processing, intersymbol interference, data compression, and cryptography. He has consulted in the areas of communications, information theory, and general engineering for various industries and laboratories. He is the liaison at Stanford for its Industrial Affiliates Program in Information Systems, a program designed to increase interaction between industry and the university.

Dr. Hellman is a member of Tau Beta Pi and Eta Kappa Nu. He is President of the San Francisco Section's IEEE Information Theory Group's Chapter, and was Publications Chairman for the 1972 Information Theory Symposium. He is an Associate Editor for Communication Theory of the IEEE TRANSACTIONS ON COMMUNICATIONS.

Packet Switching in a Multiaccess Broadcast Channel: Performance Evaluation

LEONARD KLEINROCK, FELLOW, IEEE, AND SIMON S. LAM, MEMBER, IEEE

Abstract—**In this paper, the rationale and some advantages for multiaccess broadcast packet communication using satellite and ground radio channels are discussed. A mathematical model is formulated for a "slotted ALOHA" random access system. Using this model, a theory is put forth which gives a coherent qualitative interpretation of the system stability behavior which leads to the definition of a stability measure. Quantitative estimates for the relative instability of unstable channels are obtained. Numerical results are shown illustrating the trading relations among channel stability, throughput, and delay. These results provide tools for the performance evaluation and design of an uncontrolled slotted ALOHA system. Adaptive channel control schemes are studied in a companion paper.**

INTRODUCTION

IN THIS and a forthcoming paper [1], a packet switching technique based upon the random access concept of the ALOHA System [2] will be studied in detail. This technique, referred to as slotted ALOHA random access, enables efficient sharing of a data communication channel by a large population of users, each with a bursty data stream. This packet switching technique may be applied to the use of satellite and ground radio channels for computer–computer and terminal–computer communications, respectively [3]–[10]. The multiaccess broadcast capabilities of these channels render them attractive solutions to two problems: 1) large computer–communication networks with nodes distributed over wide geographic areas, and 2) large terminal access networks with potentially mobile terminals.

The objective of this study is to develop analytic models and methods for the evaluation and optimization of the channel performance of a slotted ALOHA system. The problem of performance evaluation is addressed in this paper. In [1], we present dynamic channel control procedures as solutions to some of the issues considered herein.

In this paper, the rationale for multiaccess broadcast packet communication is first discussed. The mathematical model to be considered is then described. Following that, a theory is proposed which explains the dynamic and stochastic channel behavior. In particular, we display the delay-throughput performance curves obtained under the assumption of equilibrium conditions [6]. We then demonstrate that a slotted ALOHA channel often exhibits "unstable behavior." A stability definition is proposed which characterizes stable and unstable channels. A stability measure (FET) is then defined which quantifies the relative instability of unstable channels. An algorithm is given for the calculation of FET. Finally, numerical results are shown which illustrate the trading relations among channel stability, channel throughput, and average packet delay. Our main concern in this paper is the consideration of the stability issue and its effect on the channel throughput-delay performance.

MULTIACCESS BROADCAST PACKET COMMUNICATION

Rationale

For almost a century, circuit switching dominated the design of communication networks. Only with the higher speed and lower cost of modern computers did packet communication become competitive. It was not until approximately 1970 that the computer (switching) cost dropped below the communication (bandwidth) cost in a packet switching network [11]. This also marked the first appearance of packet switched computer–communication networks [2], [12].

Circuit switching is relatively inefficient for computer

Paper approved by the Associate Editor for Computer Communication of the IEEE Communications Society for publication after presentation at the 7th Hawaii International Conference on System Sciences, Honolulu, Hawaii, January 8–10, 1974. Manuscript received June 30, 1974; revised September 30, 1974. This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract DAHC 15-73-C-0368.

L. Kleinrock is with the Department of Computer Science, University of California, Los Angeles, Calif. 90024.

S. S. Lam is with the IBM Thomas J. Watson Research Center, Yorktown Heights, N. Y. 10598.

communications, especially over long distances. Measurement studies [13] conducted on time-sharing systems indicate that both computer and terminal data streams are *bursty.* Depending on the channel speed, the ratio between the peak and the average data rates may be as high as 2000 to 1 [5]. Consequently, if a high-speed point-to-point channel is used, the channel utilization may be extremely low since the channel is idle most of the time. On the other hand, if a low-speed channel is used, the transmission delay is large.

The above dilemma is caused by channel users imposing bursty random demands on their communication channels. By the law of large numbers in probability theory, the total demand at any instant from a large population of independent users is, with high probability, approximately equal to the sum of their average demands (i.e., a nearly deterministic quantity). Thus, if a channel is dynamically shared in some fashion among many users, the required channel bandwidth to satisfy a given delay constraint may be much less than if the users are given dedicated channels. This concept is known as *statistical load averaging* and has been applied in many computer–communication schemes to various degrees of success. These schemes include: polling systems [14], loop systems [15], asynchronous time division multiplexing (ATDM) [16], and the store-and-forward packet switching concepts [17]–[19] implemented in the ARPA network [12].

