSMA) previously introduced and analyzed by the authors in [1], [2], and [8]. Since signals for the single carrier frequency which overlap in time result in information destruction (unless a spread spectrum method such as CDMA is used), packet collisions are inherent to these random-access techniques.

C. Centrally Controlled Assignment

Here there are two methods (in common usage for wire networks): contention and polling [9]. They both require the presence of a central station performing the control. In a *contention network,* the terminal makes a request to transmit: if the channel is free, transmission goes ahead; if it is not free, the terminal must wait. The station schedules the transmissions either in a prearranged sequence (according to some scheduling scheme) or in the sequence in which the requests were made. In the *polling* technique, the station asks (polls) the terminals one by one as to whether they have anything to transmit. For this, the station may have a polling list giving the order in which terminals are polled. When a polling message is sent to the next terminal in sequence, and if the terminal has some data to transmit, it goes ahead; if not, a negative reply (or absence of reply) is received by the station, and the next terminal is polled. These controlled techniques are readily applicable to radio channels as well; in this case, they require that only the central station be within range and in line of sight of all terminals.

The emphasis in this paper is to consider controlled techniques for packet radio channels and the comparison of their performance with that of the known fixed and random assignment techniques. For this, we first give, in Section II, a simple comparative study between FDMA and slotted ALOHA showing quantitatively the superiority of each over different parameter ranges. In Section III, we review the performance of a simple polling scheme known as roll-call polling [9]. In Section IV, we introduce and analyze a new efficient contention technique which we refer to as split-channel reservation multiple access (SRMA). Our goal is to compare these various schemes on an analytic basis, and this we do throughout the paper.

II. RANDOM ACCESS VERSUS FIXED ASSIGNMENT

It has long been recognized that fixed allocation of a scarce communication resource is extremely wasteful when M is large and the terminals are bursty. On the other hand, providing the bandwidth as a single high-speed channel to a large number of users allows us to take advantage of the powerful "large number laws" which state that with very high probability, the demand at any instant will be approximately equal to the sum of the average demands of that population. To illustrate this quantitatively, we wish to compare FDMA[2] with the simple random-access scheme known as slotted ALOHA [6], [7], [10].

The performance measures considered in this comparison and throughout the paper are channel throughput (or bandwidth utilization) and average packet delay. The average *packet delay* $\mathcal{D}$ is defined as the average time from when a packet is generated until it is successfully received at the station.

To analyze FDMA, we adopt the following assumptions: a) an assumed finite (but large) population of M users; b) each user generates a new fixed-length packet (of b_m bits) according to a Poisson process at a rate Λ packets/s; c) the total channel has a bandwidth of W Hz modulated at 1 bit/Hz·s (giving a channel capacity of W bit/s). Thus, with M users in this FDMA mode, each is assigned a channel of W/M bit/s [see Fig. 1(a)]; we neglect any loss due to guard bands, etc. Each such channel behaves as an $M/D/1$ queueing system giving an average time in system $\mathcal{D}$ (waiting plus transmission) as follows [11]:

$$\mathcal{D} = \frac{\frac{\rho}{\Lambda}\left(1 - \frac{\rho}{2}\right)}{1 - \rho} \tag{1}$$

where $\rho = M\Lambda b_m/W$.

We are assuming that queueing is permitted at each of the M terminals. We note that a finite population model with M users, each at rate Λ and with queueing permitted, produces fewer collisions in random access than does the infinite population since each terminal avoids conflicts among its own packets. However, the analysis for slotted ALOHA assumes an infinite population of users with an aggregate input rate of $M\Lambda$ packet/s and this, therefore, produces an upper bound on delay. (See Fig. 1(b) and (c).)

Slotted ALOHA with an infinite population has been thoroughly analyzed by Kleinrock and Lam. Neglecting the propagation delay, and letting the maximum retransmission delay be an integer number K of packet slots (the retransmission delay being uniformly distributed over the K slots), the delay $\mathcal{D}$ is then given by [7], [12]

$$\mathcal{D} = \left[1 + \frac{E(K+1)}{2}\right]\frac{b_m}{W} \text{ (in seconds)} \tag{2}$$

[2] Although the delays in both TDMA and FDMA are of the same order of magnitude, they do differ. The delay for FDMA is given by [see (1)]: $\mathcal{D} = [(\rho^2/\Lambda)/2(1-\rho)] + M(b_m/W)$, the first term accounting for the queueing delay and the second, for the service time (the transmission time of the packet on the user-assigned channel). The delay for TDMA can be shown to be $\mathcal{D} = (M/2)(b_m/W) + [(\rho^2/\Lambda)/2(1-\rho)] + (b_m/W)$, where the first term accounts for user slot synchronization (under a Poisson arrival process assumption), the second accounts for the queueing delay, and the third accounts for the transmission of the packet over the channel, that is, the user's slot size. Thus TDMA provides delays smaller than FDMA by $(M/2 - 1)(b_m/W)$. We consider only FDMA in the comparisons of this section.

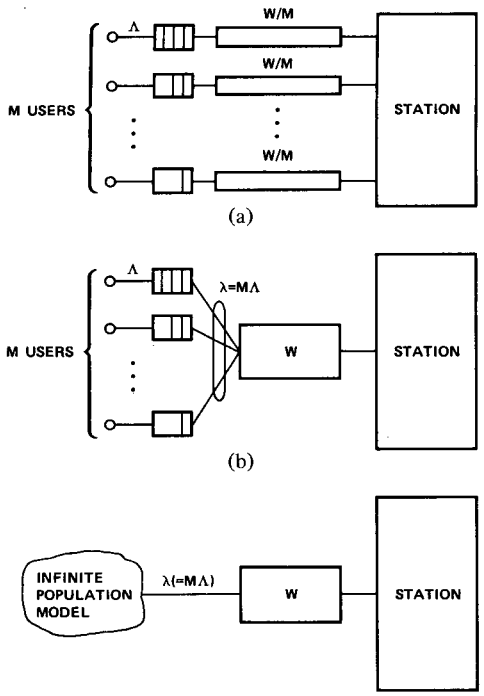


Fig. 1. Channel configurations: (a) FDMA, (b) random access, (c) infinite population model.

where

$$E = \frac{1 - q_n}{q_t}$$

$$q_n = \left[e^{-\frac{G}{K}} + \frac{G}{K} e^{-G} \right]^K e^{-S}$$

$$q_t = \frac{e^{-\frac{G}{K}} - e^{-G}}{1 - e^{-G}} \left[e^{-\frac{G}{K}} + \frac{G}{K} e^{-G} \right]^{K-1} e^{-S}$$

$$S = G \frac{q_t}{q_t + 1 - q_n}.$$

S and Λ are related as follows:

$$S = \frac{M \Lambda b_m}{W}.$$

The normalized packet delay $\mathcal{D}$ (in units of $T = b_m/W$, the packet transmission time) is shown in Fig. 2, versus the normalized input rate S (also referred to as throughput, under steady-state conditions) for various values of K. For each value of S, we note that an optimum value of K can be selected so as to achieve minimum delay. The lower envelope of all delay curves provides the ultimate throughput-delay performance of slotted ALOHA (shown dashed in Fig. 2). It is to be noted that the maximum utilization is limited to $1/e = 0.37$ of the total available bandwidth.

Equation (1) for FDMA is compared with the results for delay in slotted ALOHA with an infinite population as follows. We consider the (M, Λ) plane in Fig. 3, in which we represent constant $\mathcal{D}$ contours. Comparing the delay per-

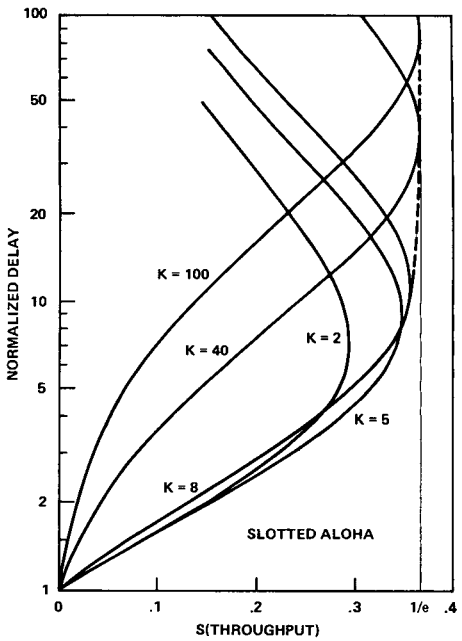


Fig. 2. Delay in slotted ALOHA channels.

