# Multiaccess Protocols in Packet Communication Systems

FOUAD A. TOBAGI, MEMBER, IEEE

*(Invited Paper)*

***Abstract*—The need for multiaccess protocols arises whenever a resource is shared by many independent contending users. Two major factors contribute to such a situation: the need to share expensive resources in order to achieve their *efficient utilization*, or the need to provide a *high degree of connectivity* for communication among independent subscribers (or both). In data transmission systems, the communication bandwidth is often the prime resource, and it is with respect to this resource that we view multiaccess protocols here. We give in this paper a unified presentation of the various multiaccess techniques which we group into five categories: 1) fixed assignment techniques, 2) random access techniques, 3) centrally controlled demand assignment techniques, 4) demand assignment techniques with distributed control, and 5) mixed strategies. We discuss their applicability to different environments, namely, satellite channels, local area communication networks and multihop store-and-forward broadcast networks, and their applicability to different types of data traffic, namely stream traffic and bursty traffic. We also present the performance of many of the multiaccess protocols in terms of bandwidth utilization and message delay.**

## I. INTRODUCTION

THE need for multiaccess protocols arises whenever a resource is shared (and thus accessed) by a number of independent users. One main reason contributing to such a situation is the need to *share scarce and expensive resources*. An excellent example is typified by time-sharing systems. Time-sharing was developed in the 1960's to make the powerful processing capability of a large computer system available to a large population of users, each of whom has relatively small or infrequent demands so that a dedicated system cannot be economically justified. Two advantages are gained: the smoothing effect of large populations on the demand resulting from the law of large numbers and a lower cost per unit of service resulting from the (almost always existing) economy of scale.

A second major reason contributing to the multiaccess of a common resource by many independent entities is the need for communication among the entities; we refer to this as the *connectivity requirement*. An excellent example today is the telephone system, the main purpose of which is to provide a high degree of connectivity among its subscribers. The multiaccess protocol used in the telephone system is conceptually simple; it merely consists of placing a request for connection to one or several parties, a request which gets honored by the system if all the required resources are available.

Manuscript received September 6, 1979; revised January 7, 1980. This work was supported by the Defense Advanced Research Projects Agency under Contract MDA 903-79-C-0201, Order A03717, monitored by the Office of Naval Research.

The author is with the Computer Systems Laboratory, Stanford, University, Stanford, CA 94305.

### *Packet Communication*

Let us now consider data communication systems, the subject of interest in this paper. Communications engineers have long recognized the need to multiplex expensive transmission facilities and switching equipment. The earliest techniques for doing this were synchronous time-division multiplexing and frequency-division multiplexing. These methods assign a fixed subset of the time-bandwidth space to each of several subscribers and are very successful for stream-type traffic such as voice. With computer traffic however, usually characterized as *bursty*, fixed assignment techniques are not nearly so successful, and to solve this problem, *packet communication systems* have been developed over the past decade [1]-[7]. Packet communication is based on the idea that part or all of the available resources are allocated to one user at a time but for just a short period of time. Here each component of the system is itself a resource which is multiaccessed and shared by the many contending users. To achieve sharing at the component level, customers are required to divide their messages into small units called packets which carry information regarding the source and the intended recipient.

One type of packet communication network, known as the *point-to-point store-and-forward* network, is one where packet switches are interconnected by point-to-point data circuits according to some topological structure. Packets are transmitted independently and pass asynchronously from one switch to another until they reach their destination. The multiplexing of packets on a channel is done by queueing them at each switch until the outgoing channel is free. Typical examples are the ARPANET [7], the Cigale subnetwork [8], TELENET [9], and DATAPAC [10].

Another type of packet transmission network is the (single-hop) *multiaccess/broadcast* network typified by the ALOHA network [11], SATNET [12], and ETHERNET [5]. Here a *single* transmission medium is shared by all subscribers; the medium is allocated to each subscriber for the time required to transmit a single packet. The inherent single-hop broadcast nature of these systems achieves full connectivity at small additional cost. Each subscriber is connected to the common channel through a smart interface which listens to all transmissions and absorbs packets addressed to it.

Yet a third type of packet network can be identified. It is the (multihop) *store-and-forward multiaccess/broadcast* type which combines the features exhibited (and problems encountered) in the two types just mentioned. The best and perhaps only example of this type is the packet radio network (PRNET) sponsored by the Advanced Research Projects Agency [13], [14]. The PRNET is an extension of the

0090-6778/80/0400-0468$00.75 © 1980 IEEE

Authorized licensed use limited to: Sterne Kessler Goldstein Fox. Downloaded on October 19,2022 at 20:45:59 UTC from IEEE Xplore. Restrictions apply.

*ALOHA* network in that it includes many added features such as direct communication by a ground radio network between *mobile* users over *wide* geographical areas, coexistence with possibly different systems in the same frequency band, antijam protection, etc. The key requirement of direct communication over wide geographical areas renders store-and-forward switches, called repeaters, integral components of the system. Furthermore, for easy communication among mobile users and for rapid deployment in military applications, all devices employ omnidirectional antennas and share a high-speed radio channel; hence the multiaccess/broadcast nature of the system.

The main issue of concern here is how to control access to a common channel to efficiently allocate the available communication bandwidth to the many contending users. The solutions to this problem form the set of protocols known as *multiaccess protocols*. These protocols and their performance differ according to the environment in question and the system requirements to be satisfied. We devote the next few paragraphs to summarizing the basic relevant characteristics underlying these environments.

Consider first *satellite channels*. A satellite transponder in a geostationary orbit above the earth provides long-haul communication capabilities. It can receive signals from any earth station in its coverage pattern and can transmit signals to all such earth stations (unless the satellite uses spot beams). Full connectivity and multidestination addressing can both be readily accommodated. The many characteristics regarding data rates, error rates, satellite coverage, channelization, and design of earth stations have been fully discussed in a recent paper by Jacobs *et al.* [12]. Perhaps the most important characteristic relevant to this discussion is the inherent long propagation delay of approximately 0.25 s for a single hop. This delay which is usually long compared to the transmission time of a packet, has a major impact on the bandwidth allocation techniques and on the error and flow control protocols.

In *ground radio* environments, the propagation delay is relatively short compared to the transmission time of a packet, and as we shall see in the sequel, this can be of great advantage in controlling access to a common channel. It is important however to distinguish single-hop environments where direct full connectivity is assumed to prevail, and more complex user environments where, due to geographical distance and/or obstacles opaque to UHF signals, limited direct connectivity is achieved. Clearly, the latter situation is significantly more complex as it gives rise to a multihop system where global control of system operation and resource allocation (whether centralized or distributed) is much harder to accomplish. Another dimension of complexity results from the fact that, unlike satellite environments where earth stations are stationary, ground radio systems must also support mobile users. With mobile users, not only does demand on the system exhibit relatively fast dynamic changes, but the radio propagation characteristics are subject to important variations in received signal strength so that system connectivity is at all times difficult to predict; with these considerations it is important to devise access schemes and system control mechanisms that allow the system to adapt itself to these changes. Furthermore, multipath effects in urban environments can be so disastrous that special signaling schemes, such as spread spectrum, may be in order [14]. Finally, another point of growing concern today is RF spectrum utilization. This is becoming an increasingly predominant factor in determining the structure of radio systems, both in satellite and ground environments. A packet radio system which allows the dynamic allocation of the spectrum to a large population of bursty mobile user needs flexible high performance multiaccess schemes which can take advantage of the law of large numbers, and which permit coexistence of the system with other (possibly different) systems in the same frequency band.

Finally, we consider *local area communication* systems. These span short distances (ranging from a few meters up to a few kilometers) and usually involve high data rates. The transmission medium can be privately owned and inexpensive, such as twisted pair or coaxial cable. Local area environments are characterized by a large and often variable number of devices requiring interconnection, and these are often inexpensive. These situations call for communication networks with simple topologies and simple and inexpensive connection interfaces that can provide great flexibility in accommodating the variability in the environment and that achieve the desired level of reliability. With these constraints, we again face the situation in which a high bandwidth channel is to be shared by independent users. Short propagation delays and high data rates are the main characteristics that are exploited in devising multiaccess schemes appropriate to local area environments.

Multiaccess schemes are evaluated according to various criteria. The performance characteristics that are desirable are, first of all, high bandwidth utilization and low message delays. But a number of other attributes are just as important. The ability for an access protocol to simultaneously support traffic of different types, different priorities, with variable message lengths, and differing delay constraints is essential as higher bandwidth utilization is achieved by the multiplexing of all traffic types. Also, to guarantee proper operation of schemes with distributed control, robustness, defined here as the insensitivity to errors resulting in misinformation, is also most desirable.

Having so far discussed briefly the basic characteristics and system requirements underlying the various communication environments, we now proceed with a discussion of the multiaccess protocols appropriate to these environments.

## II. MULTIACCESS PROTOCOLS

Multiaccess protocols differ by the static or dynamic nature of the bandwidth allocation algorithm, the centralized or distributed nature of the decision-making process, and the degree of adaptivity of the algorithm to changing needs. Accordingly, these protocols can be grouped into five classes. The first class, labeled *fixed assignment techniques*, consists of those techniques which allocate the channel bandwidth to the users in a static fashion, independently of their activity. The second class is that of *random access techniques*. In this class the entire bandwidth is provided to the users as a single channel to be accessed randomly; since collisions may result which degrade the performance of the channel, improved

Authorized licensed use limited to: Sterne Kessler Goldstein Fox. Downloaded on October 19,2022 at 20:45:59 UTC from IEEE Xplore. Restrictions apply.

performance can be achieved by either synchronizing users so that their transmissions coincide with the boundaries of time slots, by sensing carrier prior to transmission, or both. The third and fourth classes correspond to *demand assignment* techniques. Demand assignment techniques require that explicit control information regarding the users' need for the communication resource be exchanged. A distinction is made between those techniques in which the decision-making is centralized (constituting the third class in question), and those techniques in which all users individually execute a distributed algorithm based on control information exchanged among them. The latter constitute the fourth class. The fifth class, labeled *adaptive strategies and mixed modes*, includes those techniques which consist of a mixture of several distinct modes, and those strategies in which the choice of an access scheme is itself adaptive to the varying need, in the hope that near-optimum performance will be achieved at all times.

We describe here the various protocols known today, either implemented or proposed, and discuss their performance and applicability to the different environments introduced in Section I. For this we consider the (conceptually) simplest situation consisting of $M$ users wishing to communicate over a channel. This situation arises typically in a satellite communication environment or in a single-hop ground radio environment.

### A. Fixed Assignment Techniques

Fixed assignment techniques consist of allocating the channel to the user, independently of their activity, by partitioning the time-bandwidth space into slots which are assigned in a static predetermined fashion. These techniques take two common forms: *orthogonal*, such as frequency division multiple access (FDMA) or synchronous time division multiple access (TDMA), and "*quasi-orthogonal*" such as code division multiple access (CDMA).

1) *FDMA and TDMA*: FDMA consists of assigning to each user a fraction of the bandwidth and confining its access to the allocated subband. Orthogonality is achieved in the frequency domain. FDMA is relatively simple to implement and requires no real time coordination among the users.

TDMA consists of assigning fixed predetermined channel time slots to each user; the user has access to the entire channel bandwith, but only during its allocated slots. Here, signaling waveforms are orthogonal in time.

In the author's opinion, a number of disadvantages exist for FDMA when compared to TDMA. FDMA wastes a fraction of the bandwidth to achieve adequate frequency separation. FDMA is also characterized by a lack of flexibility in performing changes in the allocation of the bandwidth and certainly the lack of broadcast operation. The major disadvantages in TDMA are the need to provide A/D converters for overlap traffic such as voice, and rapid burst synchronization and sufficient burst separation to avoid time overlap. However, it has been shown that guard bands of less than 200 ns are achievable (as in INTELSAT's MAT-1 TDMA system, for example) and many operational systems are moving towards the use of TDMA [16]. Timing at an earth station is provided by a global time reference established either explicitly by a reference station, or implicitly by measurement of the propagation delay from the earth station to the transponder. In order to allow the TDMA modems to acquire frequency, phase, bit timing and bit framing synchronization for each received burst, a preamble is included in front of each burst requiring typically from 100 to 200 bit times. Thus clearly, TDMA is more complex to implement than FDMA, but an important advantage is the connectivity which results from the fact that all receivers listen to the same channel while senders transmit on the same common channel at different times. Accordingly, many network realizations, both in ground and satellite environments, are easier to accomplish [12], [14].

From the performance standpoint it has also been established that TDMA is superior to FDMA in many cases of practical interest. I. Rubin has shown that the random variable representing packet delay is always larger in FDMA than in TDMA [17] for comparable systems. Lam derived the average message delay for a TDMA system with multipacket messages and a nonpreemptive priority queue discipline [18]. There, too, it was shown that TDMA is superior to FDMA.

For both FDMA and TDMA, the fixed preallocation of the frequency or time resource does not have to be equal for all users, but can be tailored to fit their needs (assumed constant). Kosovych studied two TDMA implementations [19]. In the first, called *contiguous assignment*, the users are cyclically ordered in the time sequence in which they have access to the channel. Each user is periodically assigned its *own* fixed time duration. In the second implementation, called *distributed allocation*, all access periods are of equal time duration, but the frequency of accesses can be different from one user to the other. It was shown that for situations in which the transmission overhead (defined as guard time and synchronization preamble time) is large, the contiguous fixed assignment implementation is better suited and provides substantially better performance than distributed fixed assignments, while when the transmission overhead is small, distributed fixed assignments provide slightly better performance.

Finally we note that, even though the allocation can be tailored to the relative need of each user, fixed allocation can be wasteful if the users' demand is highly bursty, as we shall explicitly see in the sequel. Given these limitations, one may increase the channel utilization beyond FDMA and TDMA by using asynchronous time division multiple access (ATDMA), also known as statistical multiplexing [70]. Basically the technique consists of switching the allocation of the channel from one user to another only when the former is idle and the latter is ready to transmit data. Thus the channel is *dynamically* allocated to the various users according to their need. The performance of ATDMA in packet communication systems corresponds to that of a work-conserving single server queueing system, and is the best we can achieve under unpredictable demand. Unfortunately, it is not always possible to accomplish the necessary coordination among the users. This mode of multiplexing is possible only when several colocated users (such as at the same earth station) are sharing a single point-to-point channel.