We are currently facing an enormous growth in computer networks [20]. To design cost-effective computer–communication networks for the future, new techniques are needed which are capable of providing efficient high-speed computer–computer and terminal–computer communications in a large network environment. The application of packet switching techniques to radio communication (both satellite and ground radio channels) appears to provide a solution.

Radio is a multiaccess broadcast medium. That is, a signal generated by a radio transmitter may be received over a wide area by any number of receivers. This is referred to as the *broadcast* capability. Furthermore, any number of users may transmit signals over the same channel. This is referred to as the *multiaccess* capability. (However, if two signals at the same carrier frequency overlap in time at a radio receiver[1], we assume that neither is received correctly. This destructive interference is the key issue in studying the multiaccess radio channel used in a packet switching mode.) Thus, a single ground radio channel provides a completely connected network topology for a large number of nodes within range of each other. Similarly, a satellite transponder in a geostationary orbit above the earth acts as a radio repeater. Any number of earth stations may transmit signals up to the satellite at one carrier frequency (the multiaccess channel). Any signal received by the satellite transponder is beamed back to earth at another frequency (the broadcast channel). This broadcasted signal may be received by all earth stations covered by the transponder beam. Thus, a satellite channel (consisting of both carrier frequencies) provides a completely connected network topology for all earth stations covered by the transponder beam.

Consider the use of packet communication in a computer–communication network environment to support large populatons of (bursty) users over a wide area. We can then identify and summarize the following advantages of satellite and ground radio channels over conventional wire communications.

1) Elimination of Complex Topological Design and Routing Problems: Topological design and routing problems are very complex in networks with a large population of users. Existing implementations suitable for a (say) 50 node network may become totally inappropriate for a 500 node network required to perform the same functions [21]. On the other hand, ground radio and satellite channels used in the multiaccess broadcast mode provide a completely connected network topology, since every user may access any other user covered by the broadcast.

2) Wide Geographical Areas: Wire communications become expensive over long distances (e.g., transcontinental, transoceanic). Even on a local level, the communication cost for an interactive user on an alphanumeric console over distances of over 100 miles may easily exceed the cost of computation [2]. On the other hand, satellite and radio communications are relatively distance independent, and are especially suitable for geographically scattered users.

3) Mobility of Users: Since radio is a multiaccess broadcast medium, it is possible for users to move around freely. This consideration will soon become important in the development of personal terminals in future telecommunication systems [22] as well as in aeronautical and maritime applications [23].

4) Large Population of Active and Inactive Users: In wire communications, the system overhead usually increases with the number of users (e.g., polling schemes). The maximum number of users is often bounded by some hardware limitation (e.g., the fan-in of a communications processor). In radio communication, since each user is merely represented by an ID number, the number of active users is bounded only by the channel capacity and there is no limitation to the number of inactive (but potentially active) users beyond that of a finite address space.

5) Flexibility in System Design: A radio packet communication system can become operational with two or three users. The size of the user population can be increased up to the channel capacity. More users can be accommodated by increasing the radio channel bandwidth. In other words, the communication system can be expanded or contracted without major changes in the basic system design and operational schemes.

6) Statistical Load Averaging: Wire communication links are more efficiently utilized in a store-and-forward packet switched network than in a circuit switched network. However, at any instant, there may be unused channel capacity in some parts while congestion exists in

[1] This event will be referred to as a *channel collision.*

other parts of the network. The application of packet switching techniques to a single high-speed satellite or radio channel permits the total demand of all user input sources to be statistically averaged at the channel. Note also that each user transmits data at the wide-band channel rate.

7) *Multiaccess Broadcast Capability:* This capability in radio communication may be useful for certain multipoint-to-multipoint communication applications.

The Multiaccess Channel Model

Consider a radio communication system such as a packet switched satellite system [5]-[10] or the ALOHA System [2]. In each case, there is a *broadcast* channel for point-to-multipoint communication and a *multiaccess* channel shared by a large number of users. Since the broadcast channel is used by a single transmitter, no transmission conflict will arise. All nodes covered by the radio broadcast can receive on the same frequency, picking out packets addressed to themselves and discarding packets addressed to others.

The problem we are faced with is how to effect time-sharing of the multiaccess channel among all users in a fashion which produces an acceptable level of performance. As soon as we introduce the notion of sharing in a packet switching mode, we must be prepared to resolve conflicts which arise when simultaneous demands are placed upon the channel. There are two obvious solutions to this problem: the first is to form a queue of conflicting demands and serve them in some order; the second is to "lose" any demands which are made while the channel is in use. The former approach is taken in ATDM and in store-and-forward networks assuming that storage may be provided economically at the point of conflict. The latter approach is adopted in the ALOHA System random access scheme; in this system, in fact, *all* simultaneous demands made on the radio channel are lost.

Let us define *channel throughput rate* S_{out} to be the average number of correctly received packet transmissions per packet transmission time (assuming stationary conditions). We also define *channel capacity* $S_{\max}$ to be the maximum possible channel throughput rate. The channel capacity of a pure ALOHA multiaccess channel was shown by Abramson to be $1/2e \simeq 18$ percent for a fixed packet size [2]. Under similar assumptions, Gaarder showed that a pure ALOHA channel with a fixed packet size is always superior (in terms of channel capacity) to one with different packet sizes [24].