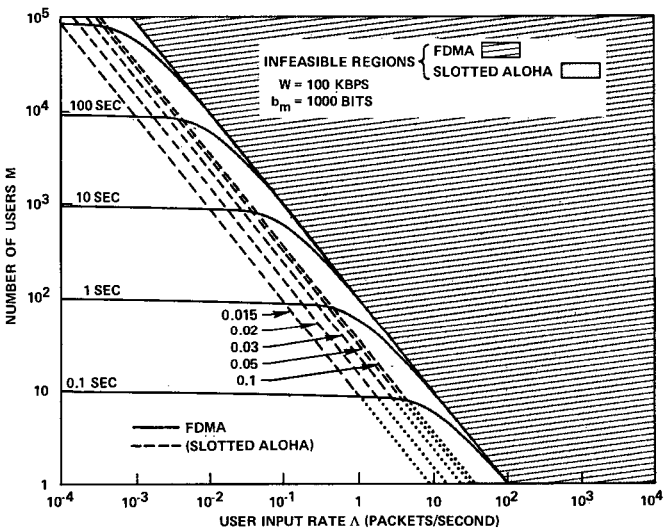


Fig. 3. FDMA and slotted ALOHA random access: performance with 100 kbit/s bandwidth.

formance of the two systems, we note that with bursty users (small Λ), slotted ALOHA can support many more users than FDMA, for the same packet delay. For example, at $\mathcal{D} = 0.1$ s, slotted ALOHA can support a number of users which is over three orders of magnitude greater than the number that FDMA can support when $\Lambda = 10^{-3}$ packet/s; as Λ increases (i.e., as the burstiness decreases), this difference is reduced until at $\Lambda \approx 5$ the two systems can support roughly an equal number of users. Beyond this point, FDMA is superior. This crossover point clearly depends upon the value of $\mathcal{D}$ examined. In fact, slotted ALOHA can support total traffic only in the range $M\Lambda b_m/W < 1/e \cong 0.37$ and beyond that, FDMA will always be superior until it too saturates at $M\Lambda b_m/W = 1$; this tradeoff is clearly evident in Roberts' curves [13].

The above result can be alternatively presented in the following manner. Let M be some large number, say 1000. Fig. 4 shows constant $\mathcal{D}$ contours in the (W, Λ) plane. Again we note

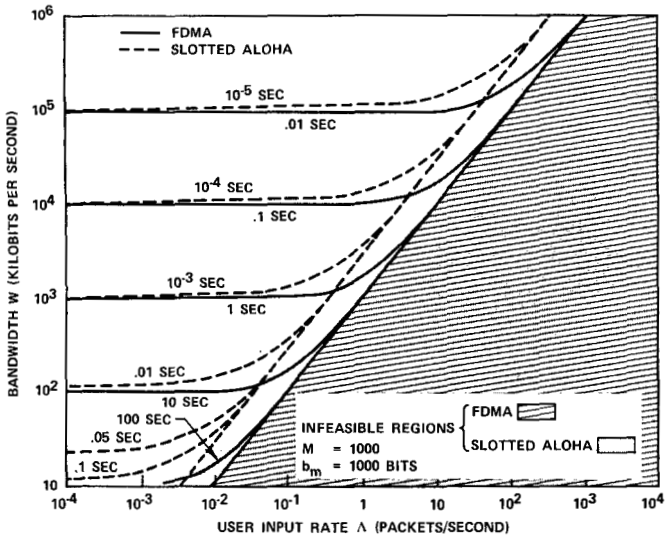


Fig. 4. FDMA and slotted ALOHA random acess: bandwidth requirements for 1000 terminals.

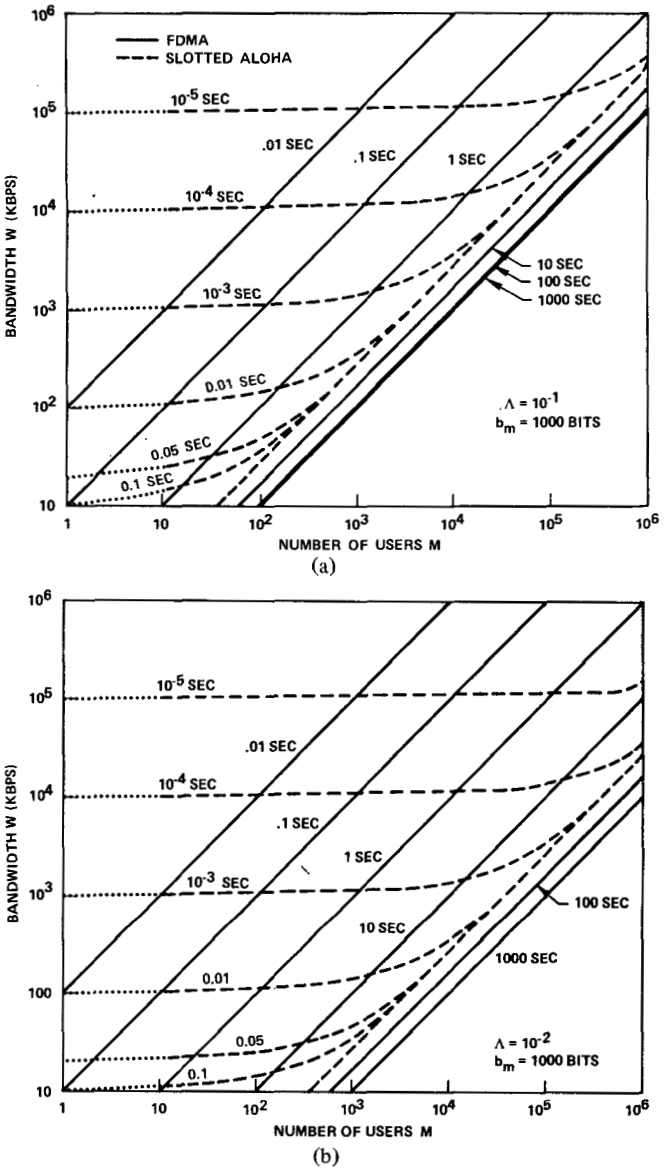


Fig. 5. FDMA and slotted ALOHA random access: performance for fixed Λ. (a) $\Lambda = 10^{-1}$ packets/s; (b) $\Lambda = 10^{-2}$ packets/s.

that with bursty users, in order to achieve the same small delay, FDMA requires a bandwidth far greater than does slotted ALOHA (by as much as three orders of magnitude). This factor is exactly equal to M as $\Lambda \to 0$ since in this region queueing effects are insignificant; in this limit the delay $\mathcal{D}$ is simply the packet transmission time (observe the flatness of the curves in Figs. 3 and 4), which for FDMA is $\mathcal{D} = Mb_m/W$ and which for slotted ALOHA is $\mathcal{D} = b_m/W$. It is also obvious here, for the same total bandwidth W, that FDMA will give M times the delay as compared to slotted ALOHA. This gain diminishes as Λ increases, until finally as $M\Lambda b_m/W \to 1/e$ the situation reverses as mentioned above.

Finally, let us fix Λ and consider the delay contours in the (W, M) plane. Fig. 5 (a) and (b) corresponds to $\Lambda = 10^{-1}$ and $\Lambda = 10^{-2}$ packet/s. Such input rates correspond again to bursty users. We note again that in order to support a large number of users, FDMA requires a larger bandwidth for the same delay performance.

It is all too evident from the above comparison that random access is by far superior to FDMA (or TDMA) when the environment consists of a large population of bursty users. The fixed channel assignment in FDMA is effective in preventing collisions but succeeds in this at the expense of poor utilization of each channel since the smoothing effect of a large population is absent.

However, it is known that slotted ALOHA itself does not use the channel as efficiently as we might hope. This prompted us to inquire as to other, superior, random-access modes. In previous papers [1], [2] we introduced and analyzed the CSMA modes and their extensions, particularly suitable for ground packet radio environments characterized by a propagation delay between source-destination pairs which is very small compared to the packet transmission time. In CSMA, one attempts to avoid collisions by listening to (i.e., sensing) the channel carrier due to another user's transmission. Among the various protocols studied, we consider for the purpose of this study, only the *nonpersistent protocol* because of its simplicity in analysis and implementation, as well as its relatively high efficiency. Briefly, the idea in nonpersistent CSMA is to limit repeated interference among packets by always rescheduling (into the future) a packet which finds the channel busy. Thus a ready terminal senses the channel and operates as follows.

1) If the channel is sensed idle, it transmits the packet.

2) If the channel is sensed busy, then the terminal schedules the retransmission of the packet to some later time, and then repeats the algorithm.

The performance of CSMA was further shown to be highly dependent on the mutual sensing ability of all terminals. The existence of *hidden terminals* (terminals which are out of sight or out of range) significantly degrades the performance of CSMA. To eliminate this problem a natural extension of CSMA, namely, the busy-tone multiple access (BTMA) was considered and was shown to provide an efficient solution. The performance of these systems as determined in the aforementioned references will be considered later in the final comparisons.

Next we consider a polling scheme applied to packet radio. Following that we analyze a dynamic reservation technique and compare it to many of the others.