2) *CDMA*: Unlike FDMA and TDMA, code division multi-

Authorized licensed use limited to: Sterne Kessler Goldstein Fox. Downloaded on October 19,2022 at 20:45:59 UTC from IEEE Xplore. Restrictions apply.

ple access allows overlap in transmission both in the frequency and time coordinates. It achieves orthogonality by the use of different signaling codes in conjunction with matched filters (or equivalently, correlation detection) at the intended receivers. Multiple orthogonal codes are obtained at the expense of increased bandwidth requirements (in order to spread the waveforms); this also results in a lack of flexibility in interconnecting all users (unless, of course, matched filters corresponding to all codes are provided at all receivers). However, CDMA has the advantage of allowing the coexistence of several systems in the same band, as long as different codes are used for different systems. Moreover, it is also possible to separate, by "capture," time overlapping signaling waveforms with the same code, thus achieving connectivity and efficient spectrum utilization. This interesting possibility falls into the class of random access techniques and is addressed in the following subsection.

### *B. Random Access Techniques*

In computer communication, much data traffic is characterized as bursty e.g., interactive terminal traffic. Burstiness is a result of the high degree of randomness seen in the message generation time and size, and of the relatively low-delay constraint required by the user. If one were to observe the user's behavior over a period of time, one would see that the user requires the communications resources rather infrequently; but when he does, he requires a rapid response. That is, there is an inherently large peak-to-average ratio in the required data transmission rate. If fixed subchannel allocation schemes are used, then one must assign enough capacity to each subscriber to meet his peak transmission rates with the consequence that the resulting channel utilization is low. A more advantageous approach is to provide a single sharable high-speed channel to the large number of users. The strong law of large numbers then guarantees that, with a very high probability, the demand at any instant will be approximately equal to the sum of the average demands of that population. As stated in the introduction, packet communication is a natural means to achieve sharing of the common channel. When dealing with shared channels in a packet-switched mode, one must be prepared to resolve conflicts which arise when more than one demand is placed upon the channel. For example, in packet-switched radio channels, whenever a portion of one user's transmission overlaps with another user's transmission, the two collide and "destroy" each other (unless a code division multiple-access scheme is used). The existence of some positive acknowledgment scheme permits the transmitter to determine if his transmission is successful or not. The problem is how to control the access to the common channel in a fashion which produces, under the physical constraints of simplicity and hardware implementation, an acceptable level of performance. The difficulty in controlling a channel which must carry its own control information has given rise to the so-called random-access protocols, among others. We describe these here by considering again single-hop environments.

1) *ALOHA [20]-[22]*: Historically, the *pure ALOHA* protocol was first used in the ALOHA system, a single-hop terminal access network developed in 1970 at the University of Hawaii, employing packet-switching on a radio channel [11], [20]. The simplest of its kind, pure ALOHA permits a user to transmit any time it desires. If they do so, and within some appropriate time-out period it receives an acknowledgment from the destination (the central computer), then it knows that no conflict occurred. Otherwise it assumes that a collision occurred and it must retransmit. To avoid continuously repeated conflicts, the retransmission delay is randomized across the transmitting devices, thus spreading the retry packets over time. A slotted version, referred to as *slotted ALOHA*, is obtained by dividing time into slots of duration equal to the transmission time of a single packet (assuming constant-length packets)[21], [22]. Each user is required to synchronize the start of transmission of its packets to coincide with the slot boundary. When two packets conflict, they will overlap completely rather than partially, providing an increase in channel efficiency over pure ALOHA. Due to conflicts and idle channel time, the maximum channel efficiency available using ALOHA is less than 100 percent, 18 percent for pure ALOHA and 36 percent for sloted ALOHA. Both schemes are theoretically applicable to satellite, ground radio and local bus environments. The slotted version has the advantage of efficiency, but in multihop ground radio, it has the disadvantage that synchronization may be hard to achieve.

Although the maximum achievable channel utilization is low, the ALOHA schemes are superior to fixed assignment schemes when there is a large population of bursty users. This point is illustrated in comparing the performance of FDMA with that of slotted ALOHA when $M$ users, each of which generates packets at a rate of $\Lambda$ packets per second, share a radio channel of W·Hz [23]. Figs. 1 and 2 display the constant delay contours in the $(M, \Lambda)$ and $(W, \Lambda)$ planes, respectively, showing the important improvement gained in terms of bandwidth required, population size supported, and delay achieved when the users are bursty.

2) *Carrier Sense Multiple Access (CSMA) [24], [25]*: In ground radio environments the channel can be characterized as wideband with a propagation delay between any source-destination pair that is small compared to the packet transmission time. In such an environment one may attempt to avoid collisions by listening to the carrier due to another user's transmission before transmitting, and inhibiting transmission if the channel is sensed busy. This feature gives rise to a random access scheme known as carrier sense multiple access (CSMA) [24], [25]. While in the ALOHA scheme only one action could be taken by the terminals, namely, to transmit, here many strategies are possible so that many CSMA protocols exist differing according to action that a terminal takes to transmit a packet after sensing the channel. In all cases, however, when a terminal learns that its transmission had incurred a collision, it reschedules the transmission of the packet according to the randomly distributed delay. At this new point in time, the transmitter senses the channel again and repeats the algorithm dictated by the protocol. There are two main CSMA protocols known as *nonpersistent* and *p-persistent* CSMA depending on whether the transmission by a station which finds the channel busy

Authorized licensed use limited to: Sterne Kessler Goldstein Fox. Downloaded on October 19,2022 at 20:45:59 UTC from IEEE Xplore. Restrictions apply.

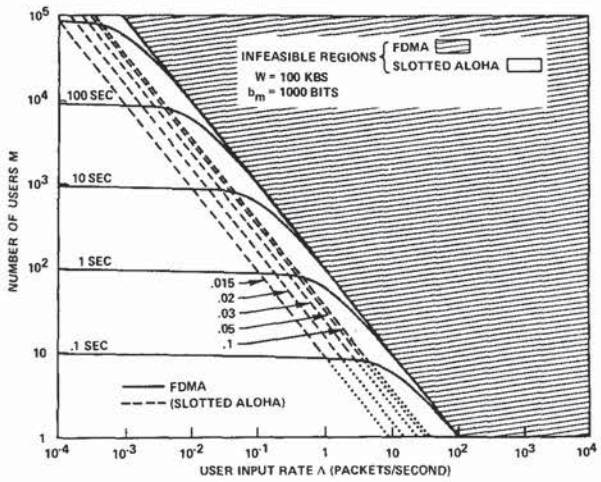


Fig. 1. FDMA and slotted ALOHA access: performance with 100 kbits/s bandwidth and 1000 bit packets [23].

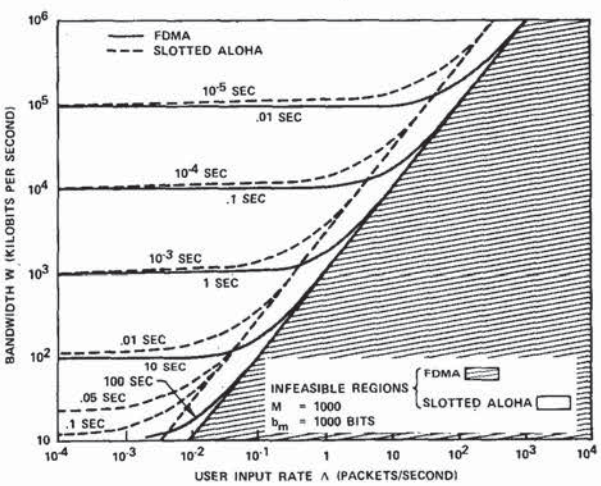


Fig. 2. FDMA and slotted ALOHA random access: bandwidth requirements for 1000 terminals. Contours are for constant delay [23].

is to occur later or immediately following the current one with probability $p$. Many variants and modifications of these two schemes have also been proposed. Thus, in nonpersistent CSMA, 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 according to the retransmission delay distribution. At this new point in time, it senses the channel and repeats the algorithm described.

The 1-persistent CSMA protocol, a special case of $p$-persistent CSMA, was devised in order to (presumably) achieve acceptable throughput by never letting the channel go idle if some ready terminal is available. More precisely, a ready terminal senses the channel and operates as follows:

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

2) If the channel is sensed busy, it waits until the channel goes idle and then immediately transmits the packet with probability one (i.e., persisting on transmitting with $p = 1$).

A slotted version of these CSMA protocols can also be considered in which the time axis is slotted and the slot size is $\tau$ s where $\tau$ is the maximum propagation delay among all pairs. Note that this definition of a slot is different from that used in the description of slotted ALOHA. Here a packet transmission time is equivalent to several slots. We make this distinction by referring to a slot of size $\tau$ s as a "minislot." All terminals are synchronized and are forced to start transmission only at the beginning of a minislot. When a packet's arrival occurs in a minislot, the terminal waits until the next minislot boundary and operates according to the protocols described above.

In the case of a 1-persistent CSMA, we note that whenever two or more terminals become ready during a packet transmission period, they wait for the channel to become idle (at the end of that transmission) and then they all transmit with probability one. A conflict will also occur with probability one. The idea of randomizing the starting time of transmission of packets accumulating at the end of a transmission period seems reasonable for interference reduction and throughput improvement. Thus we have the $p$-persistent scheme which involves including an additional parameter $p$, the probability that a ready packet persists ($1 - p$ being the probability of delaying transmission by $\tau$ seconds, the propagation delay). The parameter $p$ is chosen to reduce the level of interference while keeping the idle periods between any two consecutive nonoverlapped transmission as small as possible.

More precisely, the $p$-persistent CSMA protocol consists of the following: the time axis is minislotted and the system is synchronized such that all terminals begin their transmission at the beginning of a minislot. If a ready terminal senses the channel idle, then with probability $p$, the terminal transmits the packet; and with probability $1 - p$, the terminal delays the transmission of the packet by $\tau$ seconds (i.e., one minislot). If at this new point in time, the channel is still detected idle, the same process is repeated. Otherwise some packet must have started transmission, and the terminal in question schedules the retransmission of the packet according to the retransmission delay distribution (i.e., acts as if it had conflicted and learned about the conflict). If the ready terminal senses the channel busy, it waits until it becomes idle (at the end of the current transmission) and then operates as above.

Packet broadcasting technology has also been shown to be very effective in satisfying many local area in-building communication requirements. A prominent example is ETHERNET, a local communication network which uses CSMA on a tapped coaxial cable to which all the communicating devices are connected [5]. The device connection interface is a passive cable tap so that failure of an interface does not prevent communication among the remaining devices. The use of a single coaxial cable achieves broadcast communication. The only difference between this and the single-hop radio is that, in addition to sensing carrier, it is possible for the transceivers, when they detect interference among several transmissions (including their own), to abort the transmission

Authorized licensed use limited to: Sterne Kessler Goldstein Fox. Downloaded on October 19,2022 at 20:45:59 UTC from IEEE Xplore. Restrictions apply.

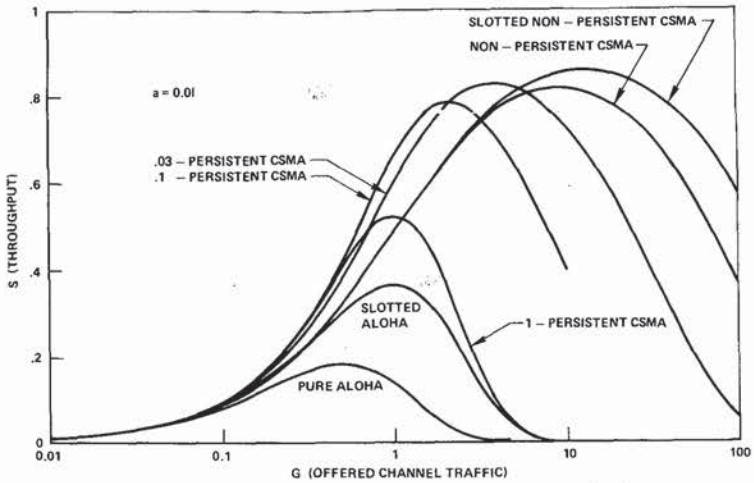


Fig. 3. Throughput for the various random access modes (propagation delay $a = 0.01$) [24].

of colliding packets. This is achieved by having each transmitting device compare the bit stream it is transmitting to the bit stream it sees on the channel. This variation of CSMA is referred to as carrier sense multiple access with collision detection (CSMA-CD) [26].

3) *Performance of Random Access*: Many theoretical studies have been carried out to determine the performance of these random access schemes [20]-[22], [24]-[29]. We summarize here the most important results. Let $S$ denote the aggregate rate of packet generation from the entire population of users, $G$ the rate of packet transmissions (new and repeated, hence $G \geqslant S$), and $D$ the packet delay (defined as the time elapsed between the time that the packet is originated and the time it is successfully received at the destination), all normalized to the (fixed) packet transmission time $T$. Analytic and simulation models provide us, for each random access scheme, with a relationship between $S$ and $G$ (displayed in Fig. 3), and the throughput delay tradeoff (displayed in Fig. 4) for a normalized propagation delay $a = \tau/T = 0.01$. We note that the behavior of these schemes is typical of contention systems, namely that the throughput increases as the offered channel traffic increases from zero, but reaches a maximum value for some optimum value of $G$, and then constantly decreases as $G$ increases beyond that optimal value. Maximizing $S$ with respect to the channel traffic rate $G$ for each of the access modes leads to the channel capacity for that mode. From Fig. 4 we clearly note that $D$ increases as the throughput increases, and reaches infinite values as the throughput approaches the channel capacity. These results show the evident superiority of CSMA over the ALOHA scheme. The CSMA channel capacity in some cases may be as high as 90 percent of the available bandwidth. It is clear however that, as expected, the channel capacity and the throughput-delay tradeoff for the CSMA schemes degrade as the normalized propagation delay ($a = \tau/T$) increases. Fig. 5 illustrates the sensitivity of the channel capacity to $a$.