Roberts suggested that the channel may be slotted by requiring all users to synchronize[2] the leading edges of their packet transmissions to coincide with an imaginary time slot boundary at the multiaccessed radio receiver [25]. The duration of a channel time slot is chosen to be equal to a packet transmission time. The resulting scheme will be referred to as "slotted ALOHA random access" or "slotted ALOHA." In this scheme, the users transmit newly generated packets into channel time slots independently. In the event of a channel collision, the collided packets are retransmitted after *random* retransmission delays. (See Fig. 1.) The channel capacity of a slotted ALOHA channel was shown to be $1/e \simeq 36$ percent [25].

To achieve a channel throughput rate larger than the 36 percent limitation, various other multiaccess broadcast packet swiching schemes have been proposed to take advantage of special system and traffic characteristics. The reader is referred to the references [3], [7], [26] for description of these schemes.

Consider a slotted ALOHA channel. The *channel input* in a time slot is defined to be a random variable representing the total number of *new* packets transmitted by all users in that time slot. Assuming stationary conditions, the channel input rate S is the average number of new packet transmissions per time slot. The *channel traffic* in a time slot is defined to be a random variable representing the total number of packet transmissions (both *new and previously collided* packets) by all users in that time slot. Assuming stationary conditions, the channel traffic rate G is the average number of packet transmissions per time slot. The *channel throughput* (or output) in a time slot is defined to be a random variable representing the number (0 or 1) of successful packet transmissions in that time slot. Assuming stationary conditions, the channel throughput (output) rate S_{out} is the probability of exactly one packet transmission in a channel time slot.

The retransmission delay (RD) incurred by an unsuccessful packet transmission may be regarded as the sum of a deterministic component (R) and a random component. The random component is necessary since if collided packets are retransmitted after the same deterministic delay, they will collide again for sure. In a ground radio system, RD corresponds to the positive acknowledgment time-out interval [2]. In a satellite system, since each channel user listens to the satellite broadcast, one round-trip propagation time after transmitting a packet he knows whether he was successful or if a channel collision occurred. In this case, the deterministic component corresponds to a round-trip satellite propagation delay. We shall assume a noise-free channel such that a packet is received incorrectly if and only if it suffered a channel collision. In [6], a uniform probability distribution is assumed for the random component of RD such that each user retransmits a previously collided packet at random during one of the next K slots (each such slot being chosen with probability $1/K$). Thus, retransmission will take place either $R + 1$, $R + 2, \cdots$ or $R + K$ slots after the previous transmission. This is said to be the uniform retransmission randomization scheme. Under this scheme, equilibrium throughput-delay tradeoffs have been obtained for a slotted ALOHA channel with a Poisson input source (the infinite population model). Such throughput-delay contours are shown here in Fig. 2 for different values of K. Note that the minimum envelope of these contours defines the optimum channel perform-

[2] The problem of synchronizing channel users is a nontrivial one. It will not be addressed in this paper.

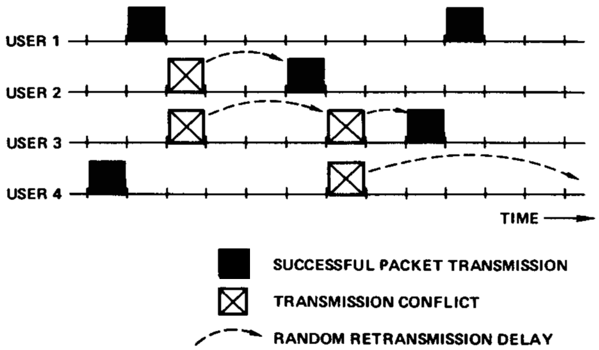


Fig. 1. Slotted ALOHA random access.

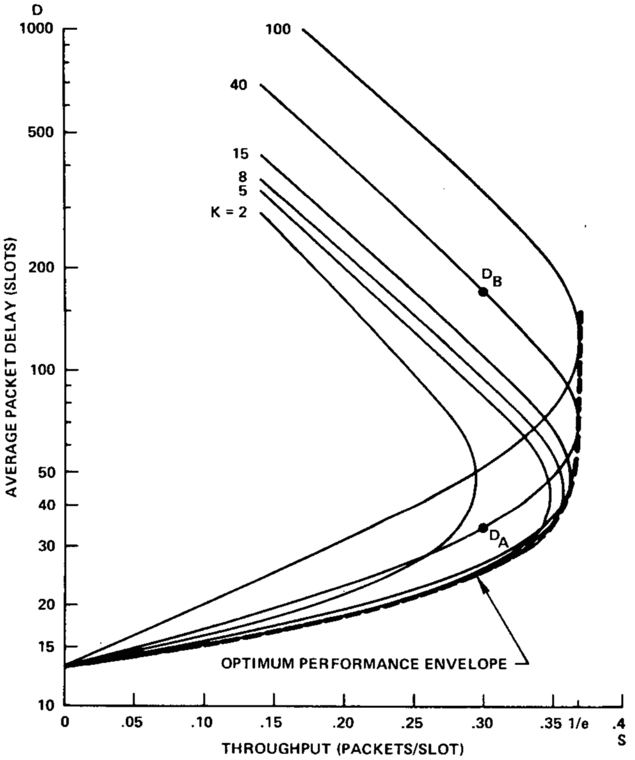


Fig. 2. Equilibrium throughput-delay tradeoff.