III. THE ROLL-CALL POLLING TECHNIQUE

Again we consider a population consisting of M identical terminals. We restrict our attention to the flow of data from the terminals to the central station. For the present analysis, we require that each terminal have a buffer of unlimited capacity. Polling messages are sent to each terminal in the population in sequence. A polling message is merely a control packet of smaller size than a message packet which queries the terminal asking if it has any data to transmit; the polling message contains information about the address (identification) of the terminal being polled. Message packets arriving at a terminal are queued in its buffer until the terminal is polled, at which time the buffer is completely emptied. Konheim and Meister [14] analyzed such a system deriving stationary distributions for queue lengths and waiting times. From this reference, we find that the expected value of the stationary queue at a terminal is given by[3]

$$E\{\text{queue length}\} = \frac{1}{2}\frac{\nu}{1-Mm} + \frac{1}{2}\frac{Mrm(1-m)}{1-Mm} \quad \text{(slots of service)} \tag{3}$$

and the stationary expected *queueing* delay that a packet incurs is given by

$$E\{\text{queueing delay}\} = \frac{1}{2}\frac{Mv}{1-Mm} + \frac{1-m}{2} + \frac{1}{2}\frac{Mr(1-m)}{1-Mm} \quad \text{(slots)} \tag{4}$$

where m and ν denote, respectively, the expectation and variance of the number of slots required to service the arrivals occuring at a buffer during a slot, and r denotes the (integer) number of slots required for synchronization (i.e., switching to the next user). We now proceed with the determination of m, ν, and r in the application of this technique to packet radio.

Let the arrival process of packets at a buffer be described by a stationary random process, namely a Poisson process. Let b_m be the number of bits per message packet (considered to be of constant length) and b_p the number of bits per polling packet. Let $L = b_m/b_p$, which we shall assume to be an integer. L is greater than one, typically 10 or 100. Let T_m and T_p be the transmission time of a message packet and a polling packet, respectively; that is, if W is the bandwidth of the channel modulated at 1 bit/Hz·s, then

$$T_m = \frac{b_m}{W} \tag{5}$$

and

$$T_p = \frac{b_p}{W}. \tag{6}$$

Let τ denote the propagation delay between the terminals and the station. We let $a = \tau/T_m$ and $b = T_p/\tau$. In packet radio environments as considered in this paper, the ratio a is small, typically 0.01.[4] Furthermore it is assumed to be identical for all terminals. The analysis requires the distinction between the two cases, $b \geqslant 1$ and $b < 1$.

We first treat the case $b \geqslant 1$ (assuming b to be an integer); here we consider the time axis to be divided into slots of size τ. In this roll-call polling scheme, the channel is assigned to a terminal until its buffer is emptied. The channel is then used for b slots to poll the next terminal in sequence. It takes one slot for the polling packet to reach the terminal and the station has to wait one additional slot (propagation from the terminal to the station) before it can decide whether to allocate the channel to the polled terminal or poll the next terminal in sequence.[5] Therefore, the scheme requires r slots for synchronization purposes and polling-packet transmission, where

$$r = b + 2. \tag{7}$$

For example, when $a = 0.01$ ($T_m = 100\tau$), we have

if $L = 100$ then $b = 1$ and $r = 3$

if $L = 10$ then $b = 10$ and $r = 12$

and when $a = 0.05$ ($T_m = 20\tau$) and $L = 10$, we have $b = 2$ and $r = 4$. Let χ be the random number of packet arrivals at a user's buffer during a slot interval. Let the Poisson arrival process at each terminal have a mean of Λ packets per slot; we have

$$\Pr\{\chi = k\} = \frac{\Lambda^k}{k!}e^{-\Lambda}. \tag{8}$$

Therefore

$$m = \Lambda T_m/\tau = \Lambda/a \tag{9}$$

$$\nu = \Lambda/a^2. \tag{10}$$

With M identical terminals, the system utilization is

[3] The time axis is divided into slots of equal size; for the purpose of our study the slot size will be appropriately chosen as explained later in the text.

[4] Consider, for example, 1000 bit packets transmitted over a channel operating at a speed of 100 kbit/s. The transmission time of a packet is then 10 ms. If the maximum distance between the source and the destination is 10 mi, then the (speed-of-light) packet propagation delay is on the order of 54 μs. Thus the propagation delay is a very small fraction ($a = 0.005$) of the transmission time of a packet. On the contrary, when one considers satellite channels, the propagation delay is a relatively large multiple of the packet transmission time ($a \gg 1$).

[5] The absence of a reply from the terminal means that the terminal has an empty buffer. The station then proceeds by polling the next one in sequence.

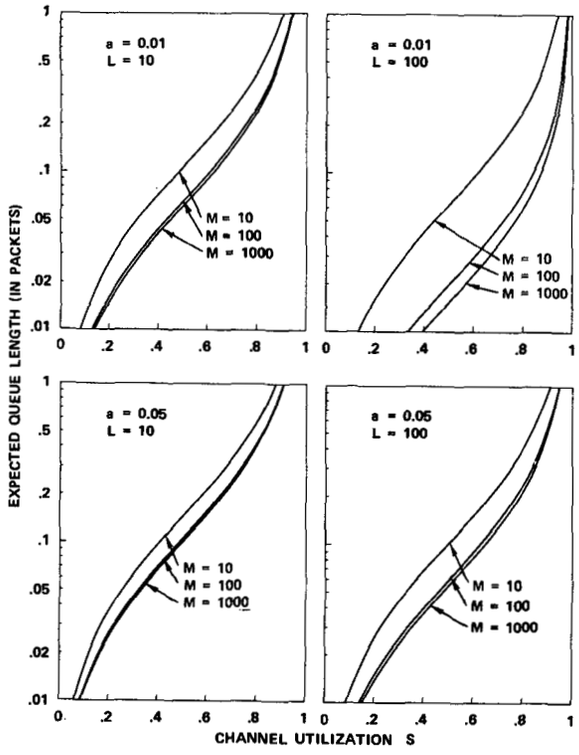


Fig. 6. Roll-call polling: expected queue length.

$$S = Mm = M\Lambda/a = \lambda/a \tag{11}$$

where

$$\lambda = M\Lambda.$$

Substituting r, m, and v in (3) and (4) by the expressions found in (7), (9), and (10) and multiplying both equations by a, we get the expected queue length and the queueing delay in *packets* and *packet transmission times*, respectively.

When $b < 1$, we redefine the slot size and consider the time axis to be divided into slots of size T_p. We assume again for simplicity that τ is an integer multiple of T_p, i.e., $1/b$ is an integer. The number r of slots required for polling and synchronization is given by

$$r = 1 + 2/b. \tag{12}$$

For example, when $a = 0.05$ and $L = 100$, we have $\tau = 5T_p$, i.e., $1/b \approx 5$; in this case $r = 11$. The expectation and variance of the number of slots required to service the arrivals occuring during a slot are now expressed as

$$m = \Lambda L \tag{13}$$

$$v = \Lambda L^2 \tag{14}$$

and therefore $S = Mm = \lambda L$.

Now substituting r, m, v in (3) and (4) by the expressions found in (12), (13), and (14) and multiplying both equations by $1/L$ we get the expected queue length and expected queueing delay in units of *packets* and *packet transmission times* respectively. We have now accounted for both cases regarding b.

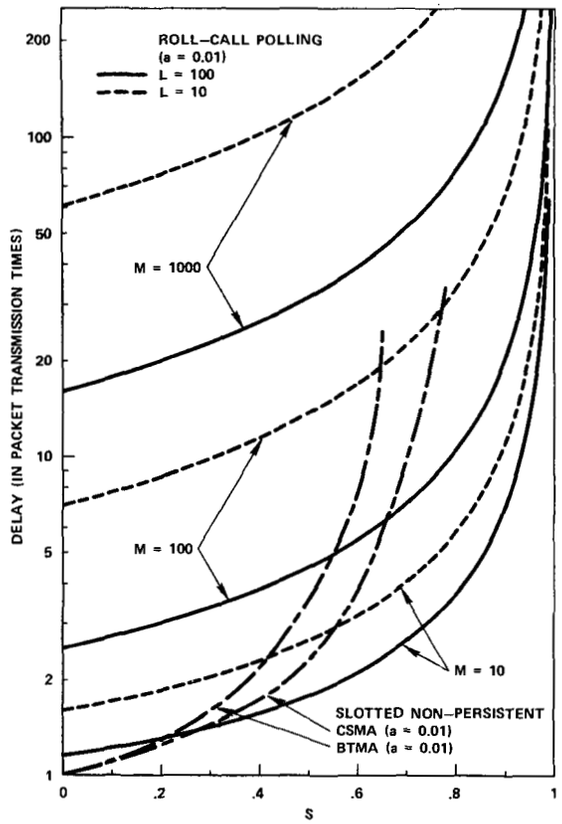


Fig. 7. Packet delay in roll-call polling ($a = 0.01$).