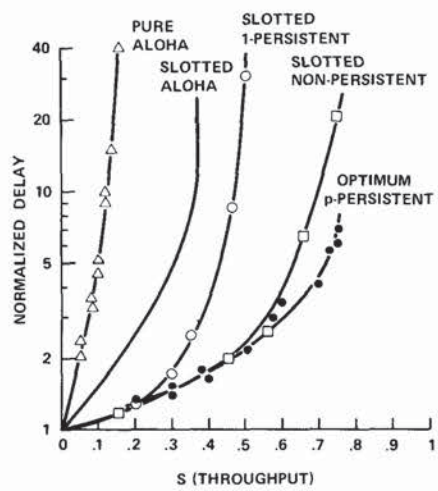


Fig. 4. CSMA and ALOHA: throughput-delay tradeoffs from simulation (propagation delay $a = 0.01$) [24].

CSMA-CD offers even more improvement. A system parameter affecting this improvement is the time required to detect collisions and abort ongoing colliding transmissions. We denote this time by ($\gamma \leqslant T$). The smaller $\gamma$, the better the improvement is. This is illustrated in Fig. 6 where we plot the channel capacity for nonpersistent CSMA versus $\gamma$ for various packet lengths (both expressed in units of $\tau$, the propagation delay). For larger $T$, CSMA provides relatively high channel capacity and thus leaves little margin for improvement; but for small $T$ (e.g., $T = 10$ times the maximum propagation delay on the broadcast bus), the relative improvement is more important (it is about 16 percent when $\gamma$ is just equal to the round trip delay). We also illustrate the improvement due to collision detection by showing packet delay versus $\gamma$ for fixed channel throughput $S$ in Fig. 7. Here, the higher the throughput is, the more significant is the improvement.

Authorized licensed use limited to: Sterne Kessler Goldstein Fox. Downloaded on October 19,2022 at 20:45:59 UTC from IEEE Xplore. Restrictions apply.

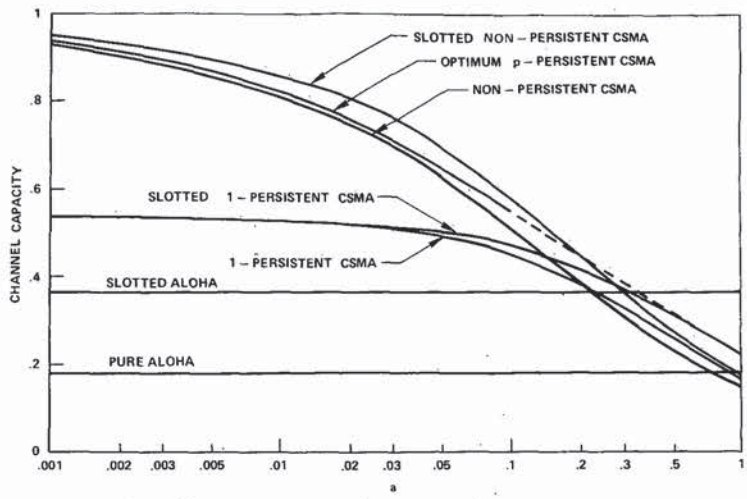


Fig. 5. CSMA and ALOHA: effect of propagation delay $a$ on channel capacity [24].

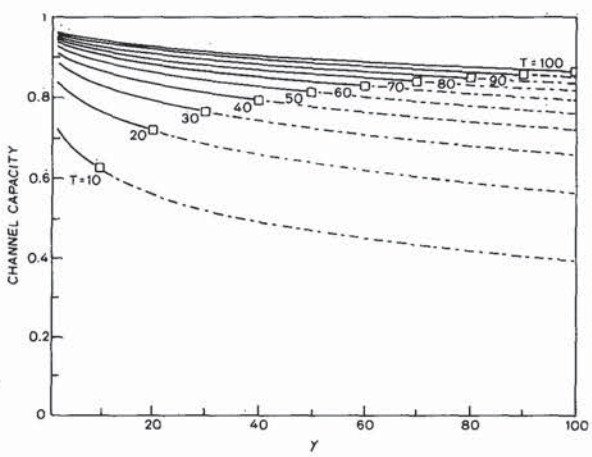


Fig. 6. Channel capacity versus $\gamma$ in nonpersistent CSMA-CD. $T$ = packet transmission time [26].

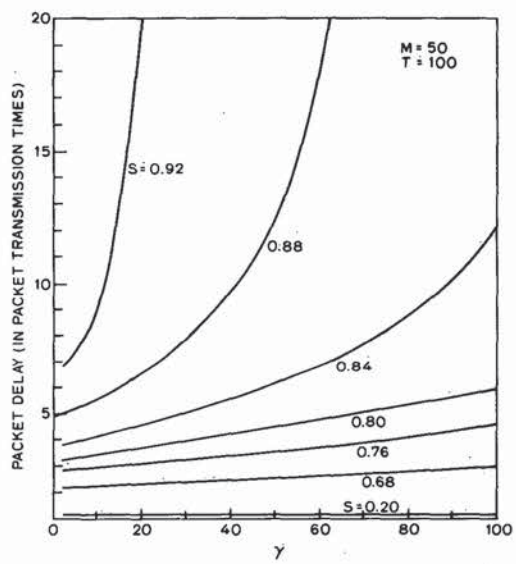


Fig. 7. Packet delay in nonpersistent CSMA-CD at fixed throughput versus collision recovery time $\gamma$ [26].

The results displayed in the above figures have two important assumed conditions, namely 1) acknowledgments are instantaneous, always received correctly and for free (i.e., do not occupy any channel time), and 2) all devices are within range and in line-of-sight of each other so that sensing of all transmissions on the channel is perfect. While Condition 1) is relevant to both ALOHA and CSMA, Condition 2) is mostly relevant to CSMA. We discuss these issues in the following.

4) *Acknowledgment Procedures and Their Effect*: Basically, errors in multiaccess radio channels are due to two major causes: 1) random noise on the radio channel and 2) multiuse interference in the form of overlapping packets. A very reliable method ensuring the integrity of the transmitted data, is the use of an error detecting (e.g., cyclic) block code in conjunction with a positive acknowledgment of each correctly received message. Each packet contains a field for the cyclic checksum in its header. Each receiver responds to a complete packet addressed to it with a correct checksum by having the destination device transmit an acknowledgment packet back to the originating terminal. This acknowledgment contains (among other things) the unique identification of the originating terminal along with a checksum to ensure the integrity of the acknowledgment packet itself.

It is all too evident that acknowledgments will use part of the total available bandwidth (our limited resource). The amount of overhead introduced, as well as the degradation in delay incurred, varies with the mode of operation. When the available bandwidth is provided as a single channel to be shared by both information and acknowledgment packets, then the channel performance will further suffer from interference between information packets and acknowledgment packets unless some kind of priority scheme is provided.

Authorized licensed use limited to: Sterne Kessler Goldstein Fox. Downloaded on October 19,2022 at 20:45:59 UTC from IEEE Xplore. Restrictions apply.

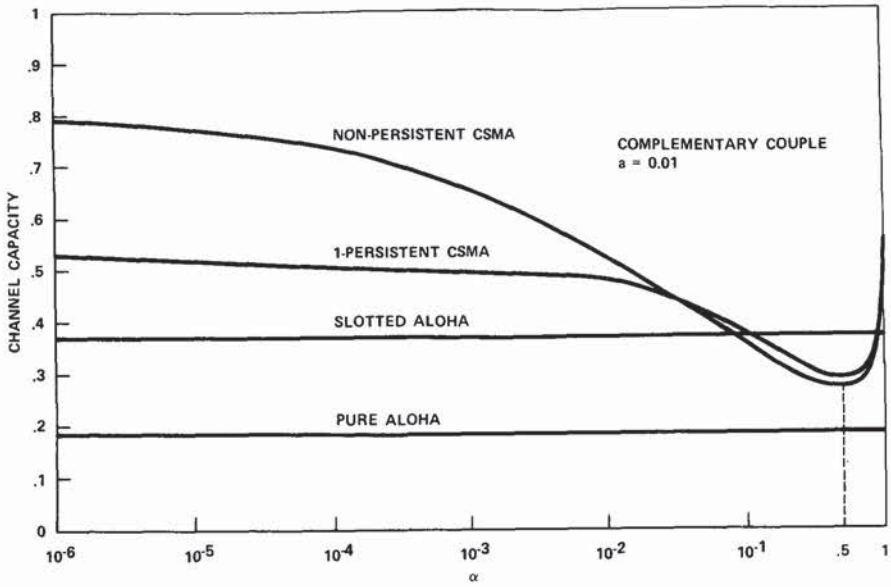


Fig. 8. Complementary couple configuration: channel capacity versus $\alpha$, the relative sizes of the two decoupled populations [28].

The degradation in channel capacity due to the overhead created by the error control traffic has been studied in [30]. It has been shown that, in a common-channel configuration with nonpriority acknowledgment traffic, the channel capacity of slotted ALOHA drops to 14 percent of the channel bandwidth. However, if by some means acknowledgment traffic can be given priority so as to guarantee its transmission free-of-conflict, then the channel capacity for slotted ALOHA can be maintained at around 26 percent (assuming here that an acknowledgment packet uses an entire slot). The effect of acknowledgment traffic on CSMA channels need not be as dramatic since it is very simple to implement schemes which give priority to acknowledgment packets. One mode of operation is as follows [30]:

1) If a terminal, with a packet ready for transmission, senses the channel idle, then the terminal transmits its packet $\tau$ seconds (the propagation delay) later if and only if the channel is still sensed idle.

2) If such a terminal senses the channel busy, then it follows the protocol in question (nonpersistent, 1-persistent,...) repeating step 1) whenever the channel is sensed idle.

3) All acknowledgment packets are transmitted immediately, without incurring the $\tau$ seconds delay.

The capacity of the nonpersistent CSMA protocol with priority acknowledgment and $a = 0.01$ drops gradually from 0.85 to about 0.45 as the acknowledgment packet size increases from 0 to a full packet size.

5) *The Hidden Terminal Problem in CSMA and the Busy-Tone Multiple Access (BTMA) [28]*: We now relax the assumption that all users are in line-of-sight and within range of each other. Typically, two terminals can be within range of the intended receiver, but out-of-range of each other or separated by some physical obstacle opaque to radio signals. The existence of hidden terminals in a radio environment significantly degrades the performance of CSMA. To illustrate this effect, consider a population of users, each of which is communicating with a central station. This station is in line of sight communication with the entire population, but this population is divided into two groups (of relative sizes $\alpha$ and $1 - \alpha$) such that the radio connectivity exists only between users in the same group. Fig. 8 displays the CSMA channel capacity versus $\alpha$, showing that the channel capacity drops drastically as $\alpha$ increases from 0 and reaches a minimum at $\alpha = 0.5$ [28].

Fortunately, in environments where all users communicate with a single central station such as in the ALOHA system, the hidden-terminal problem can be eliminated by frequency dividing the available bandwidth into two separate channels: a busy-tone channel and a message channel, thus giving rise to so called *busy-tone multiple access* (BTMA). The operation of BTMA rests on the fact that, by definition, there exists a central station which is within range and in line of sight of all users. As long as the central station senses carrier on the message channel it transmits a (sine wave) busy-tone signal on the busy-tone channel. It is by sensing carrier on the busy-tone channel that the users' terminals determine the state of the message channel. The action that a terminal takes pertaining to the transmission of the packet is again prescribed by the particular protocol being used.

In CSMA, the difficulty of detecting the presence of a signal on the message channel when this message occupies the entire bandwidth is minor and is therefore neglected. This is not realistic when we are concerned with the (statistical) detection of the (sine wave) busy tone signal on a narrow-band channel. In BTMA, the system's design involves a more complex set of system variables, namely the window detection time, the false alarm probability $F$, and the fraction of bandwidth devoted to the busy-tone signal. For a detailed analysis

Authorized licensed use limited to: Sterne Kessler Goldstein Fox. Downloaded on October 19,2022 at 20:45:59 UTC from IEEE Xplore. Restrictions apply.

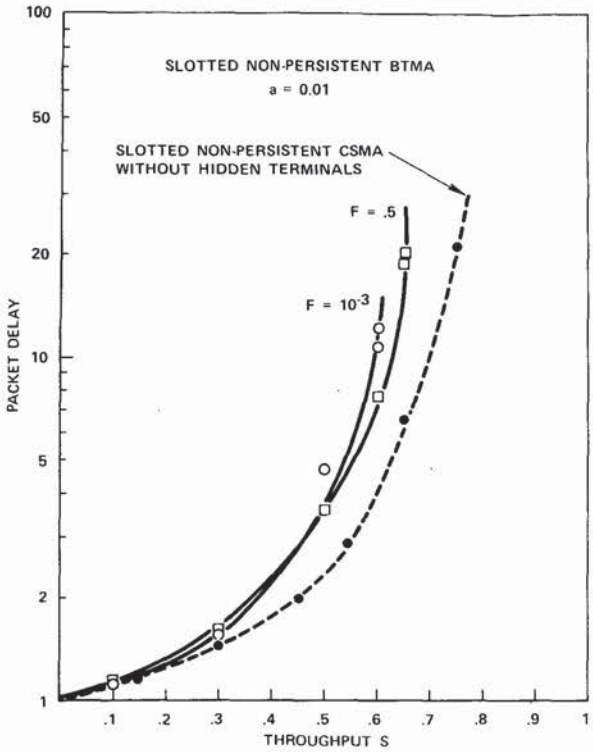


Fig. 9. BTMA: throughput-delay tradeoffs (propagation delay $a$ = 0.01, $F$ = false alarm probability) [28].

of this scheme, the reader is referred to [28]. Fig. 9 displays the throughput-delay tradeoff for BTMA in comparison to CSMA with no hidden terminals, showing the relatively good performance of BTMA.