ance. These results correspond to the use of a 50 KPBS satellite channel, 1125 bits per packet, and a satellite round-trip propagation delay of 0.27 s for all users. Thus R is equal to 12 slots and there are 44.4 slots in one second. (These numbers will be assumed throughout this paper.) In Fig. 2, D represents the average packet delay in slots. Note that the channel input rate S is equal to the channel throughput rate S_{out} under the assumption of channel equilibrium. The channel capacity $S_{\max}$ approaches $1/e$ in the limit as $K \rightarrow \infty$. For $K = 15$, it is almost there. For values of K between 8 and 15, the equilibrium throughput-delay tradeoffs are very close to the optimum performance envelope over a wide range of S.

The analytic results presented so far are based upon the assumption that the channel is in equilibrium. Referring to Fig. 2, we see that given S and K (say $K = 40$), there are two possible equilibrium solutions for D! They correspond to a small delay value D_A and a much larger delay value D_B. (We shall refer to the equilibrium point given by S, K, and D_A as the *channel operating point*, since this is the desired channel performance given S and K.) This observation suggests that the assumption of equilibrium conditions adopted in most previous analytic models [4]–[7] may not be valid.

In order to study the dynamic behavior of these channels, simulations were performed for the infinite population model [10]. Each simulation run was observed to behave in the following manner. Starting from an initially empty system, the channel stays in equilibrium at the channel operating point for a finite period of time until stochastic fluctuations give rise to some high channel traffic rate which reduces the channel throughput rate which in turn further increases the channel traffic rate. As this vicious cycle continues, the channel becomes inundated with collisions and retransmissions. At the same time, the channel throughput rate vanishes rapidly to zero. This phenomenon will be referred to as *channel saturation.* Thus, we realize that the equilibrium throughput-delay tradeoffs are not sufficient to characterize the performance of the infinite population model. A more accurate measure of channel performance must reflect the trading relations among channel stability, throughput and delay. A mathematical model with a simpler structure than that used in [6] will be defined below. This model is similar to the one studied by Metcalfe [4]. Using this model, the concepts of channel saturation and stability in a slotted ALOHA random access channel have been characterized [8], [10].

STABILITY-THROUGHPUT-DELAY TRADEOFF PERFORMANCE

In this section, a Markovian model is first formulated for a population of M channel users. The variable M is assumed to be large and may be either finite or infinite. A theory is then proposed which characterizes the instability phenomenon in the following ways.

1) Stable and unstable channels are defined.

2) In a stable channel, equilibrium throughput-delay results (as shown in Fig. 2) are achievable over an infinite time horizon. In an unstable channel, such channel performance is achievable only for some finite time period before the channel goes into saturation.

3) For unstable channels, a stability measure is defined and an efficient computational procedure for its calculation is given.

4) Using the above stability measure, the stability-throughput-delay tradeoff for unstable channels is examined.

The Markovian Model

We consider a slotted ALOHA channel with a user population consisting of M users. Each such user can be in one of two states: *blocked* or *thinking.* In the thinking state, a user generates and transmits a new packet in a time slot with probability σ. A packet which had a channel collision and is waiting for retransmission is said to be *backlogged.* The retransmission delay RD of each backlogged packet is assumed to be geometrically distributed, i.e., each backlogged packet retransmits in the current time slot with probability p. Assuming bursty users, we must have $p \gg \sigma$. From the time a user generates a packet until that packet is successfully received, the user is blocked in the sense that he cannot generate (or accept from his input source) a new packet for transmission.

Let N^t be a random variable (called the *channel backlog*) representing the total number of backlogged packets at time t. The channel input rate at time t is $S^t = (M - N^t)\sigma$. Note that S^t decreases linearly as N^t increases. The vector (N^t, S^t) will be denoted as the *channel state vector*. In this context, both M and σ may be functions of time. We shall assume M and σ to be time-invariant unless stated otherwise. In this case, N^t is a Markov process (chain) with stationary transition probabilities and serves as the state description for the system. The discrete *state space* will now consist of the set of integers $\{0,1,2,\cdots,M\}$. The *one-step state transition probabilities* of N^t are, for $i = 0,1,2,\cdots,M$,

$$p_{ij} = \text{Prob}\,[N^{t+1} = j \mid N^t = i] = \begin{cases} 0 & j \le i-2 \\ ip(1-p)^{i-1}(1-\sigma)^{M-i} & j = i-1 \\ (1-p)^i(M-i)\sigma(1-\sigma)^{M-i-1} + [1 - ip(1-p)^{i-1}](1-\sigma)^{M-i} & j = i \\ (M-i)\sigma(1-\sigma)^{M-i-1}[1-(1-p)^i] & j = i+1 \\ \begin{pmatrix} M-i \\ j-i \end{pmatrix} \sigma^{j-i}(1-\sigma)^{M-j} & j \ge i+2. \end{cases} \tag{1}$$