In Fig. 6 we plot the expected queue length for various values of L, a, and M ($L = 10$ 100; $a = 0.01$, 0.05; $M = 10$, 100, 1000). Note that for most values of S, the average queue length is less than one. The expected *total* packet delay is simply equal to

$$\mathcal{D} = E\{\text{queueing delay}\} + 1(\text{packet transmissions}).$$

In Figs. 7 and 8 we plot packet delay versus S for the cases mentioned above ($a = 0.01$, 0.05; $L = 10$, 100; $M = 10$, 100, 1000) along with the throughput-delay curves for CSMA ($a = 0.01$, 0.05) and BTMA ($a = 0.01$) obtained from [1], [8].[6] We note that for the same value of system utilization S, the delay increases with increasing values of M and decreasing values of L; this is of course due to the increase in overhead (transmission of polling messages).[7] Although polling may allow the system to achieve full utilization of the channel ($S = 1$), the delay incurred by a packet is large (mainly for the large M case which is of interest to us) rendering the poll-

[6] It is to be noted that the analysis in [1] and [8] was based on the assumption that a terminal has at most one packet at any time. This comparison is still valid since, especially when M is large and S not too close to 1, the queue length is rarely greater than one packet (see Fig. 6). One could also plot the tail of the distributions (Pr{queue length $> k$}) which can be obtained from the generating function of queue length derived in [14]; however, the expressions are extremely complex and we restrict ourselves to the expected values.

[7] Moreover, polling messages have to be longer for larger M since they have to accommodate longer addresses; this is a second-order effect.

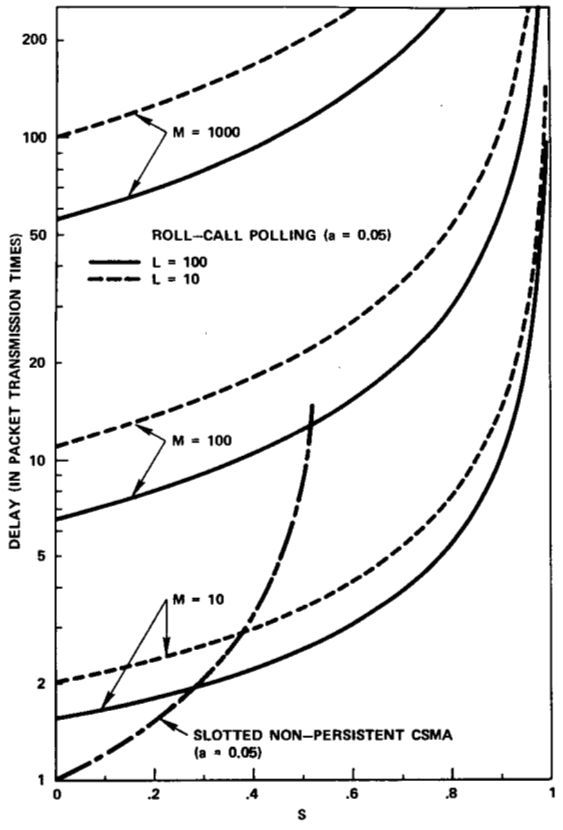


Fig. 8. Packet delay in roll-call polling ($a = 0.05$).

ing technique less attractive and CSMA and BTMA more desirable.

IV. A DYNAMIC RESERVATION TECHNIQUE

Although polling is more common in wired networks than contention, we note, from the previous section, that the former involves considerable overhead and is inefficient when M is large. An attractive alternative is to use a reservation technique which is the subject of this section and which is the key topic of this paper. In the dynamic reservation systems considered here the terminal first makes a request for service on the channel whenever it has a message packet to transmit. It is only when such a request is received at the station that this station will schedule the request; the station maintains a queue of requests and informs the terminal of its position in the queue.

However, since the radio channel is the only means of communication between terminals, the latter cannot schedule the requests themselves in order to avoid conflicts. The contention on the channel due to these request packets is exactly of the same nature as the contention due to the transmission of the message packets themselves, as seen in the random-access techniques. From previous results, random multiple-access modes suggest themselves as a method for multiplexing the requests on the channel. In order to prevent collisions between control packets (requests) and the actual message packets, the channel is either time divided or frequency divided between the two types of data.[8] In this study, we shall select the frequency division method giving rise to the so-called SRMA. The available bandwidth is divided into two channels: one used to transmit control information, the second used for the messages themselves. There are many operational modes. At first, we shall restrict ourselves to the simple one described in the following and called the request answer-to-request message scheme (RAM). In this implementation, the bandwidth allocated for control is further divided into two channels: the request channel and the answer-to-request channel. The request channel will be operated in a random-access mode (ALOHA or CSMA).

Consider now a terminal with a message ready for transmission. To initiate the sending of the message, the terminal sends, on the request channel, a request packet containing information about the address of the terminal, and, in the case of variable length or multipacket messages, the length of the message. At the correct reception of the request packet, the scheduling station computes the time at which the backlog on the message channel will empty and then transmits back to the terminal, on the answer-to-request channel, an answer packet containing the address of the answered terminal and the time at which it can start transmission.

A. Message Delay

We define again the *total message delay* as the time lapse from the moment the message is ready for transmission up to the time the transmission of the message is completed.[9] This total delay is composed of the two following components (see Fig. 9):

a) $\mathcal{D}_1$, the time for the request packet to be successfully received at the station;

b) $\mathcal{D}_2$, the time between reception of the request packet at the station and the end of the message transmission.

Let W again be the total available bandwidth (modulated at 1 bit/Hz·s). Let W_m be the bandwidth allocated to the message channel and $\theta = W_m/W$. The answer-to-request channel is an interference-free channel since the station is the only transmitter. That is, answer packets can be queued at the station and transmitted without conflicts. It is possible to give the answer-to-request channel enough bandwidth W_a such that answer packets do not incur any queueing delay at the station. Indeed, if b_r and b_a are the number of bits per request packet and answer-to-request packet, respectively, then W_a should satisfy

$$W_a \geqslant W_r \frac{b_a}{b_r} \tag{15}$$

where W_r is the bandwidth assigned to the request channel. Since the answer-to-request packet also constitutes the positive acknowledgement for the request packet, we note that the time-out to receive an acknowledgement for the request packet is fixed and simply equal to $T_a + 2\tau$ where T_a is the

[8] The reservation scheme presented by Roberts [13] for packet-switched satellite channels is based on a time-division scheme.

[9] The transmission of message packets on the message channel is free from interference. It is further assumed that the message channel is noiseless and incurs no packet loss.

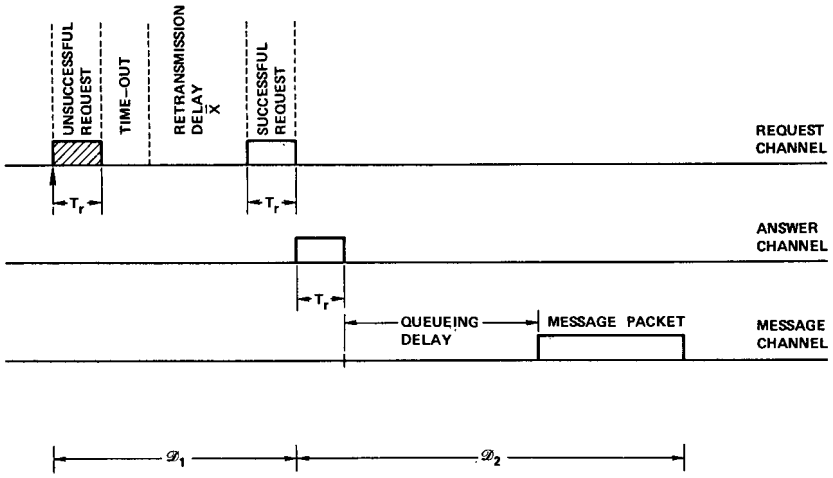


Fig. 9. SRMA.

transmission time of the answer packet, and τ is the one-way propagation delay, assumed to be identical for all terminals.[10]

B. Statement of the Problem

The problem here is twofold. Given that the bandwidth is the limiting resource:

a) find the maximum throughput;

b) for a given throughput, find the optimal bandwidth assignment; that is, the bandwidth assignment which minimizes the total message delay.

C. Delay Analysis

We have so far introduced the following notation:

- W: total bandwidth available;
- W_m: bandwidth assigned to message channel;
- W_r: bandwidth assigned to request channel;
- W_a: bandwidth assigned to answer-to-request channel;
- b_m: number of bits in a message packet (or average number of bits in message in the cases of variable length messages or multipacket messages);
- b_r: number of bits in a request packet;
- b_a: number of bits in an answer-to-request packet;
- θ: fraction of total bandwidth assigned to the message channel (W_m/W).