6) *Dynamic Behavior and Dynamic Control of Random Access Schemes*: The performance results reported upon above were based on renewal theory and probabilistic arguments, assuming that steady-state conditions exist. If one examines in more detail the $(S, G)$ relationships displayed above, one can see that the steady-state may not exist because of an inherent instability of these random-access techniques. This instability is simply explained by the fact that statistical fluctuations in the offered traffic increase the level of mutual interference among transmissions which in turn increases total $G$ which increases the frequency of collisions, and so forth. Such positive feedback causes the throughput to decrease to very low values. Extensive simulation runs performed on a slotted ALOHA channel with an *infinite population* of users have indeed shown that the assumption of channel equilibrium is not strictly speaking valid; in fact after some finite period of quasi-stationary conditions, the channel will drift into saturation with probability one [31]. Thus a more accurate measure of channel performance must reflect the tradeoffs among stability, throughput, and delay. To that effect, Markov models have been formulated to analyze slotted ALOHA and CSMA when $M$ interactive users contend for the channel [31]-[34]. These models permit one not only to derive analytic expressions for the average throughput-delay performance, but also to understand the dynamic behavior of these systems. In particular, it was observed that even in a finite population environment, if the retransmission delay is not sufficiently large, then the stationary performance attained is significantly degraded (low throughput, very high delay), so that, for all practical purposes, the channel is said to have failed; it is then called an unstable channel. With an infinite population, stationary conditions just do not exist; the channel is always unstable, thus confirming the results obtained from simulation, as just discussed. For unstable channels, Kleinrock and Lam [32] defined a stability measure which consists of the average time the system takes, starting from an empty channel, to reach a state determined to be critical. In fact, this critical state partitions the state space into two regions: a safe region, and an unsafe region in which the tendency is towards degraded performance. The stability measure is the average first exit time (FET) into the unsafe region. As long as the system operates in the safe region, the channel performance is acceptable; but then, of course, it is only usable over a finite period of time with an average equal to FET. For more details concerning the determination of FET and the numerical results, the reader is referred to [32], [33].

In the above discussion, it was furthermore assumed that the system parameters were all fixed, time invariant, and state-independent. These systems are referred to as *static*. It is often advantageous to design systems that dynamically adapt to time-varying input and to system state changes, thus providing improved performance. Dynamic adaptability is achieved via dynamic control consisting of time and state dependent parameters. The basic problem then is to find the control functions which provide the best system performance. Markov decision theory has successfully been applied by Lam and Kleinrock to the design and analysis of control procedures suitable to slotted ALOHA in particular and random-access techniques in general [35]. Two main types of control are proposed: an *input* control procedure (ICP) consisting of either accepting or rejecting all new packets generated in the current slot, and a *retransmission* control procedure (RCP) consisting of selecting a retransmission delay; in both cases the action taken is a function of the current system state, defined as the number of active users with outstanding packets. In order to implement such control schemes, each channel user must individually estimate the channel state by observing the channel outcome over some period of time. The control is of a distributed nature, as there is no central station monitoring and broadcasting state information or control actions. In the context of slotted ALOHA, Lam and Kleinrock give some heuristic control-estimation algorithms which prove to be very satisfactory [35]. With appropriate modification and extensions, these algorithms can be applied to CSMA channels as well. These algorithms are best suited to fully connected single-hop type environments. The dynamic control problem in multihop environments is more complex and little progress has yet been made in this area.

7) *Capture*: In the preceding discussions it was assumed that whenever two packet transmissions overlap in time, these packets destroy each other. This assumption is pessi-

Authorized licensed use limited to: Sterne Kessler Goldstein Fox. Downloaded on October 19,2022 at 20:45:59 UTC from IEEE Xplore. Restrictions apply.

mistic as it neglects *capture* effects in radio channels. Capture can be defined as the ability for a receiver to successfully receive a packet (with nonzero probability) although it is partially or totally overlapped by another packet transmission. Capture is mainly due to a discrepancy in receive power between two signals allowing the receiver to correctly receive the stronger; both distance and transmit power contribute to this discrepancy. Clearly capture improves the overall network performance, and, by the means of adaptive transmit power control, it allows one to achieve either fairness to all users, or intentional discrimination. Some of these effects have been addressed in [27], [36].

8) *Spread Spectrum Multiple Access (SSMA)*: Spread spectrum multiple access (SSMA) is the most common form of CDMA whereby each user is assigned a particular code sequence which is modulated on the carrier with the digital data modulated on top of that. Two common forms exist: the frequency-hopped SSMA and the phase-coded SSMA. In the former, as its name indicates, the frequency is periodically changing according to some known pattern; in the latter the carrier is phase modulated by the digital data sequence and the code sequence. SSMA has many applications: it is useful in satellite communications, mobile ground-radio, and computer communication networks [37]. In [14], Kahn *et al.* addressed many of the issues concerning the use of SSMA in packet radio systems. Security, coexistence with other systems, and ability to counteract the effects of multipath are key factors contributing to the choice of SSMA in the PRNET; however one main point of interest in this presentation is the benefit of capture in asynchronous SSMA. Even when several users employ the same code, the effect of interference is minimized by the "capture effect," defined here as the ability of the receiver to "lock on" one packet while all other overlapping packets appear as noise. The receiver locks on a packet by correctly receiving the preamble appended in the front of the transmitted packet. As long as the preamble of different packets do not overlap in time, and the signal strength of the late packets is not too high, capture of the earliest packet can be guaranteed with a high probability. In essence SSMA allows a packet to be captured at the receiver, while CSMA allows a user to capture the channel. CSMA can still be used in conjunction with SSMA. This mode will have the benefit of keeping away all users within hearing distance of the transmitter and thus help keep the capture effect and antijamming capability of the system at the desired level. For a complete discussion of all these issues, the reader is referred to [14].

### C. Centrally Controlled Demand Assignment

We have so far discussed the two extremes in the bandwidth allocation spectrum as far as control over the user's access right is concerned: the tight fixed assignment which has the most rigid control, is nonadaptive to dynamically varying demand, and can be wasteful of capacity if small-delay constraints are to be met; and random access which involves no control, is simple to implement, is adaptive to varying demand, but which, in some situations, can be wasteful of capacity due to collisions. In this and the following subsections, we examine demand assignment techniques which require that explicit information regarding the need for the communication resource be exchanged. We distinguish those demand assignments which are controlled by a central scheduler from those which employ a distributed algorithm executed by all users. We address centrally controlled assignments in the present subsection.

1) *Circuit Oriented Systems*: In these systems, the bandwidth is divided into FDMA or TDMA subchannels which are assigned on demand. The satellite SPADE system, for example, has a pool of FDMA subchannels which get allocated on request [38]. It uses one subchannel operated in a TDMA fashion with one slot per frame permanently assigned to each user to handle the requests and releases of FDMA circuits. Intelsat's MAT-1 system uses the TDMA approach [39]. TDMA subchannels are periodically reallocated to meet the varying needs of earth stations.

The Advanced Mobile Phone Service (AMPS), recently introduced by Bell Laboratories, is yet another example of a centrally controlled FDMA system [40]. The uniqueness of this system, however, lies in an efficient management of the spectrum based on space division multiple access (SDMA). That is, each subchannel in the pool of FDMA channels is allocated to different users in separate geographical areas, thus considerably increasing the spectrum utilization. To accomplish space division, the AMPS system has a cellular structure and uses a centralized handoff procedure (executed by a central office) which reroutes the telephone connections to other available subchannels as the mobile users move from one cell to another.

Given the significant setup times required in allocating subchannels, the above systems are attractive only when applications have stream-type traffic. When traffic is bursty, we again turn to packet-oriented systems, such as in the following.

2) *Polling Systems*: In packet oriented systems, polling is one of two modes used to centrally control access to the communication bandwidth, again provided as a single high-speed channel. A central controller sends polling messages to the terminals, one-by-one, asking the polled terminal to transmit. For this the station may have a polling list giving the order in which the terminals are polled. If the polled terminal has something to transmit, it goes ahead; if not, a negative reply (or absence of reply) is received by the controller, which then polls the next terminal in sequence. Polling requires this constant exchange of control messages between the controller and the terminals, and is efficient only if 1) the round-trip propagation delay is small, 2) the overhead due to polling messages is low, and 3) the user population is not a large bursty one. Polling has been analyzed by Konheim and Meister [41], and their analysis has been applied to the environment of $M$ users sharing a radio channel in [23]. Denoting by $L$ the ratio of the data message length to the polling message length, and by $a$ the ratio of propagation delay to message transmission time, Fig. 10 displays numerical results corresponding to some typical values of $L$ and $a$. These curves show that indeed as the population size increases, thus containing more and more bursty users, the performance of polling degrades significantly. Channel utilization can reach

Authorized licensed use limited to: Sterne Kessler Goldstein Fox. Downloaded on October 19,2022 at 20:45:59 UTC from IEEE Xplore. Restrictions apply.

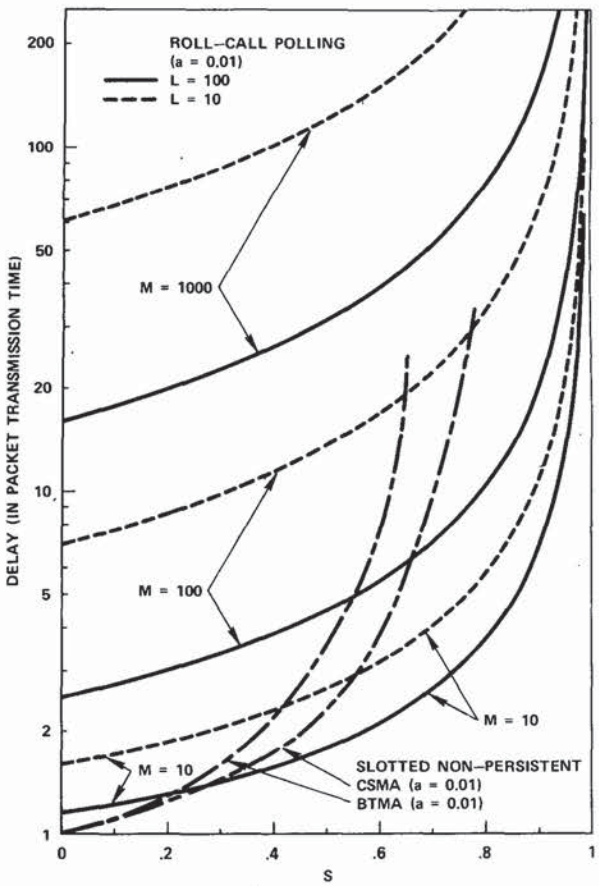


Fig. 10. Packet delay in roll-call polling. $L$ = ratio of data message length to polling message length, $a$ = normalized propagation delay, $M$ = number of stations [23].

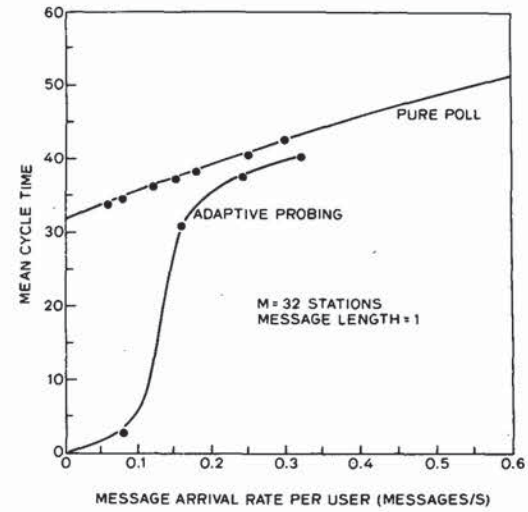


Fig. 11. Polling and adaptive probing: mean cycle time versus message arrival rate (simulation results—32 stations) [42].

100 percent of the channel bandwidth if the terminals are allowed to empty their buffers when they are polled. But as a result, the variance of packet delay can become intolerably large.

3) *Adaptive Polling or Probing [42]*: The primary limitation of polling in lightly loaded systems is the high overhead incurred in determining which of the terminals have messages. In order to decrease this overhead, a modified polling technique, based on a tree searching algorithm, and referred to as *probing*, has been proposed [42]. This technique assumes that the central controller can *broadcast* signals to all terminals. First the controller interrogates all terminals asking if any of them has a message to transmit, and repeats this question until some terminals respond by putting a signal on the line. When a positive response is received, the central station breaks down the population into subsets (according to some tree structure) and repeats the question to each of the subsets. This can be performed simply, for example by using binary addresses for the terminals and by transmitting as probing signal the common prefix of the addresses of a group of terminals. The process is continued until the terminals having messages are identified. When a single terminal is interrogated, it transmits its message.

Assume that the number of terminals is a power of 2, say $M = 2^n$. Let a cycle be recursively defined as the time required for the polling and transmission of all messages that were generated in the preceding cycle. If a single terminal has a message to transmit, probing requires $2n + 1$ inquiries per cycle as opposed to $2^n$ for conventional polling; but if all terminals have messages, probing requires $2^{n+1} - 1$ inquiries as opposed to $2^n$ for conventional pollings. To avoid incurring such a penalty when the system is heavily loaded, the probing technique can be made adaptive whereby the controller starts a cycle by probing smaller groups as the probability of terminals having messages increases. In particular, the group size may be considered a function of the duration of the immediately preceding polling cycle. Simulation of the adaptive probing technique has shown that this scheme is always superior to polling in that its mean cycle time is always smaller than that of polling. Fig. 11 displays the mean cycle time (obtained from simulation) as a function of the message arrival rate for both polling and probing [42]. Reference [42] did not provide any results concerning message delay, but it is intuitively clear that the smaller the mean cycle time is, the lower is the average delay.