For the infinite population model in which $M \to \infty$ and $\sigma \to 0$ such that $M\sigma = S$ which is constant and finite, the above equation becomes

$$p_{ij} = \begin{cases} 0 & j \le i-2 \\ ip(1-p)^{i-1}\exp(-S) & j = i-1 \\ (1-p)^iS\exp(-S) + [1 - ip(1-p)^{i-1}]\exp(-S) & j = i \\ S\exp(-S)[1-(1-p)^i] & j = i+1 \\ \dfrac{S^{j-i}}{(j-i)!}\exp(-S) & j \ge i+2. \end{cases} \tag{2}$$

The assumption that RD has a memoryless geometric distribution permits a simple state description for the mathematical model. However, this assumption implies that RD has a zero deterministic component ($R = 0$). In a satellite channel this obviously represents an approximation. (However, it may be physically realizable in radio communications over short distances in which channel propagation delays are negligible compared to a packet transmission time.) A (geostationary) satellite channel has a round-trip propagation delay of 0.27 s, which necessitates a state description consisting of the channel history for at least R consecutive time slots. The difficulty in mathematical analysis using such a state description was illustrated in [10]. However, simulation results have shown that the slotted ALOHA channel performance (in terms of average throughput and delay) is dependent primarily upon the *average* retransmission delay ($\overline{\text{RD}}$) and quite insensitive to the exact probability distributions considered [10]. In order to use the analytic results of the Markovian model here to predict the throughput-delay performance of a slotted ALOHA channel with nonzero R, it is necessary to use a value of p in the Markovian model which gives the same $\overline{\text{RD}}$. For example, to approximate a slotted ALOHA channel with uniform retransmission randomization, we must let

$$p = \frac{1}{R + (K+1)/2} \tag{3}$$

such that $\overline{\text{RD}} = R + (K+1)/2$ in both cases.

We define the length of time for which a packet is backlogged to be the backlog time of the packet and denote the *average backlog time* by D_b. To obtain the average packet delay (as defined in [6]), we must add to D_b, $R + 1$ time slots, which represent the delay incurred by each successful transmission. Thus, we have

$$D = D_b + R + 1. \tag{4}$$

Numerical results in this paper will be expressed in terms of K (rather than p) through use of (3) and (4) for comparison with previous results for channel performance [6].

The Theory

Conditioning on $N^t = n$, the expected channel throughput $S_{\text{out}}(n,\sigma)$ is the probability of exactly one packet transmission in the tth time slot. Thus,

$$S_{\text{out}}(n,\sigma) = (1-p)^n(M-n)\sigma(1-\sigma)^{M-n-1} + np(1-p)^{n-1}(1-\sigma)^{M-n}. \quad (5)$$

For the infinite population model, i.e., in the limit as $M \uparrow \infty$ and $\sigma \downarrow 0$ such that $M\sigma = S$ is finite and the channel input is Poisson distributed at the constant rate S, the above equation reduces to

$$S_{\text{out}}(n,S) = (1-p)^n S \exp(-S) + np(1-p)^{n-1} \exp(-S). \quad (6)$$

This expression is very accurate even for finite M if $\sigma \ll 1$ and if we replace $S = M\sigma$ by $S = (M-n)\sigma$. We assume that the condition $\sigma \ll 1$ (which implies bursty users) is always satisfied in problems of interest to us.

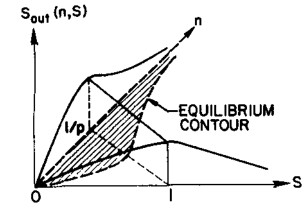


Fig. 3. Throughput surface above the (n,S) plane.

In Fig. 3, for a fixed K we sketch $S_{\text{out}}(n,S)$ as a three-dimensional surface above the (n,S) plane. Note that there is an *equilibrium contour* in the (n,S) plane defined as the locus of points on which the channel input rate S is equal to the expected channel throughput $S_{\text{out}}(n,S)$ given by (6). In the crosshatched region enclosed by the equilibrium contour, $S_{\text{out}}(n,S)$ exceeds S; elsewhere, S is greater than $S_{\text{out}}(n,S)$. In Fig. 4, a family of equilibrium contours for various K are displayed. We see that if we increase the average retransmission delay (by increasing K or equivalently decreasing p), the equilibrium contour moves upwards. We show below that these equilibrium contours play a crucial role in determining the stability behavior of the channel.