In addition, we introduce the following notation:

- T_m: transmission time of a message on the message channel, $T_m = b_m/W_m$;
- T_r: transmission time of a request packet on the request channel, $T_r = b_r/W_r$;
- T_a: transmission time of an answer packet on the answer to request channel, $T_a = b_a/W_a$;
- $\eta_a = b_a/b_m$;
- $\eta_r = b_r/b_m$.

In this analysis we assume that the M users (M large) collectively form an independent Poisson source with an aggregate mean packet generation rate of λ packets/s. Under steady-state conditions, λ is also the channel throughput. The maximum generation rate that the total bandwidth W can ever handle is W/b_m. The normalized throughput (average number of packets per transmission time of a packet on the entire bandwidth) denoted again by S is then expressed as

$$S = \frac{\lambda b_m}{W} . \tag{16}$$

Since both control packets contain the same type of information, it is reasonable to assume that $b_a = b_r$ and therefore $\eta_a = \eta_r = \eta$. We further let $W_r = W_a$ and hence $T_r = T_a$. In this case we have

$$W_r = W_a = \frac{(1-\theta)W}{2} . \tag{17}$$

Consider the request channel operated in a random-access mode. The expected delay incurred by a request packet is readily obtained from the simulation results presented in [1], [8]. These throughput-delay tradeoffs are normalized with respect to the packet transmission time on the channel under consideration, namely, T_r for this case. Let S_r denote the normalized input rate on the request channel; we have

$$S_r = \lambda T_r = \frac{2\eta S}{1-\theta} . \tag{18}$$

If the request channel is operated under an ALOHA mode, and letting $\mathcal{D}_{\text{ALOHA}}(S_r)$ denote the delay as a function of the input rate in an ALOHA channel, then

$$\mathcal{D}_1 = \mathcal{D}_{\text{ALOHA}}(S_r) \cdot T_r = \mathcal{D}_{\text{ALOHA}}(S_r) \cdot \frac{2\eta}{1-\theta} \left[\frac{b_m}{W} \right] \quad \text{(seconds)}. \tag{19}$$

Similarly, if the request channel is operated under the nonpersistent CSMA protocol, and if $\mathcal{D}_{\text{NPCSMA}}(S_r, a_r)$ denote the delay as a function of the normalized input rate (S_r) and the normalized propagation delay (a_r) as displayed in Fig. 10,

[10] Propagation delays are not shown in Fig. 9.

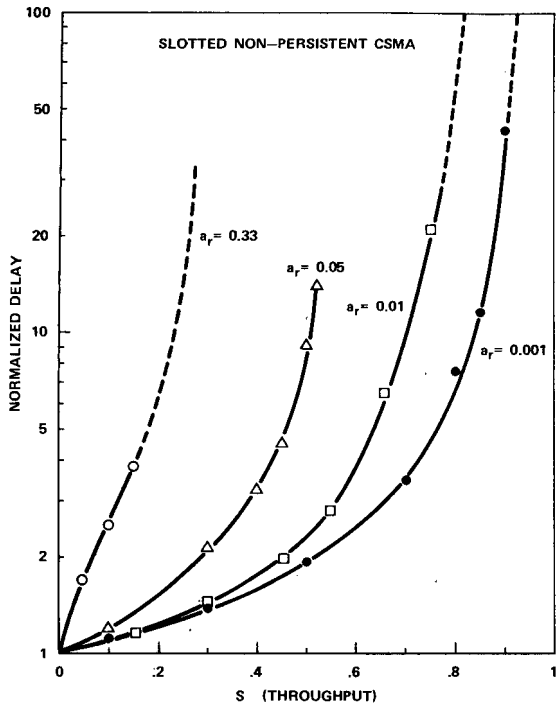


Fig. 10. Slotted nonpersistent CSMA: packet delay for various a_r.

then

$$\mathcal{D}_1 = \mathcal{D}_{\text{NPCSMA}}(S_r, a_r)\frac{2\eta}{1-\theta}\left[\frac{b_m}{W}\right] \quad \text{(seconds)} \tag{20}$$

where S_r is given in (18) and a_r is expressed as

$$a_r = \frac{\tau}{T_r} = \frac{(1-\theta)\tau}{2\eta}\frac{W}{b_m} \tag{21}$$

(In the case where W = 100 kbit/s, b_m = 1000 bit and τ = 100 μs, then $a_r = [(1-\theta)/2\eta] \times 10^{-2}$.)

To estimate the delay $\mathcal{D}_2$, we make the following assumption: the output of the random-access request channel (defined as the process corresponding to the arrival of successful requests at the station) is Poisson with a mean of λ requests per second. In order to verify the above assumption, we examine the interdeparture times (i.e., time between successive successful packets) of the nonpersistent CSMA simulator when $a_r = 0.01$. For this we plot in Fig. 11 (a) and (b) the histograms for various values of S_r, along with the density function of the exponential distribution with mean $1/S_r$ for comparison. We note that except for interdepartures in the range of one or two packet transmission times, the match is acceptable and that the smaller is S_r, the more valid is the assumption.

Under this assumption, the message channel can be modeled as an M/G/1 queueing system, in which the arrival process is the (assumed) Poisson output of the request channel, shifted in time by $T_a + 2\tau$ s (the time-out period). Therefore, using the Pollaczek-Khinchin formula [11] the expected delay $\mathcal{D}_2$ is given by

$$\mathcal{D}_2 = T_a + T_m + 2\tau + \frac{\rho_m T_m(1+c_m{}^2)}{2(1-\rho_m)} \tag{22}$$

where ρ_m is the utilization of the queueing system (message channel) and is given by

$$\rho_m = \lambda T_m = \frac{S}{\theta} \tag{23}$$

and where c_m is the coefficient of variation of the message service time T_m. Finally, for fixed or exponential message length (with an average of b_m bits), we have

$$\mathcal{D}_2 = \left[\frac{2\eta}{1-\theta} + \frac{1}{\theta} + \frac{S/\theta^2}{\delta(1-S/\theta)}\right]\frac{b_m}{W} + 2\tau \quad \text{(seconds)} \tag{24}$$

where

$$\delta = \begin{cases} 2 & \text{if we have deterministic message length} \\ 1 & \text{if we have exponentially distributed message length.} \end{cases}$$

The expected total message delay is therefore

$$\mathcal{D} = \mathcal{D}_1 + \mathcal{D}_2.$$

D. Maximum Bandwidth Utilization

Let C_r denote the capacity of the request channel. The following two constraints must always be satisfied:

$$S_r \leqslant C_r$$

$$\rho_m \leqslant 1.$$

For a given bandwidth assignment θ, the maximum input rate S is determined by the tighter of the two above constraints. The maximum bandwidth utilization (also called the system capacity and denoted by C_{SRMA}) is therefore obtained as the solution of the following program:

$$\begin{cases} C_{\text{SRMA}} = \max\limits_{0\leqslant\theta\leqslant 1} S \\ \text{subject to:} \\ S_r = \dfrac{2\eta S}{1-\theta} \leqslant C_r \\ \rho_m = \dfrac{S}{\theta} \leqslant 1 \end{cases} \tag{25}$$

which can be expressed as

$$C_{\text{SRMA}} = \max_{\theta}\left[\min\left(\frac{(1-\theta)C_r}{2\eta}, \theta\right)\right]. \tag{26}$$

The solution is obtained when the following condition is satisfied:

$$\frac{(1-\theta)C_r}{2\eta} = \theta. \tag{27}$$

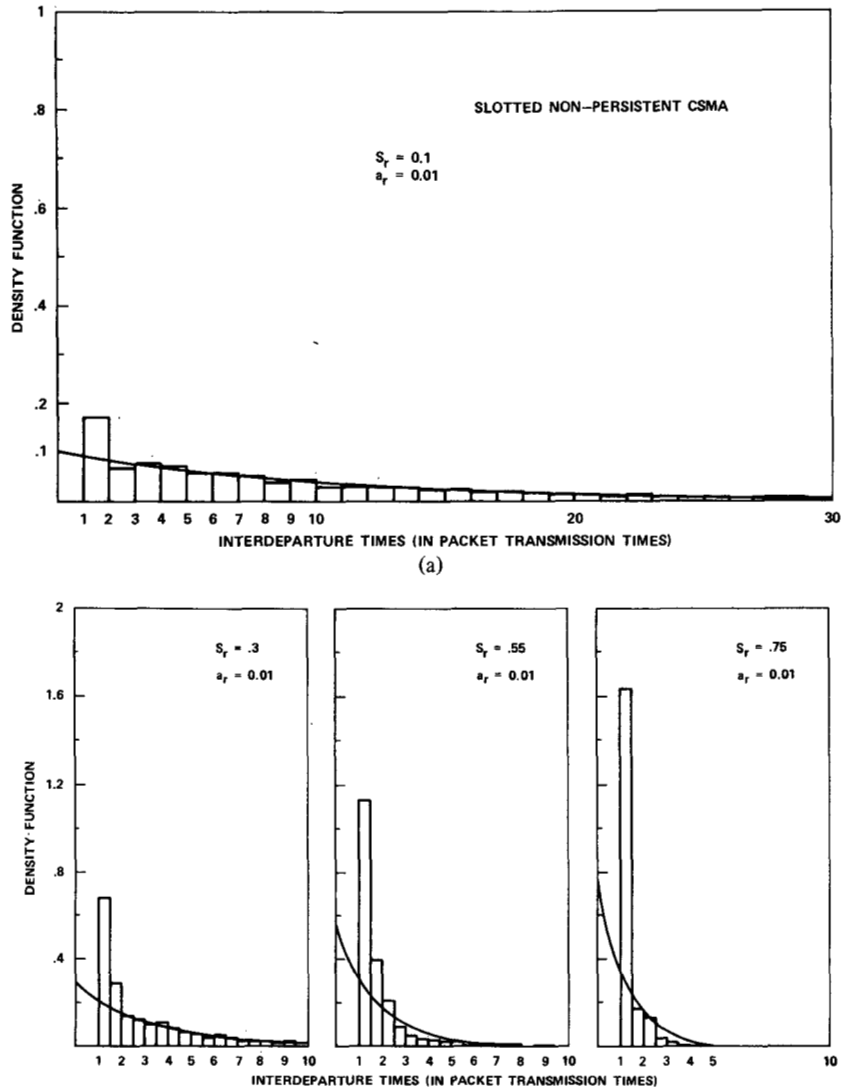


Fig. 11. Histograms of interdeparture times in slotted nonpersistent CSMA. (a) $S_r = 0.1$. (b) $S_r = 0.3, 0.55, 0.75$.

In ALOHA-reservation systems, C_r is constant ($1/2e$ for pure ALOHA, $1/e$ for slotted ALOHA); C_{SRMA} is then easily expressed as

$$C_{\mathrm{SRMA}} = \frac{C_r}{2\eta + C_r} \tag{28}$$

which leads to

$$C_{\mathrm{SRMA}} = \begin{cases} \dfrac{1}{1+4\eta e} & \text{for pure ALOHA} \\ \dfrac{1}{1+2\eta e} & \text{for slotted ALOHA.} \end{cases} \tag{29}$$

In carrier sense SRMA systems, C_r is a function of a_r which itself, by (21), is a function of θ. Given $C_r(a_r)$ (see Fig. 10) the solution of (27) can easily be determined numerically or graphically. To illustrate this, we plot in Fig. 12 the two equations

$$S = \theta$$

$$S = \frac{(1-\theta)C_r(a_r)}{2\eta}$$

which define the space of feasible solutions for nonpersistent carrier sense SRMA. C_{SRMA} lies at the intersection of these two constraints.

E. Numerical Results and Discussion

System Capacity: In Fig. 13 we plot the SRMA system capacity versus η (which represents a relative measure of the overhead due to control information) for the following access modes: pure ALOHA SRMA, slotted ALOHA SRMA, slotted nonpersistent carrier sense SRMA ($\tau W/b_m = 0.01, 0.05$). In addition, we show the system capacity for both ALOHA and CSMA modes. We note that the system capacity in SRMA reaches 1 for very small η. A case of interest considered throughout this paper corresponds to $b_m = 1000$ bits and b_r, anywhere from 10 to 100 bits (b_r is directly related to the

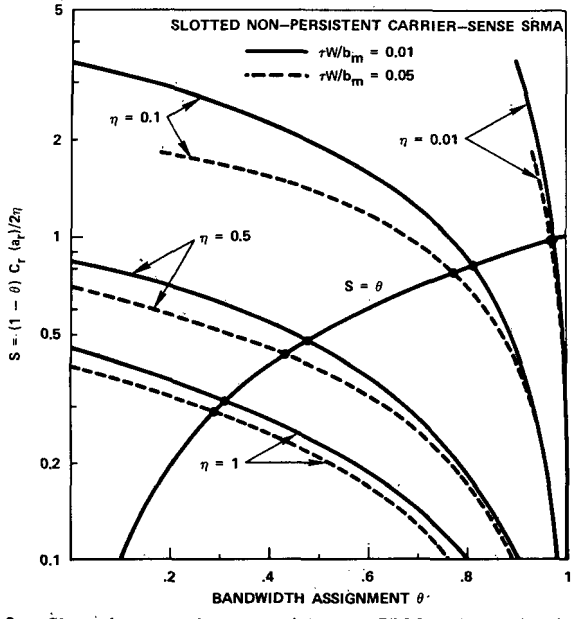


Fig. 12. Slotted nonpersistent carrier sense SRMA: determination of maximum channel utilization.

number of terminals in the population, since addressing information increases logarithmically with increasing M). Thus, the interesting range for η is 0.01 to 0.1. For $\eta > 0.01$, the effect on the system capacity of the random access used to operate the request channel is important: a high improvement is gained when the request channel is operated in slotted nonpersistent CSMA as compared to ALOHA. On the other hand, in comparing the capacity of SRMA to the capacity of random access modes, we note that SRMA is superior for relatively small values of η, and it is these which are of practical interest.

Delay Considerations: Let us restrict ourselves here to $\tau W/b_m = 0.01$. For given η and S, the total message delay $\mathcal{D}$ is a function of θ, the bandwidth assignment. As an example, this is shown in Fig. 14, for slotted nonpersistent carrier sense SRMA with fixed message length (packet) and $\tau W/b_m = 0.01$. Similar plots can be obtained for other random-access modes used for the request channel. For each value of S, we see that θ must lie in a feasible range denoted as $[\theta_{\min}, \theta_{\max}]$, where $\theta_{\min}$ is determined by the constraint $\rho_m = 1$ and then $\theta_{\min} = S$, and $\theta_{\max}$ is determined by the constraint $S_r = C_r(a_r)$. For small values of θ (θ close to $\theta_{\min}$), the major part of delay is due to $\mathcal{D}_2$; for θ close to $\theta_{\max}$, it is due to $\mathcal{D}_1$. The optimal bandwidth assignment is defined as the value of θ which minimizes total delay. We note that the higher the load is, the more critical is the choice of θ_{opt}. (Bearing in mind that random access is more unstable when the load is higher,[11] one tends to choose a value of θ slightly below the optimum, even though the delay is then slightly larger.) The minimum delay for ALOHA-SRMA and slotted nonpersistent carrier sense SRMA is shown in Fig. 15 as a function of S for various values of η. In comparing the two systems between themselves, we again note an important improvement in using CSMA for the request channel. The improvement is more important when larger values of η are involved.

Finally, in Fig. 16 we compare carrier sense SRMA with the random-access modes ALOHA, CSMA, and BTMA; we note that unless η is large (0.1 and above), there is a value of S below which CSMA or BTMA performs better than SRMA and above which the opposite is true. This is mainly due to the following facts.

a) For small S, reservation systems exhibit delays larger than one packet transmission time due to the transmission time of both the request and the answer-to-request packets.

b) For $\eta < 0.1$, CSMA and BTMA exhibit a system capacity lower than the capacity of SRMA (see Fig. 13).

For reference purposes, we also plot in Fig. 16 the $M/D/1$ curve which corresponds to the absolutely best performance one can achieve under statistical load; it consists of a system where one is able to buffer the demands placed on the channel therefore scheduling the transmission of packets (at no cost) in such a way as to avoid conflicts.

Moreover, in comparing SRMA to polling, we note from Figs. 7 and 16 that when $M > 100$ (large population case), SRMA far exceeds the performance of polling.

F. Another Implementation

Another version of SRMA, called the RM scheme, consists of the following implementation. The total available bandwidth is divided into only two channels: the request channel and the message channel. As before, the request channel will be operated in a random-access mode. A terminal with a message ready for transmission sends a request packet (containing its ID) on the request channel. When correctly received by the scheduling station, the request packet joins the request queue. Requests may be serviced on a "first come first served" basis (or any other scheduling algorithm). When the message channel is available, an answer packet (containing the ID of a queued terminal scheduled for transmission) is transmitted by the station on the message channel. After hearing its own ID repeated by the station, the terminal starts transmitting its message on the message channel. If a terminal does not hear its own ID repeated by the scheduling station within a certain appropriate time after the request is sent, the original transmission of the request packet is assumed to be unsuccessful (conflicted with). The request packet is then retransmitted.

The time to receive an answer to a correctly received request packet is equal to the delay incurred by a request packet waiting in the request queue for the message channel to be available. This delay is a random variable. Since the terminal repeats the request if it does not receive any answer within a certain time-out interval (even though the request may have already been correctly received), we note that the terminal undertakes some "additional transmissions" of a request packet following the successful one; the shorter the time-

[11] Random-access channels exhibit unstable behavior at most input loads as shown by Kleinrock and Lam [4]. In this last reference, the dynamic behavior and stability of an ALOHA channel are considered; quantitative estimates for the relative stability of the channel are given, which indicate the need for special control procedures to avoid a collapse. Optimal control procedures have been found [12], [15]. It has been shown [8] that CSMA exhibits similar unstable behavior. However, contrary to ALOHA channels where steady-state performance is badly degraded when true stability must be guaranteed, CSMA provides excellent stable performance even with as large a population as 1000 terminals! Furthermore, the application of adaptive channel control can further improve channel performance. For more details, the reader is referred to the forthcoming Part IV of this series on packet switching in radio channels [16].