4) *Split-Channel Reservation Multiple Access (SRMA) [23]*: An attractive alternative to polling is the use of explicit reservation techniques. In dynamic reservation systems, it is the terminal which makes a request for service on some channel whenever it has a message to transmit. The central scheduler manages a queue of requests and informs the terminal of its allocated time.

Since the channel is the only means of communication among terminals, the main problem here is, once again, how to communicate the request to the central scheduler. The contention on the channel of these request packets is of exactly the same nature as the contention of the data packets themselves. Fixed assignment and random access techniques suggest themselves, but it is clear from previous results that random access modes for multiplexing the requests on the channel would be

Authorized licensed use limited to: Sterne Kessler Goldstein Fox. Downloaded on October 19,2022 at 20:45:59 UTC from IEEE Xplore. Restrictions apply.

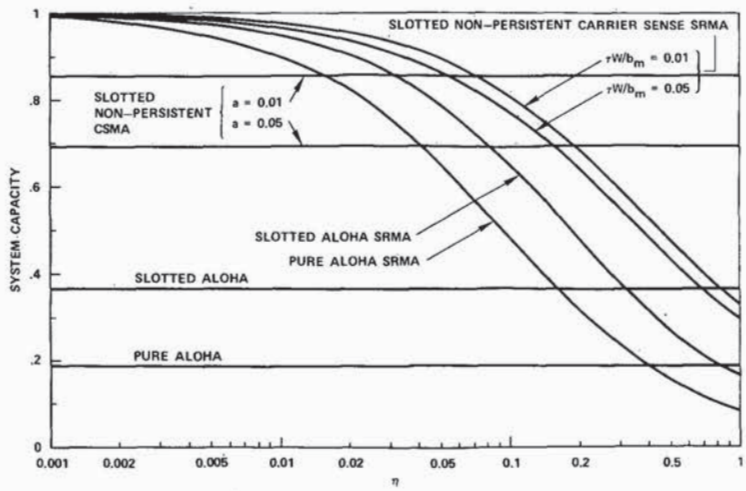


Fig. 12. SRMA: channel capacity versus $\eta$, ratio of request packet length to data packet length (normalized propagation delay of 0.01 and 0.05) [23].

more efficient. Furthermore, in order to prevent collisions between the requests and the actual message packets, the available bandwidth is either time divided or frequency divided between the two types of data. In the split-channel reservation multiple access (SRMA) scheme, frequency division of a ground radio channel is considered [23]. The available bandwidth is divided into two channels: one used to transmit control information, the second used for the data messages themselves. With this configuration, there are many operational modes. In the request/answer-to-request/message scheme (RAM), the bandwidth allocated for control is further divided into two channels: the request channel and the answer-to-request channel. The request channel is operated in a random access mode (ALOHA or CSMA). Upon correct reception of the request packet, the scheduling station computes the time at which the backlog on the message channel will empty and transmits an answer packet back to the terminal, on the answer-to-request channel, containing the address of the terminal and the time at which it can start transmission. Another version of SRMA, called the RM scheme, consists of having only two channels: the request channel and the message 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. The request packet is then retransmitted.

We now examine the performance of SRMA. Let $\eta$ denote the ratio of request packet length to data packet length, this representing a measure of the overhead due to control information. In Fig. 12 we plot the (RAM) SRMA system capacity versus $\eta$ for the following access modes: pure ALOHA SRMA, slotted ALOHA SRMA, and slotted nonpersistent carrier sense SRMA. In addition, we show the system capacity for both ALOHA and CSMA. We note that the system capacity in SRMA reaches 1 for very small $\eta$. Typical values for $\eta$ fall in the range (0.01, 0.1). Fig. 12 shows that a high improvement is gained when the request channel is operated in slotted nonpersistent CSMA as compared to ALOHA. The delay for ALOHA-SRMA and slotted nonpersistent carrier sense SRMA (normalized to $b_m/W$, where $W$ denotes again the total channel bandwidth, and $b_m$ is the number of bits per packet) is shown in Fig. 13 as a function of $S$ for various values of $\eta$. We again note an important improvement in using CSMA for the request channel. Finally, in Fig. 14 we compare carrier sense SRMA with the random access modes ALOHA, CSMA, BTMA, and $M/D/1$, the perfect scheduling with fixed size packets and Poisson sources. We note that unless $\eta$ 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.

5) *Global Scheduling Multiple Access (GSMA) [43]*: GSMA is a conflict-free reservation multiaccess scheme suitable for a high-speed data bus, which is based on the time-division concept for reservation. Here too a scheduler oversees all scheduling tasks. The users, all connected to the same line, listen for scheduling assignments and transmit in accordance with the slot allocation initiated by the scheduler. The channel time is divided into frames (of variable lengths). A frame is partitioned into two subframes: a subframe of status slots statically assigned to the users (in a fixed TDMA mode) to request data slot allocation, and a subframe of data slots, each sufficient to transmit a data packet of $P$ bits. The fixed assignment of the status slots removes the need to transmit users' ID's and thus reduces the size of these slots. In each frame, a user can be allocated a number of data slots which does not exceed the number of packets generated at the user during the preceding frame or a maximum number specified, whichever is smaller. As a consequence

Authorized licensed use limited to: Sterne Kessler Goldstein Fox. Downloaded on October 19,2022 at 20:45:59 UTC from IEEE Xplore. Restrictions apply.

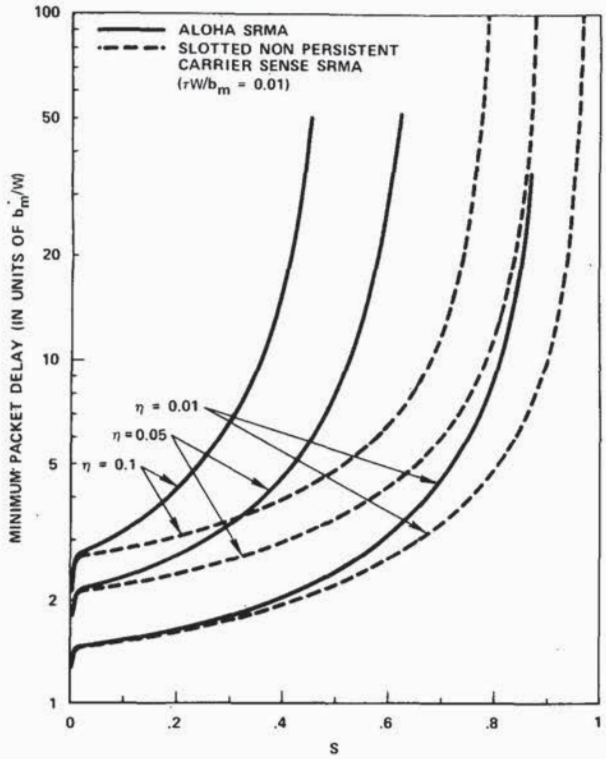


Fig. 13. Packet delay in SRMA (normalized propagation delay = 0.01) [23].

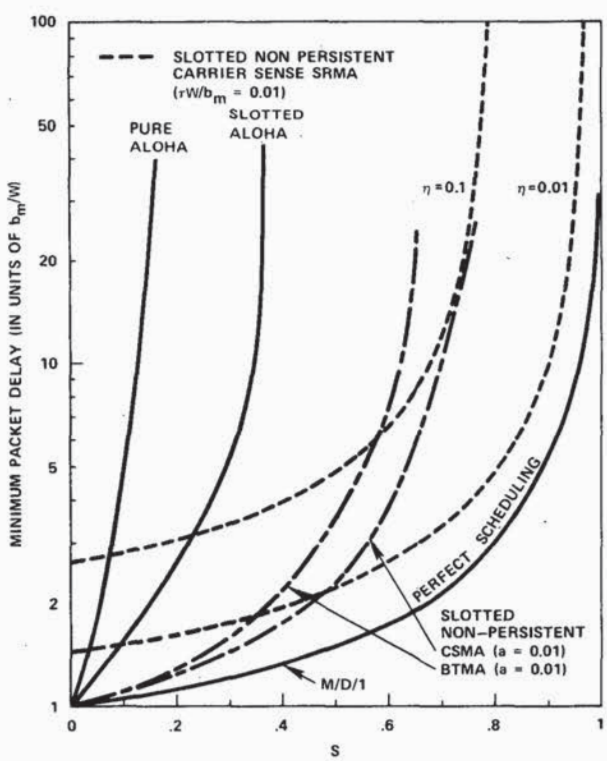


Fig. 14. Comparison of various schemes (parameters defined as in Figs. 12 and 13) [23].

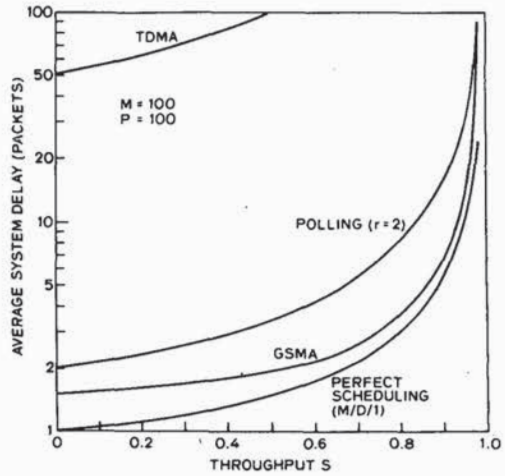


Fig. 15. GSMA: throughput delay performance ($M$ = number of stations, $P$ = number of bits per data packet) [43].

each active user is guaranteed at least one slot per frame. Fig. 15 displays the performance of GSMA (with $P = 100$ and number of stations $M = 100$) in comparison to polling (for some typical parameter values regarding the polling overhead $r$) and $M/D/1$ (the perfect scheduling). This illustrates some improvement gained in GSMA over polling [43].

### D. Demand Assignment with Distributed Control

There are two reasons why distributed control is desirable. The first is *reliability*; with distributed control the system is not dependent on the proper operation of a central scheduler. The second is improved *performance*, especially when dealing with systems with long propagation delays, such as those using satellite channels. Indeed, if an earth station were to play the role of a scheduler, the minimum packet delay in a packet reservation scheme would be three times the round-trip propagation delay. (Of course, this can be decreased if on-board processing is available.) With distributed control, this minimum delay can be brought down to twice the round-trip delay or less without affecting the bandwidth utilization. Clearly, in slotted ALOHA, the best random access scheme available for satellite channels, the minimum packet delay is exactly one round-trip delay; but this is guaranteed only for a channel utilization approaching zero! In fact, the inherent long propagation delay in satellite channels is really the nasty characteristic that makes this environment "more distributed" than the single-hop ground radio or local area environments. In the latter, we have seen that efficient random access schemes, such as CSMA, are available; and the shorter the propagation delay, the better the CSMA performance. With zero propagation delay, collisions in CSMA can be completely avoided and CSMA's performance then corresponds to that of an $M/D/1$ queue,[1] the best we can achieve under random demand. In fact, as observed in [44], when the propagation

[1] This corresponsdence applies to CSMA with fixed size packets and Poisson sources.

Authorized licensed use limited to: Sterne Kessler Goldstein Fox. Downloaded on October 19,2022 at 20:45:59 UTC from IEEE Xplore. Restrictions apply.

delay is zero we no longer have a distributed environment, and the cost of creating a common queue disappears.

The basic element underlying all distributed algorithms is the need to exchange control information among the users, either explicitly or implicitly. Using this information, all users then execute independently the same algorithm resulting in some coordination in their actions. Clearly, it is essential that all users receive the same information regarding the demand placed on the channel and its usage in order to achieve a global optimum, and thus distributed algorithms are most attractive in fully connected systems. This attribute is not always present in ground radio environments, but certainly exists in satellite environments due to their inherent broadcast nature.[2] The long-delay/broadcast combination of attributes has been one of the reasons why many distributed control algorithms have been proposed in the context of satellite environments. We examine in this subsection distributed control algorithms suitable for each of our three environments (satellite, ground radio and local area), starting with satellite channels.

1) *Reservation-ALOHA [45]*: Reservation-ALOHA for a satellite channel is based on a slotted time axis, where the slots are organized into frames of equal size. The duration of a frame must be greater than the satellite propagation delay. A user who has successfully accessed a slot in a frame is guaranteed access to the same slot in the succeeding frame and this continues until the user stops using it. "Unused" slots, however, are free to be accessed by all users in a slotted ALOHA contention mode. An unused slot in the *current* frame is a slot which, in the *preceding* frame, either was idle or contained a collision. (Note again the effect of long delays on the control procedure.) Users need to simply maintain a history of the usage of each slot for just one frame duration. Since no request is explicitly issued by the user, this scheme has been referred to as an *implicit reservation* scheme. Clearly Reservation-ALOHA is effective only if the users generate stream type traffic or long multipacket messages. Its performance will degrade significantly with single packet messages, as every time a packet is successful the corresponding slot in the following frame is likely to remain empty.

2) *A First-in First-out (FIFO) Reservation Scheme [46]*: In this scheme, reservations are made explicitly. Time division is used to provide a reservation subchannel. The channel time is slotted as before, but every so often a slot is divided into $V$ small slots which are used for the transmission of reservation packets (as well as possibly acknowledgments and small data packets); these packets contend on the $V$ small slots in a slotted ALOHA mode. All other slots are data slots and are used on a reservation basis, free of conflict. The frequency of occurrence of reservation slots can be made adaptive to the load on the channel and the need to make new reservations. This adaptivity can be achieved as a result of the time-division of bandwidth allocation between reservations and data packets.

[2] This is valid unless the satellite uses spot beams, in which case we may lose on the connectivity requirement but gain the benefits of space division multiple access (SDMA).