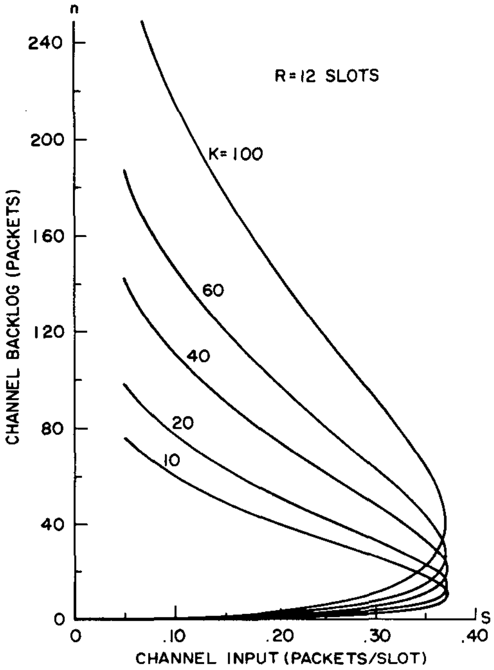


Fig. 4. Equilibrium contours in the (n,S) plane.

Given an equilibrium contour in the (n,S) plane, we first consider the dynamic behavior of the channel subject to *time-varying* inputs using a *fluid approximation* interpretation. The following example serves to illustrate the underlying concepts.

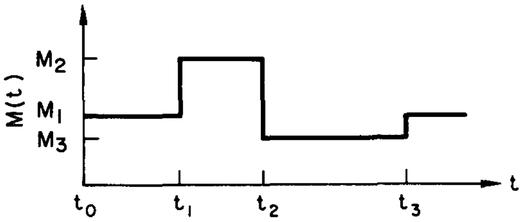


Fig. 5. $M(t)$.

Consider the case in which σ is constant while $M = M(t)$ is a function of time as shown in Fig. 5. We use the fluid approximation for the trajectory of the channel state vector (N^t,S^t) in the (n,S) plane as sketched in Fig. 6. Recall that $S^t = (M - N^t)\sigma$. The arrows indicate the "fluid" flow direction which depends on the relative magnitudes of the instantaneous channel throughput rate $S_{\text{out}}(n,S)$ and the channel input rate S. Two possible cases are shown corresponding to different values of the amplitude M_3, of the input pulse in Fig. 5. The solid line (Case 1) represents a trajectory which returns to the original state on the equilibrium contour despite the input pulse. The dashed line (Case 2) represents a less fortunate situation in which the decrease in the channel input rate at time t_2 is not sufficient to bring the trajectory back into the "safe" region (shown shaded) in which $S < S_{\text{out}}(n,S)$; eventually, the channel "fails" as a result of an increasing backlog and a vanishing channel throughput.

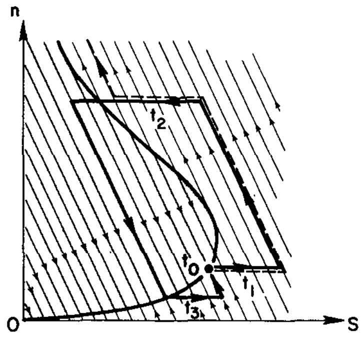


Fig. 6. Fluid approximation trajectories.

The above example demonstrates channel saturation due to a time-varying input. Let us now study the conditions under which the slotted ALOHA channel with a *stationary* input (constant M and σ) can go into saturation as a result of statistical fluctuations.

Assume that M and σ are constant. The trajectory of (N^t,S^t) is constrained to lie on the straight line $S = (M - n)\sigma$ called the *channel load line* which intercepts the n-axis at $n = M$ and has a slope equal to $-1/\sigma$. We now propose the following definition for characterizing stable and unstable channels.

The Stability Definition: A slotted ALOHA channel is said to be *stable* if its load line intersects (nontangentially) the equilibrium contour in exactly one place. Otherwise, the channel is said to be *unstable.*

Examples of stable and unstable channels are shown in Fig. 7. Arrows on the channel load lines indicate directions of fluid flow given by the fluid approximation. In other words, the arrows point in the direction of increasing backlog size if $S > S_{\text{out}}(n,S)$ and in the direction of decreasing backlog size if $S_{\text{out}}(n,S) > S$.

Each channel load line may have one or more equilibrium points. A point on the load line is said to be a *stable equilibrium point* if it acts as a "sink" with respect to fluid flow. It is a *globally stable equilibrium point* if it is the only stable equilibrium point on the channel load line. Otherwise, it is a *locally stable equilibrium point.* (Each stable equilibrium point is identified by a dot on channel load lines in Fig. 7 except in Fig. 7(c), where one of the stable equilibrium points is at $n = \infty$.) An equilibrium point is said to be an *unstable equilibrium point* if fluid flow emanates from it. Thus, the channel state N^t sitting on such a point will drift away from it given the slightest perturbation. The stability definition given above is equivalent to defining a stable channel to be one whose channel load line has a globally stable equilibrium point.