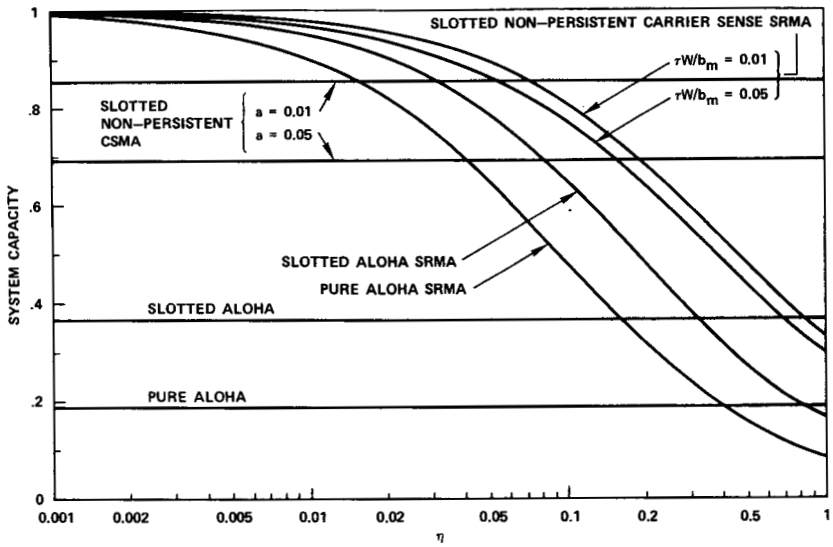


Fig. 13. SRMA: channel capacity versus η.

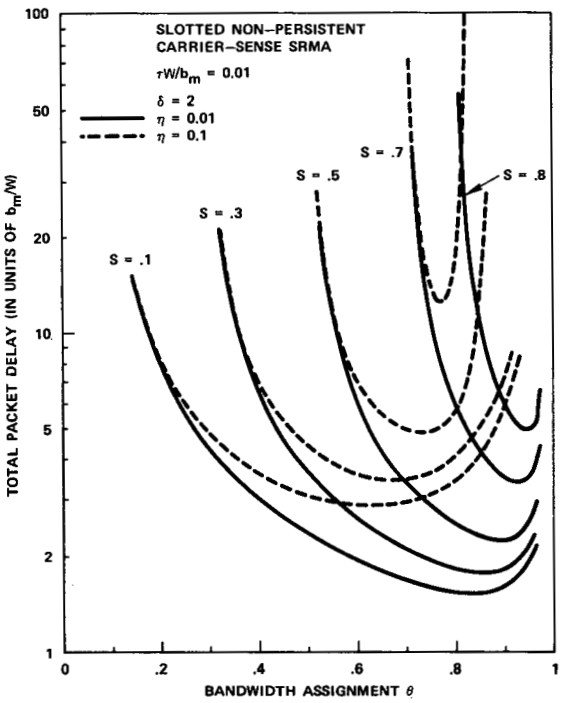


Fig. 14. Slotted nonpersistent carrier sense SRMA: packet delay versus bandwidth assignment.

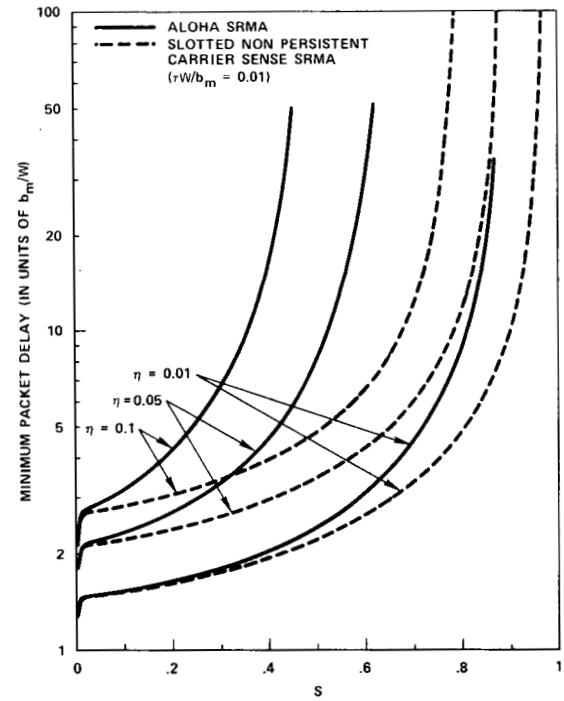


Fig. 15. Minimum packet delay in SRMA.

out period is, the larger is the traffic on the request channel and hence the smaller is the probability of success. On the other hand, the larger the time-out period is, the larger is the probability of success of a request packet, but the longer is the retransmission delay in case of conflict. The problem here consists of the following.

a) For a given load S and a given capacity assignment θ, find the optimum time-out which minimizes the delay for a request packet to be correctly received.

b) For a given load, find the optimal capacity assignment which minimizes tht total packet delay.

This problem has been studied through simulation. The large number of system variables (θ, S, η, time-out period, and retransmission delays on the request channel) render the experimental design task a rather tedious one. It was carried out only to the extent of showing that the performance of the RM scheme is comparable and even slightly superior to the RAM scheme, as considered above. This is summarized in Table I.

V. CONCLUSION

In this paper we reviewed various ways of allocating a channel of limited bandwidth to M user terminals communicating with a station; we also introduced and analyzed a contention system suitable to ground packet radio networks called SRMA. These many modes were compared with regard to throughput and delay.

When we have a large population of bursty users, random

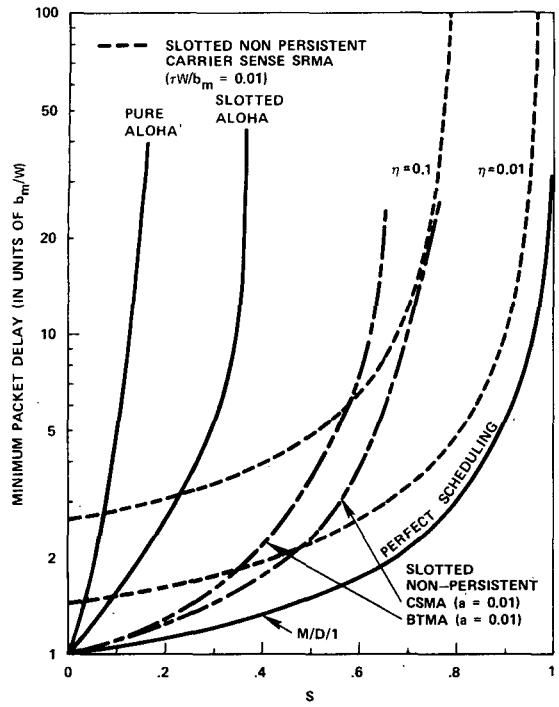


Fig. 16. Final comparison.

TABLE I
MINIMUM DELAY FOR BOTH RAM AND RM SCHEMES

	$\eta = 0.01$		$\eta = 0.1$	
S	RAM	RM	RAM	RM
0.3	1.8	1.5	3.4	2.3
0.7	3.5	2.75	13	6.2

access was shown to be far superior to a fixed-channel assignment; polling was also shown to be inferior to random access due to the large overhead caused by the need for control and slot synchronization. SRMA, on the other hand, represents an interesting scheme since it is both simple and efficient over a large range. From the final comparison performed in the previous section, it is to be noted that there exists no scheme which is consistently superior to all others. The performance of each is dependent on the several system parameters (a, η, S); so also is the selection of the best scheme.

REFERENCES

[1] L. Kleinrock and F. Tobagi, "Packet switching in radio channels: Part I—Carrier sense multiple access modes and their throughput delay characteristics," *IEEE Trans. Commun.*, vol. COM-23, pp. 1400-1416, Dec. 1975.

[2] F. Tobagi and L. Kleinrock, "Packet switching in radio channels: Part II—The hidden terminal problem in carrier sense multiple access and the busy tone solution," *IEEE Trans. Commun.*, vol. COM-23, pp. 1417-1433, Dec. 1975.

[3] N. Abramson, "The ALOHA system—Another alternative for computer communications," in *1970 Fall Joint Comput. Conf., AFIPS Conf. Proc.*, vol. 37. Montvale, NJ, 1970, pp. 281-285.

[4] L. Kleinrock and S. Lam, "Packet switching in a multiaccess broadcast channel: Performance evaluation," *IEEE Trans. Commun.*, vol. COM-23, pp. 410-423, Apr. 1975.