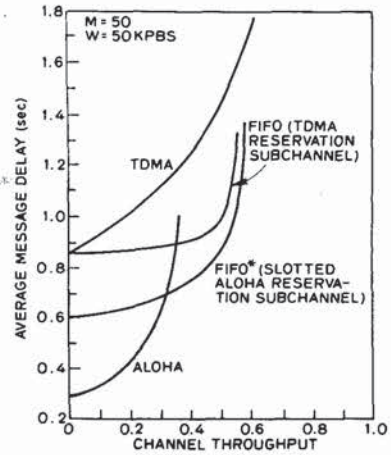


Fig. 16. Slotted ALOHA, TDMA, and FIFO reservation: delay throughput tradeoff for 50 users and single-packet messages in a satellite environment [66].

To execute the reservation mechanism properly, each station must maintain information on the number of outstanding reservations (the "queue in the sky") and the slots at which its own reservations begin. These are determined by the FIFO discipline based on the successful reservations received. Each successful reservation can accommodate up to a design maximum of, say, eight packets, thus preventing stations from acquiring exclusivity of the channel for long periods of time. To maintain synchronization of control information at the proper time, and to acquire the correct count of packets in the queue if out-of-sync conditions do occur, each station sends, in its data packet, information regarding the status of its queue. This information is also used by new stations which need to join the queue. The robustness of this system is achieved by a proper encoding of the reservation packets to increase the probability of their correct reception at *all* stations. Furthermore, to limit the effect of errors, a station reacquires synchronization if it detects a collision in one of its reserved slots or an error in a reservation packet.

Fig. 16 compares the throughput-delay tradeoff of the FIFO reservation scheme (operated with either a TDMA or a slotted ALOHA reservation subchannel) to that of TDMA and slotted ALOHA [66]. FIFO-Reservation offers delay improvements over TDMA. When compared to ALOHA, we note that higher system capacity is achieved but at the expense of a higher delay at low channel throughputs (due to a higher overhead).

3) *A Round-Robin (RR) Reservation Scheme [47]*: The basis of this scheme is fixed TDMA assignment, but with the major difference that "unused" slots are assigned to the active stations on a round-robin basis. This is accomplished by organizing packet slots into equal size frames of duration greater than the propagation delay and such that the number of slots in a frame is larger than the number of stations. One slot in each frame is permanently assigned to each station. To allow other stations to know the current state (used or unused) of its own slot, each station is required to transmit

Authorized licensed use limited to: Sterne Kessler Goldstein Fox. Downloaded on October 19,2022 at 20:45:59 UTC from IEEE Xplore. Restrictions apply.

information regarding its own queue of packets piggybacked in the data packet header (transmitted in the previous frame.) A zero count indicates that the slot in question is free. All stations maintain a table of all stations' queue lengths, allowing them to allocate among themselves free unassigned slots in the current frame. Round-robin is the discipline proposed by Binder [47], but other scheduling disciplines can be used as well. A station recovers its slot by deliberately causing a conflict in that slot which other users detect. For a station which was previously idle, initial acquisition of queue information is required and is achieved by having one of the stations transmit its table at various times. However, it is interesting to note that in this scheme, while acquiring queue synchronization, a station can always reclaim and use its own assigned slot.

The above three schemes have been proposed for satellite channels. All assumed fixed size slots, and thus can be implemented in systems which have been built for synchronous TDMA. The effect of large propagation delay is important. Framing is used in two of the schemes to deal with it, with the frame duration being equal to or longer than the propagation delay. Due to their dynamic nature, these protocols perform better than synchronous TDMA. However, when compared to random access (namely ALOHA here), they offer higher capacity, but also higher delay at low throughput. If used in systems with small propagation delay, such as ground radio, then they will perform significantly better, and are expected to have a performance comparable to SRMA. In fact, due to the inherent small propagation delay in ground radio environments, other access modes with distributed control are also possible if all devices are in line-of-sight and within range of each other. We describe these in the following.

4) *Minislotted Alternating Priorities (MSAP) [48]*: MSAP is a conflict-free multiple access scheme suitable for a *small* number of data users. In essence, MSAP is a "carrier-sense" version of polling with distributed control. The time axis is slotted with the minislot size again equal to the maximum propagation delay. All users are synchronized and may start transmission only at the beginning of a minislot. Users are considered to be ordered from 1 to $M$. When a packet transmission ends, the alternating priorities (AP) rule assigns the channel to the same user who transmitted the last packet (say user $i$) if he is still busy; otherwise the channel is assigned to the next user in sequence (i.e., user ($i$, mod $M + 1$)). The latter (and all other users) detects the end of transmission of user $i$ by sensing the absence of carrier over one minislot. At this new point in time, either user ($i$ mod $M + 1$) starts transmission of a packet (which will be detected by all other users) or he is idle in which case a minislot is lost and control of the channel is handed to the next user in sequence. The overhead at each poll in this scheme is simply one minislot.

Scheduling rules other than AP are also possible, namely round-robin (RR) or random order (RO). MSAP, however, exhibits the least overhead incurred in switching control between users. On the other hand, MSRR may be more suitable to environments with unbalanced traffic since then smaller users will be guaranteed more frequent access than with MSAP. These scheduling rules have also appeared in the literature as BRAM, the broadcast recognizing access method. For details, see [72].

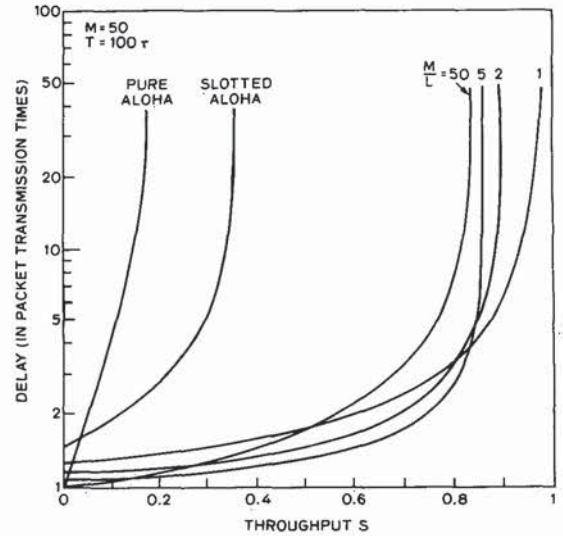


Fig. 17. Assigned-slot listen-before transmission protocol: throughput delay tradeoff for 50 users and $T = 100$ (propagation time $a = 0.01$) [49].

5) *The Assigned-Slot Listen-Before-Transmission Protocol [49]*: MSAP, being a "carrier sense" version of polling, behaves like polling. In particular, as the system load decreases, the overhead incurred in locating a nonidle user increases, and so does the delay. The assigned-slot listen-before-transmission protocol has been proposed to improve on MSAP by allowing several users to share common minislots. In such a case, there exists a tradeoff between the time wasted in collisions, and the time wasted in control overhead. Time is divided into frames, each containing an equal number of minislots (say, $L$). To each minislot of a frame is assigned a given subset of $M/L$ users. A user with a packet ready for transmission in a frame can sense the channel only in his assigned minislot. If the channel is sensed idle, transmission takes place; if not, the packet is rescheduled for transmission in a future frame. A packet transmission spans $T$ slots. The parameter $M/L$ is adjusted according to the load placed on the channel. For high throughput, $M/L = 1$ is found to be optimum. In fact, with $M/L = 1$, the scheme becomes a conflict-free one which approaches MSAP and gives nearly identical results [49]. For very low throughput, $M/L = M$ (i.e., $L = 1$) is found to be optimum; this corresponds to pure CSMA. In between the two extreme cases intermediate values of $M/L$ are optimum. Fig. 17 displays the throughput-delay performance of this scheme for various values of $M/L$ when $M = 50$ and $T = 100$. It also shows how this scheme (and thus, MSAP) compare to CSMA.

6) *Distributed Tree Retransmission Algorithms in Packet Broadcast Channels [71]*: In many of the multiaccess protocols examined above, conflict resolution is achieved by retransmitting randomly in the future. Such a rescheduling discipline in slotted ALOHA achieves a 36 percent bandwidth utilization, but exhibits some sort of instability unless the rescheduling is controlled, as discussed in Section II-B. Tree algorithms

Authorized licensed use limited to: Sterne Kessler Goldstein Fox. Downloaded on October 19,2022 at 20:45:59 UTC from IEEE Xplore. Restrictions apply.

are based on the observation that a contention among several active sources is completely resolved if and only if all the sources are somehow subdivided into groups such that each group contains at most one active source. (Such observation is similar to that made in the probing technique discussed in Section II-C3). In its simplest form, the tree algorithm consists of the following. Each source corresponds to a leaf on a *binary* tree. The channel time axis is slotted and the slots are grouped into pairs. Each slot in a pair corresponds to one of the two subtrees of the node being visited. Starting with the root node of the tree, we let all terminals in each of the two subtrees of the root transmit in their corresponding slot. If any of the two slots contains a collision, then the algorithm proceeds to the root of the subtree corresponding to the collision and repeats itself. This continues until all the leaves are separated into sets such that each of them contains at most one packet. This is known to all users, as the outcome of the channel is either a successful transmission or an idle slot. Collisions caused by the left subtree (1st slot of a pair) are resolved prior to resolving collisions in the right subtree. This scheme provides a maximum throughput of 0.347 packets/slot, and all moments of the delay are finite if the aggregate packet arrival rate is less than 1/3 packets/slot [71].

Clearly, a binary tree is not always optimum. If, each time we return to the root node, we allow the tree to be reconfigured according to the current traffic conditions, it can be shown that the optimum tree is binary everywhere except for the root node whose optimum degree depends on traffic conditions [71]. The dynamic scheme achieves a throughput of 0.430 packets/slot, and all the moments of the delay are finite for $\lambda < 0.430$ packets/slot. Tree algorithms are implementable in both ground radio and satellite channels as long as the broadcast capability is available.

7) *Distributed Control Algorithms in Local Area Networks*: In addition to the random access schemes described previously, the above two algorithms are also applicable to local area (broadcast) *bus* networks as these exhibit the required characteristics of small propagation delay and full connectivity. But in local area communication, a slightly different topology, namely the *ring* (or loop), has also been widely considered. In the ring topology, messages are not broadcast but rather passed from node to node along unidirectional links, until they return to the originating node. A simple scheme suitable for a ring consists of passing the access right sequentially from node to node around the ring. (Note that in a ring, the physical locations of the nodes define a natural ordering among them.) One implementation of this scheme is exemplified by the Distributed Computing System's network where an 8-bit *control token* is passed sequentially around the ring [50]. Any node with a ready message may, upon receiving the control token, remove the token from the ring, send the message, and then pass on the control token. Another implementation consists of providing a number of *message slots* which are continuously transmitted around the ring. A message slot may be empty or full; a node with a ready message waits to see an empty slot pass by, marks it as full, and uses it to send its message [51]-[53]. A still different strategy is known as the *register insertion* technique [3], [54], [55]. Here a message to be transmitted is first loaded into a shift register. If the ring is idle, the shift register is just transmitted. If not, the register is inserted into the network loop at the next point separating two adjacent messages: the message to be sent is shifted out onto the ring while an incoming message is shifted into the register. The shift register can be removed from the network loop when the transmitted message has returned to it. The insertion of a register has the effect of increasing the transport delay of messages on the ring.

### *E. Adaptive Strategies and Mixed Modes*

We have so far examined quite a large number of multiaccess schemes and compared their performance. One thing is clear: each of these schemes has its advantages and limitations. No one scheme performs better than all others over the entire range of system throughput (except, of course, the hypothetical perfect scheduling, which is clearly unachievable in a distributed environment). If a scheme performs nearly as well as perfect scheduling at low input rates, then it is plagued by a limited achievable channel capacity. Conversely, if a scheme is efficient when the system utilization is high, the overhead accompanying the access control mechanism becomes prohibitively large at low utilization. Although some characteristics of a system (propagation delay, channel speed, etc.) are unlikely to vary during operation, it is certain that the load placed upon the system will be time varying. In the case of a single subscriber type (say with periodic traffic, stream-type traffic, or bursty traffic) the volume of the traffic may be varying; if several subscriber types are simultaneously present, the volume of traffic introduced by each, and therefore the proportional mix of traffic types, may also be time varying.

We have discussed at several points in this paper the dynamic control of a specific access scheme which improved its performance to a certain extent; but such an adaptive control did not change the nature of the access scheme nor the nature of its limitation. Dynamically controlled random access schemes provide improved packet delay over uncontrolled versions, but still exhibit channel capacity less than 1. The adaptive polling technique decreased the overhead at low throughput but only to a certain extent. Actually, what one really needs is a strategy for choosing an access mode which is itself adaptive to the varying need so that optimality is maintained at all times. Clearly, in order to accomplish adaptivity, a certain amount of information is needed by the distributed decision makers. The type and amount of information required by an adaptive strategy, as well as the implementation of the information acquisition mechanism are among the most crucial factors in determining the performance and robustness of the strategy. A great deal of effort has been spent in recent years on such adaptive strategies. We devote this subsection to schemes which fall into this category.

1) *The URN Scheme [56]*: We start with this more recent scheme because of its simplicity, elegance, and the smoothness by which it adapts to varying loads. It has been proposed in the context of a ground radio fully connected environment in which the time axis is divided into packet slots, and all

Authorized licensed use limited to: Sterne Kessler Goldstein Fox. Downloaded on October 19,2022 at 20:45:59 UTC from IEEE Xplore. Restrictions apply.

users are synchronized. Assuming that all users know the exact number $n$ of busy users, the scheme consists of giving full access right (i.e., the right to transmit with probability 1) to some number $k$ of users. A successful transmission will result if there is exactly one busy user among these $k$. The probability of such an event is maximized when $k = [M/n]$, where $[M/n]$ denotes the integer part of $M/n$. This is in contrast to the controlled slotted ALOHA scheme where all users are given the same partial access right: the right to transmit with probability $p = 1/n$. Assume the system is lightly loaded (for instance $n = 1$); a large number of users are given access right (in the example $n = 1$, the number is $k = M$), but only a few and hopefully only one will make use of it (in the example $n = 1$, a successful transmission takes place). As the load increases, $k$ decreases and the access right is gradually restricted. For the extreme case of $n = M$, $k = 1$ and the scheme converges to TDMA. If the sampling of $k$ is random, the urn scheme converges to random TDMA; if the sampling of $k$ is without repetitions from slot to slot until all users have been sampled once, the urn scheme converges to round-robin TDMA.