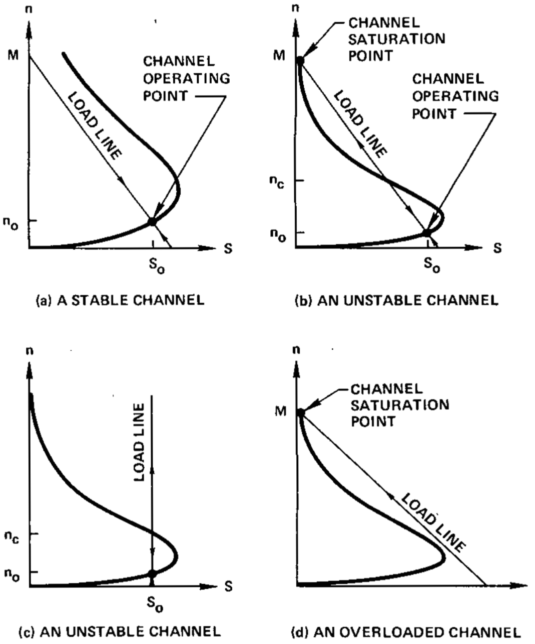
In Fig. 7(a), we show the channel load line of a stable channel. The globally stable equilibrium point on the load line, (n_o,S_o), will be referred to as the *channel operating point.* If M is finite, a stable channel can always be achieved by using a sufficiently large K (see Fig. 4). Of course, a large K implies that the equilibrium backlog size n_o is large; the corresponding average packet delay may be too large to be acceptable. Since the Markov chain N^t has a finite state space and is irreducible (assuming $p,\sigma > 0$), a stationary probability distribution always exists [27], [28]. The stationary probability distribution $\{P_n\}_{n=0}^M$ of N^t can be computed by solving the following set of linear simultaneous equations

$$P_j = \sum_{i=0}^{M} P_i p_{ij} \qquad j = 0,1,\cdots,M$$

and

$$\sum_{i=0}^{M} P_i = 1$$

where the state transition probabilities p_{ij} are given by (1). The steady-state channel throughput rate S_{out} and expected channel backlog $\bar{N}$ can then be obtained from



(a) A STABLE CHANNEL (b) AN UNSTABLE CHANNEL

(c) AN UNSTABLE CHANNEL (d) AN OVERLOADED CHANNEL

Fig. 7. Stable and unstable channels.

$$S_{\text{out}} = \sum_{n=0}^{M} S_{\text{out}}(n,\sigma) P_n \tag{7}$$

and

$$\bar{N} = \sum_{n=1}^{M} nP_n. \tag{8}$$

Numerical results have shown that these values of S_{out} and $\bar{N}$ for a stable channel are closely approximated by the equilibrium S_o and n_o at the channel operating point, and also by the equilibrium throughput-delay values in Fig. 2 for the infinite population model. For example, suppose $K = 60$, $M = 200$, and $1/\sigma = 536.1$; the equilibrium channel throughput rate at the channel operating point is $S_o = 0.346$. In Fig. 9 below (to be described later), we see that the steady-state channel throughput rate computed by using (7) is $S_{\text{out}} = 0.344$. For the same example, $\bar{N}$ is calculated to be 15.4 slots. By Little's result [27], the average backlog time is

$$D_b = \frac{\bar{N}}{S_{\text{out}}} = \frac{15.4}{0.344} = 44.8 \text{ slots.}$$

Applying (4), we get $D = 44.8 + 13 = 57.8$ slots. Now given $S_o = 0.346$, the $K = 60$ equilibrium throughput-delay contour for the infinite population model [6] gives $D = 56.5$ slots.

In Fig. 7(b), we show the channel load line of an unstable channel. The point (n_o,S_o) is again the desired channel operating point since it yields the larger channel throughput and smaller average packet delay between the two locally stable equilibrium points on the load line. In fact, the other locally stable equilibrium point, having a huge backlog and virtually zero throughput, corresponds to the channel saturation state; it will be referred to as the *channel saturation point.* Although it has a stationary probability distribution, N^t will "flip-flop" between the two locally stable equilibrium points in the following manner. Starting from an empty channel ($N^0 = 0$) quasi-stationary conditions will prevail at the operating point (n_o,S_o). The channel, however, cannot maintain equilibrium at this point indefinitely since N^t is a random process; that is, with probability one, the channel backlog N^t crosses the unstable equilibrium point n_c in a finite time, and as soon as it does, the channel input rate S exceeds $S_{\text{out}}(n,S)$. Under this condition, N^t will drift toward the saturation point. Although there is a nonzero probability that N^t may return below n_c, all our simulations show that the channel state N^t accelerates up the channel load line producing an increasing backlog and a vanishing throughput rate. Since the saturation point is a locally stable equilibrium point, quasi-stationary conditions will prevail there for some finite (but probably very long) time period. In this state, the communication channel can be regarded as having failed. (In a practical system, external control should be applied at this point to restore proper channel operation.) Thus, the two locally stable equilibrium points on the load line of an unstable channel correspond to the channel being "up" or "down". An unstable channel may be acceptable if the average channel up time is large and external control is available to bring the channel back up whenever it goes down.