[5] R. E. Kahn, "The organization of computer resources into a packet radio network," in *1975 Nat. Comput. Conf., AFIPS Conf. Proc.*, vol. 44. Montvale, NJ, 1975, pp. 177-186.

[6] N. Abramson, "Packet switching with satellites," in *1973 Nat. Comput. Conf., AFIPS Conf. Proc.*, vol. 42. Montvale, NJ, 1973, pp. 695-702.

[7] L. Kleinrock and S. Lam, "Packet-switching in a slotted satellite channel," in *1973 Nat. Comput. Conf., AFIPS Conf. Proc.*, vol. 42. Montvale, NJ, 1973, pp. 703-710.

[8] F. Tobagi, "Random access techniques for data transmission over packet switched radio networks," Ph.D. dissertation, Computer Science Dep., School of Engineering and Applied Science, Univ. California, Los Angeles, Dec. 1974, UCLA-ENG 7499.

[9] J. Martin, *Teleprocessing Network Organization.* Englewood Cliffs, NJ: Prentice-Hall, 1970.

[10] L. Roberts, ARPANET Satellite System Notes 8 (NIC Document #11290) and 9 (NIC Document #11291), available from the ARPA Network Information Center, Stanford Research Institute, Menlo Park, CA.

[11] L. Kleinrock, *Queueing Systems, Vol. I, Theory; Vol. II, Computer Applications.* New York: Wiley-Interscience, 1975, 1976.

[12] S. Lam, "Packet switching in a multi-access broadcast channel with applications to satellite communications in a computer network," School of Engineering and Applied Science, University of California, Los Angeles, Apr. 1974, UCLA-ENG 7429.

[13] L. G. Roberts, "Dynamic allocation of satellite capacity through packet reservation," in *1973 Nat. Comput. Conf., AFIPS Conf. Proc.*, vol 42. Montvale, NJ, 1973, pp. 711-716.

[14] A. G. Konheim and B. Meister, "Waiting lines and times in a system with polling," IBM J. Watson Research Center, Yorktown Heights, NY, Rep. RC 3841, May 1972.

[15] S. Lam and L. Kleinrock, "Packet switching in a multiaccess broadcast channel: Dynamic control procedures," *IEEE Trans. Commun.*, vol. COM-23, pp. 891-904, Sept. 1975.

[16] F. A. Tobagi and L. Kleinrock, "Packet switching in radio channels: Part IV—Stability considerations and dynamic control in carrier sense multiple access," to be published.

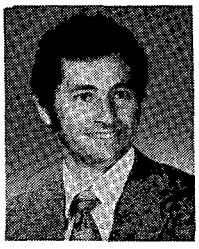
★

Fouad A. Tobagi was born in Beirut, Lebanon on July 18, 1947. He received the Engineering Degree from Ecole Centrale des Arts et Manufactures, Paris, France, in 1970 and the M.S. and Ph.D. degrees in computer science from the University of California, Los Angeles, in 1971 and 1974, respectively.

From 1971 to 1974 he was with the University of California, Los Angeles, where he participated in the ARPA Network Project as a Postgraduate Research Engineer and did research on packet radio communication. During the summer of 1972 he was with the Communications Systems Evaluation and Synthesis Group, IBM J. Watson Research Center, Yorktown Heights, NY. Since December 1974 he has been a Research Staff Project Manager with the ARPA project, Computer Science Department, University of California, Los Angeles. His current research interests include computer communication networks, and packet switching over radio and satellite networks.

From 1967 to 1970 he held a scholarship from the Ministry of Foreign Affairs of the French government. During the academic year 1972-1973 he held an Earl Anthony Fellowship.

★



Leonard Kleinrock (S'55-M'64-SM'71-F'73) was born in New York, NY, on June 13, 1934. He received the B.E.E. degree from the City College of New York, NY, in 1957, and the S.M.E.E. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1959 and 1963, respectively, while participating in the Lincoln Laboratory Staff Associate Program.

From 1951 to 1957, he was employed at the Photobell Company, Inc., New York, NY, an industrial electronics firm. He spent the summers from 1957 to 1961

at the M.I.T. Lincoln Laboratory, Lexington, first in the Digital Computer Group and later in the Systems Analysis Group. At M.I.T. he was a Research Assistant, initially with the Electronic Systems Laboratory, and later with the Research Laboratory for Electronics, where he worked on communication nets in the Information Processing and Transmission Group. After completing his graduate work at the end of 1962, he worked at Lincoln Laboratory on communication nets and on signal detection. In 1963 he accepted a position on the faculty at the University of California, Los Angeles, where he is now Professor of Computer Science. He is a referee for numerous scholarly publications, book reviewer for several publishers, and a consultant for various aerospace, research, and governmental organizations. He is principal investigator of a large contract with the Advanced Research Projects Agency (ARPA) of the Department of Defense. He has published over 70 papers and is the author of *Communication Nets; Stochastic Message Flow and Delay* (New York: McGraw-Hill, 1964), *Queueing Systems, Vol. 1: Theory* and *Vol. 2: Computer Applications* (New York: Wiley-Interscience, 1975 and 1976). His main interests are in communication nets, computer nets, data compression, priority queueing theory, and theoretical studies of time-shared systems.

Dr. Kleinrock is a member of the Tau Beta Pi, Eta Kappa Nu, Sigma Xi, the Operations Research Society of America, and the Association for Computing Machinery. He was awarded a Guggenheim Fellowship in 1971. In 1976 he received the Leonard G. Abraham Prize Paper Award for the best paper in the field of communication systems, published in the 1975 IEEE TRANSACTIONS ON COMMUNICATIONS.

A Generalization of Minimum-Shift-Keying (MSK)-Type Signaling Based Upon Input Data Symbol Pulse Shaping

MARVIN K. SIMON, SENIOR MEMBER, IEEE

Abstract—**In recent years, minimum-shift-keying (MSK) has gained increasing popularity as a modulation technique because of its desirable spectral properties. Quite often, the spectral concentration provided by MSK is not sufficient to meet requirements on out-of-band energy spillover. In these situations, one might apply additional input pulse shaping in such a way as to still maintain constant envelope signals. The properties of such MSK-type signals are the subject of this paper. Specific examples are included as illustrations of the theory both for the binary and *M*-ary cases.**

INTRODUCTION

IT is well known [1]-[3] that minimum-shift-keying (MSK) [4]-[6], which is a speical case of continuous phase frequency-shift-keying (CPFSK) [7], [8] with frequency deviation ratio equal to 0.5, is spectrally equivalent to a form of offset quadrature phase-shift-keying (OQPSK) [9]-[11] in which the symbol pulse shape is a half-cycle sinusoid rather than the usual rectangular form. Perhaps not so well known [12] is the fact that appropriate shaping of the input data symbols allows one to generate an entire class of constant-envelope MSK-type signals, whose spectral properties are more desirable in some applications than those of MSK or OQPSK. The purpose of this paper is to derive and present a set of conditions on the input pulse shaping which in turn describes the class of envelope shapes allowable. The autocorrelation function and power spectral density of this class of signals are then derived and specific examples are given to illustrate the desirable spectral properties alluded to in the above. Such properties are important considerations in system design where interchannel and intersymbol interference degradations must be kept to a minimum. Finally, the results are extended to include *M*-ary MSK which is a special case of *M*-ary CPFSK [20], [21].

Paper approved by the Editor for Data Communication Systems of the IEEE Communications Society for publication without oral presentation. Manuscript received October 20, 1975; revised February 16, 1976. This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, supported by the National Aeronautics and Space Administration under Contract NAS 7-100.

The author is with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91103.

SIGNAL CHARACTERIZATION

When antipodal data are to be transmitted at a rate $R = 1/T$ symbols/s using the MSK modulation technique, then the signal transmitted over the channel is a constant-envelope CPFSK waveform which can be expressed in the form

$$y(t) = \cos\left[\omega_c t + \frac{\pi}{2T}(t - kT)\, d_k + x_k\right] \qquad kT \leq t \leq (k+1)T \tag{1}$$

where ω_c is the carrier radian frequency in rad/s, $\{d_k = \pm 1\}$ is the antipodal data stream, and x_k is an additive phase which is constant over the kth data interval $kT \leq t \leq (k+1)T$ with a value determined by the requirement of phase continuity at the data transition instants $t = kT$ and $t = (k+1)T$. Implicit in the representation of (1) is the fact that the data sequence $\{d_k\}$ is first translated into a binary data waveform with *rectangular* shaped pulses and then frequency modulated onto the carrier.

A generalization of (1) which allows for other than rectangular shaped data pulses is as follows:

$$y(t) = \cos\left[\omega_c t + \frac{\pi}{2T}(t - kT) f_k(t) + x_k\right] \qquad (k-1)T \leq t \leq (k+1)T \tag{2}$$