Two important questions remain: how to estimate $n$, and how to reach a consensus on who the $k$ users are. In [56], Kleinrock and Yemini offer a few alternatives. One possible scheme for estimating $n$ with good accuracy is to include a single reservation minislot at the beginning of each data slot. An idle user who turns busy sends a standard reservation message of few bits. All users are able to detect the following three events: no new busy users, one new busy user, and more than one new busy user (termed an erasure). As it is impossible with this minimal overhead to estimate the exact number of new busy users when the latter is greater than one, errors in estimation result; however analysis and simulation have shown that this error is negligible, and furthermore that the scheme is insensitive to small perturbations in $n$. This last statement is even more important with respect to the robustness of the scheme since it means that all users need not have exactly the same estimate for $n$. As for coordinating the selection of the $k$ users, an effective mechanism is the use of synchronized pseudorandom generators at all users which allow them to draw the same $k$ pseudorandom numbers. Another mechanism, referred to as a round-robin slot sharing window mechanism, consists of having a window of size $k$ move over the population space. When a collision occurs, the window stops and decreases in size. When there is no collision, the tail of the window is advanced to the head of the previous window, and the size is again set to $k$ as determined by $n$.

The improvement obtained by this scheme over slotted ALOHA and TDMA can be seen in Fig. 18 where the throughput-delay performance of all these schemes is displayed for a pupulation size $M = 10$ [56].

2) *Another Adaptive Strategy for the Dynamic Management of Packet Radio Slots [57]*: Another way to achieve adaptivity is as follows. The time axis is again slotted with the slot size equal to a packet transmission time. Slots are grouped into $k$ equivalence classes or subchannels. Slots are

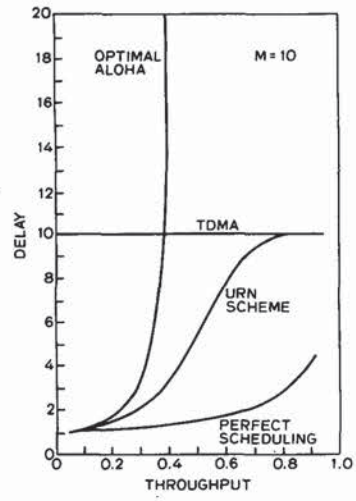


Fig. 18. Throughput-delay performance for the URN scheme (example for 10 users) [56].

furthermore grouped into frames of $m$ slots, $m \geqslant k$, each containing at least one slot for every equivalence class. Let $M$ be again the number of users. Each user is at any one time assigned to one of the $k$ equivalence classes. All stations in a given class use a random access mode to access slots assigned to their class. If CSMA is used as the contention scheme, then time slots are minislots of size $\tau$, assigned to the $k$ equivalence classes just as before. By dynamically varying the size of the frame and the assignment of slots within the frame to classes of users, one can vary the access mode to best fit the situation. At low load, for example, choosing $k = m = 1$ with all users in the same class leads to a pure random access mode of low delay. Choosing $k = m = M$ with each user constituting a separate class leads to TDMA. Increasing the parameter $k$ has the effect of decreasing the rate of collisions among users of the same class. The frame size $m$ can be used to allow a smooth changeover between the schemes. By partitioning the frame into two subframes, both contention and pure TDMA can coexist simultaneously. The information used in adapting to the situation is the collision rate and the rate of empty slots (or minislots) for the randomly accessed slots, and the rate of empty slots for the TDMA assigned slots. For example, when one minislot of a TDMA slot goes empty, the remainder of the TDMA slot may be cancelled and reassigned to some other groups (then to be used via CSMA).

Schemes other than CSMA and TDMA can be combined by this adaptive strategy. One may, for example, mix CSMA with MSRR. In [57], Ricart and Agrawala studied, via simulation, some typical adaptation algorithms of this type. Some of their simulation results for a CSMA/TDMA combination are shown in Fig. 19. These results exhibit clearly the improvement gained over the entire throughput range by using the adaptive strategy.

3) *The Reservation upon Collision Schemes (RUC) [58]*: In the Reservation upon Collision schemes, the channel time is divided into slots of fixed length which in turn are divided

Authorized licensed use limited to: Sterne Kessler Goldstein Fox. Downloaded on October 19,2022 at 20:45:59 UTC from IEEE Xplore. Restrictions apply.

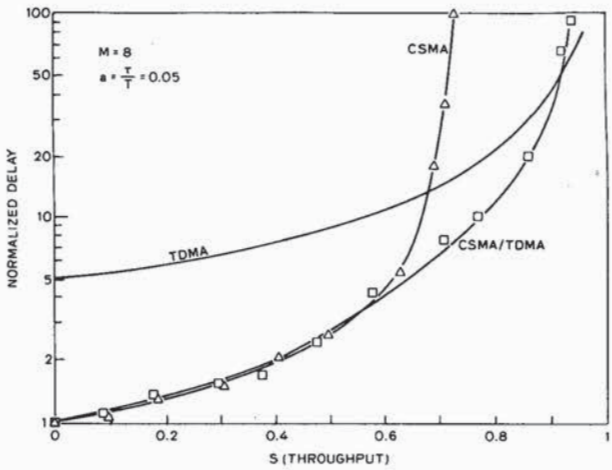


Fig. 19. Simulation results for an adaptive CSMA/TDMA strategy (eight stations, normalized propagation delay of $a = 0.05$) [57].

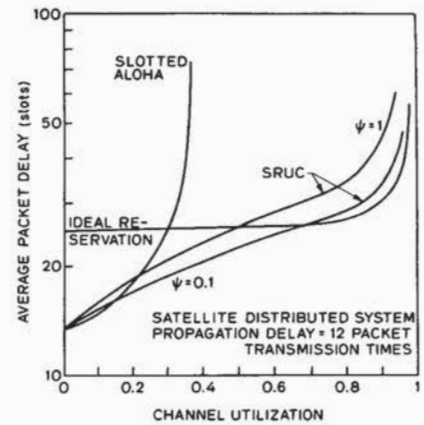


Fig. 20. Split reservation upon collision: throughput delay performance at zero overhead $\Psi = 0.1$ and $\Psi = 1$ [58].

into two parts: a data-subslot SSO for the transmission of information packets and a subslot SS1 for the transmission of (signaling) information regarding the transmitting user(s). The data subchannel can be in one of two states: the contention state or the reserved state. It is normally in the contention state and users can access the slots in a slotted ALOHA mode as long as no collisions occur. When a collision is detected, then the data subchannel switches to the reserved state and remains in that state until the queue of reservations is cleared, at which time it switches back to the contention state. That is, if a collision is detected, reservations are automatically implied for the colliding users. To accomplish this, the signaling information identifying the users must be received by all users free of interference, and thus an ingenious use of the SS1 subslots must be devised. CDMA and TDMA have been proposed in [58]. When the number of users is large, a particularly suitable approach is to consider grouping the slots into a frame of, say, $L$ slots. Each of the $L$ SS1 subslots is assigned to a group of size $M/L$ users instead of $M$ users, thus decreasing the degree of multiplexing signaling information over the SS1 subslots. TDMA or CDMA still needs to be used. In this approach, users need not transmit their identification as this is implied from the position of the SS1 subslot. However, each user has to send the number of packets transmitted in the frame, and this information requires at most $\log_2(L + 1)$ bits. This scheme is referred to as the split reservation upon collision (SRUC).

Fig. 20 shows the performance of SRUC in a satellite environment as compared to slotted ALOHA and pure reservation for two values of the overhead $\Psi$ required per frame for the signaling information. Clearly, this performance degrades as $\Psi$ increases. More detailed results can be found in [58].

Since slotted ALOHA and reservations are both suitable for satellite channels, RUC schemes are also particularly suitable for these as well as ground radio channels.

4) *Priority-Oriented Demand Assignment (PODA) [12]*: In the context of a satellite channel, PODA has been proposed as the ultimate scheme which attempts to incorporate all the properties and advantages seen in many of the previous schemes. It has provision for both implicit and explicit reservations, thus accommodating both stream and packet-type traffic. It may also integrate the use of both centralized and distributed control techniques thus achieving a high level of robustness.

Channel time is divided into two basic subframes, an information subframe and a control subframe. The information subframe contains scheduled packets and packet streams, which also contain, piggybacked, control information such as reservations and acknowledgments. The control subframe is used exclusively to send reservations that cannot be sent in the information subframe in a timely manner. In order to achieve integration of centralized and distributed assignments, the information subframe is further divided into two sections, one for each type.

Access to the control subframe (which is divided into slots accommodating fixed size control packets) can take any form that is suitable to the environment. It can be by *fixed assignment* (TDMA) if the number of stations is small (giving rise to the so-called FPODA), or by *contention* as in ALOHA if the stations have a low-duty cycle (giving rise to CPODA), or a combination of both. The boundary between the control subframe and the information subframe is not fixed, but varies with the demand placed on the channel. As in the FIFO and RR reservation schemes, distributed control is achieved by having all stations involved in this type of control keep track of their queue length information. Priority scheduling can thus be achieved. For stream traffic, a reservation is made only once, and is retained by each station in a stream queue. Centralized assignment may be used when delay is not the crucial element. This scheme has been proposed in the context of a satellite channel but may be applied to other environments as well.

5) *More on Mixed Modes*: Other studies have appeared in the literature that also deal with integrating several different access modes into the same system.

Authorized licensed use limited to: Sterne Kessler Goldstein Fox. Downloaded on October 19,2022 at 20:45:59 UTC from IEEE Xplore. Restrictions apply.

The *Mixed ALOHA Carrier Sense* (MACS) scheme consists of allowing a large user to steal, by carrier sensing, slots which are unused by a large population of small users accessing the channel in a slotted ALOHA mode [59]. Analysis has shown that the total channel utilization is significantly increased with MACS, and the throughput-delay performance of both the large user and the background ALOHA users is better with MACS than with a split-channel configuration in which the large user and the ALOHA users are each permanently assigned a portion of the channel [59].

*Group Random Access* (GRA) procedures consist of using only certain channel time-periods to allow some network terminals to transmit their information-bearing packets on a random access basis. The channel can then be utilized at other times to grant access to other terminals or other message types, by applying, as appropriate, group random access, reservation procedure or fixed assignment. The idea is simply a fixed time-division assignment among groups utilizing different access schemes. For analysis of GRA, the reader is referred to [60], [61].

Finally, we consider satellite systems with on-board processing capability. These have recently received increased attention and are being considered as a means to increase the capacity of packet satellite channels [62]-[65]. One example is typified by the integration of slotted ALOHA on several uplink channels, with TDMA on one or several downlink channels. The on-board processing capability is used to filter out all collisions and thus improve the utilization of the downlink channels. The overall spectrum efficiency is also improved especially if the ratio of uplink channels to downlink channels is properly chosen. Analysis of these disciplines is given in [62], [63]. Additional improvement over these disciplines is possible by providing buffering capability on board the satellite to smooth the input and more completely fill the downlink channels.

## III. CONCLUSION

Tremendous advances have been made in recent years in devising multiaccess schemes suitable to a variety of data communication environments. In this paper, we have briefly reviewed a large number of these protocols which we have grouped into five categories according to: 1) the degree of control exercised over the users' access; 2) the (centralized or distributed) nature of the decision-making process; and 3) the degree of adaptivity of the algorithm to the changing need. We have seen that these link level protocols have a great impact on the utilization of the communication resource in particular and the overall system performance in general. We have also briefly discussed their suitability to various traffic characteristics.

Although an attempt has been made to render the presentation complete, it is by no means exhaustive of all existing schemes, and the field is still so wide open that new schemes are constantly being introduced. Throughout the paper, an emphasis was placed on that class of packet communications that service very many bursty users, since this has been a major concern for many years. It is important, however, to note that there is a growing interest in the support of applications which lend themselves to stream-type traffic (such as packetized voice, facsimile, video data for remote conferencing, etc.) and which may also require real-time communications service on the part of the network. Moreover, with an even greater interest in integrating the many different applications onto the same network structure, it is becoming important to devise multiaccess protocols which can provide all the capabilities and features required for this integration. The adaptive strategies discussed in the paper provide an attempt at solving this problem but it is still far from being completely resolved.

Another point of great importance is the impact that these link level protocols have on the design of higher level protocols. Indeed, due to the basically different nature and behavior of some of these multiaccess schemes, one is faced with the necessity to find new ways to deal with many of the higher level functions. The routing problem in store-and-forward multiaccess/broadcast systems, for example, is significantly different from the well-known routing algorithms devised for point-to-point store-and-forward networks; here the transmitted packet should carry, at each transmission, the next node's address, and each *receiving* node has to decide as to whether to relay or ignore the packet. A discussion of routing schemes appropriate to these systems can be found in [14]. Clearly, in single-hop broadcast systems, and in local area ring architectures, the routing problem is absent.

Acknowledgment procedures may also have to be handled differently in broadcast networks. In the PRNET, for example, hop-by-hop acknowledgments can be passive, in the sense that, due to the broadcast nature of transmission, the relaying of a packet over a hop constitutes the acknowledgment for the transmission over the previous hop. Acknowledgments may also be active in the sense that an acknowledgment packet is actually created and transmitted. If acknowledgment packets are given priority, the active acknowledgment procedure has the benefit of minimizing buffering requirements at the repeaters since the acknowledgments are sent at the earliest opportunity, and possibly minimizing channel overhead since the additional transmissions beyond success resulting from delayed acknowledgments can then be kept to a minimum [67]. (In fact, it was found that if acknowledgments were instantaneous, then a few buffers in each packet radio unit appear to be sufficient to handle the storage requirements; indicating that the system becomes more channel bound than storage bound [68], [69].) In satellite environments, PODA achieves the same objective by piggybacking, whenever possible, acknowledgments on pending reservation requests, which are heard by all users, including the sender.

To conclude, we can say that despite the many advances already accomplished, this area still presents many challenging open problems, and that to best make use of the progress already achieved in link-level protocols, one also needs to turn one's attention to the many unresolved issues concerning higher level protocols.

## REFERENCES

[1] D. W. Davies, K. A. Bartlett, R. A. Scantlebury, and P. T. Wilkinson, "A digital communication network for computers giving rapid response

Authorized licensed use limited to: Sterne Kessler Goldstein Fox. Downloaded on October 19,2022 at 20:45:59 UTC from IEEE Xplore. Restrictions apply.

at remote terminals," presented at ACM Symp. Operating System Principles, Gatlinburg, TN, Oct. 1–4, 1967.

[2] W. D. Farmer, and E. E. Newhall, "An experimental distributed switching system to handle bursty computer traffic," in *Proc. ACM Conf.*, Pine Mountain, GA, Oct. 1969.

[3] M. T. Liu and C. C. Reames, "Communication protocol and network operating system design for the distributed loop computer network (DLCN)," in *Proc. 4th Annu. Symp. Computer Architecture*, Mar. 1977, pp. 193–200.

[4] M. T. Liu, "Distributed loop computer networks," in *Advances in Computer Networks*, M. Rubinoff and M. C. Yovitts, Eds. New York: Academic, 1978.

[5] R. M. Metcalfe and D. R. Boggs, "ETHERNET: Distributed packet switching for local computer networks," *Commun. Ass. Comput. Mach.*, vol. 19, pp. 395–403, 1976.

[6] L. Pouzin, "Presentation and major design aspects of the Cyclades computer network," presented at Datacom 73, ACM/IEEE, 3rd Data Commun. Symp., St. Petersburg, FL, Nov. 1973, pp. 80–87.

[7] L. G. Roberts and B. D. Wessler, "Computer network developments to achieve resource sharing," in *1970 Spring Joint Comput. Conf., Proc. AFIPS Conf.*, vol. 36, 1970, pp. 543–549.

[8] L. Pouzin, "CIGALE, The packet switching machine of the CYCLADES computer network," presented at IFIP Congress, Stockholm, Sweden, Aug. 1974, pp. 155–159.

[9] H. Opderbeck and R. B. Hovey, "Telenet—Network features and interface protocols," in *Proc. NTG-Conf. Data Networks*, Baden-Baden, West Germany, Feb. 1976.

[10] W. W. Clipshaw and F. Glave, "Datapac network review," in *Int. Comput. Commun. Conf. Proc.*, Aug. 1976, pp. 131–136.

[11] N. Abramson, "The Aloha system," in *Computer Communication Networks*, N. Abramson and F. Kuo, Eds. Englewood Cliffs, NJ: Prentice-Hall, 1973.

[12] I. M. Jacobs, R. Binder, and E. V. Hoversten, "General purpose packet satellite networks," *Proc. IEEE*, vol. 66, Nov. 1978.

[13] R. E. Kahn, "The organization of computer resources into a packet radio network," in *Nat. Comput. Conf., AFIPS Conf. Proc.*, vol. 44. Montvale, NJ: AFIPS Press, 1975, pp. 177–186; also in *IEEE Trans. Commun.*, vol. COM-25, Jan. 1977.

[14] R. E. Kahn, S. A. Gronemeyer, J. Burchfiel, and R. C. Kunzelman, "Advances in packet radio technology," *Proc. IEEE*, vol. 66, Nov. 1978.

[15] D. Clark *et al.*, "An introduction to local area networks," *Proc. IEEE*, vol. 66, Nov. 1978.

[16] W. G. Schmidt, "Satellite time-division multiple access systems: Past, present and future," *Telecommun.*, vol. 7, pp. 21–24, Aug. 1974.

[17] I. Rubin, "Message delays in FDMA and TDMA communication channels," *IEEE Trans. Commun.*, to be published.

[18] S. Lam, "Delay analysis of time-division multiple access (TDMA) channel," *IEEE Trans. Commun.*, vol. COM-25, Dec. 1977.

[19] O. Kosovych, "Fixed assignment access technique," *IEEE Trans. Commun.*, vol. COM-26, Sept. 1978.

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

[21] L. G. Roberts, "ALOHA packet system with and without slots and capture," *Comput. Commun. Rev.*, vol, 5, pp. 28–42, Apr. 1975.

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

[23] F. A. Tobagi and L. Kleinrock, "Packet switching in radio channels: Part III—Polling and (dynamic) split channel reservation multiple access," *IEEE Trans. Commun.* vol. COM-24, pp. 832–845, Aug. 1976.

[24] L. Kleinrock and F. A. 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.

[25] F. Tobagi, "Random access techniques for data transmission over packet switched radio networks," Ph.D. dissertation, Comput. Sci. Dep., School of Eng. and Appl. Sci., Univ. California, Los Angeles, Rep. UCLA-ENG 7499, Dec. 1974.

[26] F. Tobagi and V. B. Hunt, "Performance analysis of carrier sense multiple access with collision detection," in *Proc. Local Area Commun. Network Symp.*, Boston, MA, May 1979; also Stanford Electronics Lab., Comput. Syst. Lab., Tech. Rep. 173, June 30, 1979.

[27] N. Abramson, "The throughput of packet broadcasting channels," *IEEE Trans. Commun.*, vol. COM-25, pp. 117–128, Jan. 1977.

[28] 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.

[29] F. A. Tobagi, M. Gerla, R. W. Peebles, and E. G. Manning, "Modeling and measurement techniques in packet communication networks," *Proc. IEEE*, vol. 66, pp. 1423–1447, Nov. 1978.

[30] F. Tobagi and L. Kleinrock, "The effect of acknowledgment traffic on the capacity of packet-switched radio channels," *IEEE Trans. Commun.*, vol. COM-26, pp. 815–826, June 1978.

[31] S. S. Lam, "Packet switching in a multi access broadcast channel with application to satellite communication in a computer network," Ph.D. dissertation, Dep. Comput. Sci., Univ. California, Los Angeles, Mar. 1974; also in Univ. California, Los Angeles, Tech. Rep. UCLA-ENG-7429, Apr. 1974.

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

[33] F. Tobagi and L. Kleinrock, "Packet switching in radio channels: Part IV—Stability considerations and dynamic control in carrier sense multiple access," *IEEE Trans. Commun.*, vol. COM-25, pp. 1103–1120, Oct. 1977.

[34] G. Fayolle, E. Gelembe, and J. Labetoulee, "Stability and optimal control of the packet-switching broadcast channels," *J. Ass. Comput. Mach.*, vol. 24, pp. 375–386, July 1977.

[35] S. 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.

[36] J. Metzner, "On improving utilization in ALOHA networks," *IEEE Trans. Commun.*, vol. COM-24, Apr. 1976.

[37] Special Issue on Spread Spectrum Communications, *IEEE Trans. Commun.*, vol. COM-25, Aug. 1977.

[38] B. Edelson and A. Werth, "SPADE system progress and application," *COMSAT Tech. Rev.*, vol. 2, pp. 221–242, Spring 1972.

[39] W. Schmidt *et al.*, "Mat-1: INTELSAT's Experimental 700-channel TDMA/DA system," in *Proc. INTELSAT/IEEE Int. Conf. Digital Satellite Commun.*, Nov. 1969.

[40] N. Erlich, "The advanced mobile phone service," *IEEE Commun. Mag.*, vol. 17, Mar. 1979.

[41] A. G. Konheim and B. Meister, "Service in a loop system," *J. Ass. Comput. Mach.*, vol. 19, pp. 92–108, Jan. 1972.

[42] J. F. Hayes, "An adaptive technique for local distribution," *IEEE Trans. Commun.*, vol. COM-26, Aug. 1978.

[43] J. W. Mark, "Global scheduling approach to conflict-free multiaccess via a data bus," *IEEE Trans. Commun.*, vol. COM-26, Sept. 1978.

[44] L. Kleinrock, "Performance of distributed multiaccess computer communication systems," in *Proc. IFIP Congress*, 1977.

[45] W. R. Crowther, R. Rettbert, D. Walden, S. Ornstein, and F. Heart, "A system for broadcast communication: Reservation-ALOHA," in *Proc. 6th Hawaii Int. Syst. Sci. Conf.*, Jan. 1973.

[46] L. Roberts, "Dynamic allocation of satellite capacity through packet reservation," in *Proc. AFIPS Conf.*, vol. 42, June, 1973.

[47] R. Binder, "A dynamic packet switching system for satellite broadcast channels," in *Proc. ICC'75*, San Francisco, CA, June 1975.

[48] L. Kleinrock and M. Scholl, "Packet switching in radio channels: New conflict-free multiple access schemes for a small number of data users," in *ICC Conf. Proc.*, Chicago, IL, June 1977, pp. 22.1-105–22.1-111.

[49] L. W. Hansen, and M. Schwartz, "An assigned-slot listen-before-transmission protocol for a multiaccess data channel," *IEEE Trans. Commun.*, vol. COM-27, pp. 846–857, June 1979.

[50] D. C. Loomis, "Ring communication protocols," Univ. California, Dep. Inform. and Comput. Sci., Irvine, CA, Tech. Rep. 26, Jan. 1973.

[51] J. R. Pierce, "Network for block switching of data," *Bell Syst. Tech. J.*, vol. 51, pp. 1133–1143, July/Aug. 1972.

[52] A. Hopper, "Data ring at computer laboratory, University of Cambridge," *Computer Science and Technology: Local Area Networking*. Washington DC: Nat. Bur. Stand., NBS Special Publ. 500-31, Aug. 22-23, 1977, pp. 11-16.

[53] P. Zafiropoulo and E. H. Rothauser, "Signalling and frame structures in highly decentralized loop systems," *Proc. Int. Conf. on Comput. Commun.* (Washington, DC), IMB Res. Lab., Zurich, Switzerland, pp. 309–315.

[54] E. R. Hafner *et al.*, "A digital loop communication system," *IEEE Trans. Commun.*, p. 877, June 1974.

[55] M. V. Wilkes, "Communication using a digital ring," in *Proc. PACNET Conf.*, Sendai, Japan, Aug. 1975, pp. 217–255.

Authorized licensed use limited to: Sterne Kessler Goldstein Fox. Downloaded on October 19,2022 at 20:45:59 UTC from IEEE Xplore. Restrictions apply.

[56] L. Kleinrock and Y. Yemini, "An optimal adaptive scheme for multiple access broadcast communication," *ICC Conf. Proc.*, Chicago, IL, June 1977.

[57] G. Ricart and A. Agrawala, "Dynamic management of packet radio slots," presented at *Third Berkeley Workshop on Distributed Data Management and Comput. Networks*, Aug. 1978.

[58] F. Borgonovo and L. Fratta, "SRUC: A technique for packet transmission on multiple access channels," in *Proc. Int. Conf. Comput. Commun.*, Kyoto, Japan, 1978.

[59] M. Scholl and L. Kleinrock, "On a mixed mode multiple access scheme for packet-switched radio channels," *IEEE Trans Commun.*, vol. COM-27, pp. 906–911, June 1979.

[60] I. Rubin, "A group random-access procedure for multi-access communication channels," in *NTC'77 Conf. Rec. Nat. Telecommun. Conf.*, Los Angeles, CA, Dec. 1977, pp. 12:5-1–12:5-7.

[61] ——"Integrated random-access reservation schemes for multi-access communication channels," School Eng. Appl. Sci., Univ. California, Los Angeles, Tech. Rep. UCLA-ENG-7752, July 1977.

[62] J. K. DeRosa, and L. H. Ozarow, "Packet switching in a processing satellite," *Proc. IEEE*, vol. 66, pp. 100–102, Jan. 1978.

[63] R. E. Eaves, "ALOHA/TDM systems with multiple downlink capacities," *IEEE Trans. Commun.*, vol. COM-27, pp. 537–541, Mar. 1979.

[64] S. F. W. Ng and J. W. Mark, "A multiaccess model for packet switching with a satellite having some processing capability," *IEEE Trans. Commun.*, vol. COM-25, pp. 128–135, Jan. 1977.

[65] ——, "Multiaccess model for packet switching with a satellite having processing capability: Delay analysis," *IEEE Trans. Commun.*, vol. COM-26, pp. 283-290, Feb. 1978.

[66] S. S. Lam, "Satellite multi-access schemes for data traffic," in *Proc. Int. Conf. Commun.*, Chicago, IL, 1977, pp. 37.1-19–37.1-24.

[67] F. Tobagi *et al.*, "On measurement facilities in packet radio systems," in *Nat. Comput. Conf. Proc.*, New York, NY, June 1976.

[68] F. Tobagi, "Analysis of a two-hop centralized packet radio network: Part I—Slotted ALOHA," *IEEE Trans. Commun.*, vol. COM-28, pp. 196–207, Feb. 1980.

[69] ——"Analysis of a two-hop centralized packet radio network: Part II—Carrier sense multiple access," *IEEE Trans. Commun.*, vol. COM-28, pp. 208–216, Feb. 1980.

[70] W. W. Chu, "A study of asynchronous time division multiplexing for time-sharing computer systems," in *1969 Spring Joint Comput. Conf. AFIPS Conf. Proc.*, vol. 35, 1969, pp. 669–678.

[71] J. I. Capetanakis, "Tree algorithms for packet broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 505–515, Sept. 1979.

[72] I. Chlamtac *et al.*, "BRAM: The broadcast recognizing access method," *IEEE Trans. Commun.*, vol. COM-27, pp. 1183–1190, Aug. 1979.

★

**Fouad A. Tobagi** (M'77), for a photograph and biography, see page 207 of the February 1980 issue of this TRANSACTIONS.

Authorized licensed use limited to: Sterne Kessler Goldstein Fox. Downloaded on October 19,2022 at 20:45:59 UTC from IEEE Xplore. Restrictions apply.