In Figs. 8 and 9, we see how, as the number of channel users M increases, an originally stable channel becomes unstable although the channel input rate S_o at the operating point remains constant (by reducing σ). (These results are obtained by first solving for the stationary probability distribution of N^t and then applying (7) and (8).) For $S_o = 0.36$ and $K = 10$, we see that as M exceeds 80, the stationary channel throughput rate decreases and the average packet delay increases very rapidly with M. Using the $K = 10$ equilibrium contour in Fig. 4, the maximum value of M that is possible without making the channel load line intersect the equilibrium contour more than once is determined (graphically) to be $M_{\max} = 79$, which exactly gives the knees of the curves in Fig. 8. This excellent agreement provides the motivation for the stability definition proposed above. In Fig. 9, by using a larger value of K ($=60$), a larger $M_{\max}$ is possible. Note, however, that the average packet delay ($\simeq 56$ slots) for $K = 60$ is much larger than the average packet delay ($\simeq 36$ slots) for $K = 10$.

Given K and S_o, $M_{\max}$ can be obtained graphically from the equilibrium contours such as shown in Fig. 4. In Fig. 10 we show $M_{\max}$ as a function of K with S_o fixed at the maximum possible value given K. Note the linear relationship between $M_{\max}$ and K for the values shown. In Fig. 11, we illustrate how an originally unstable channel can be rendered stable by using a sufficiently large K.

In Fig. 7(c), we show the channel load line of an infinite population model. This is an unstable channel since

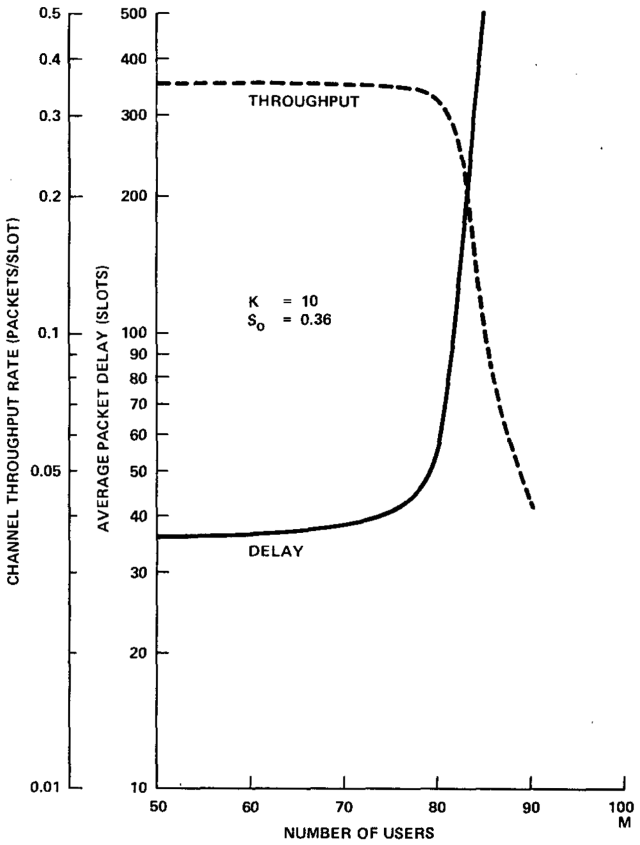


Fig. 8. Channel performance versus M at $K = 10$ and $S_o = 0.36$.

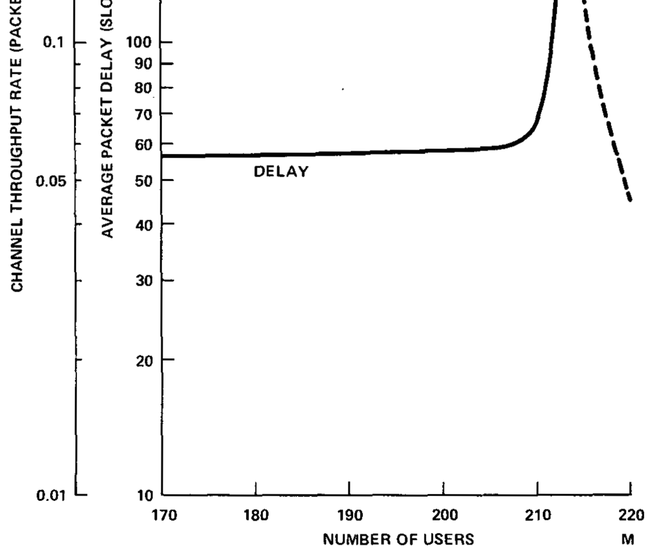


Fig. 9. Channel performance versus M at $K = 60$ and $S_o = 0.346$.

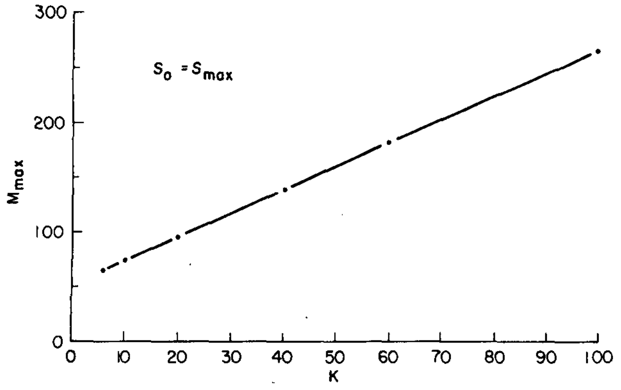


Fig. 10. $M_{\max}$ versus K.

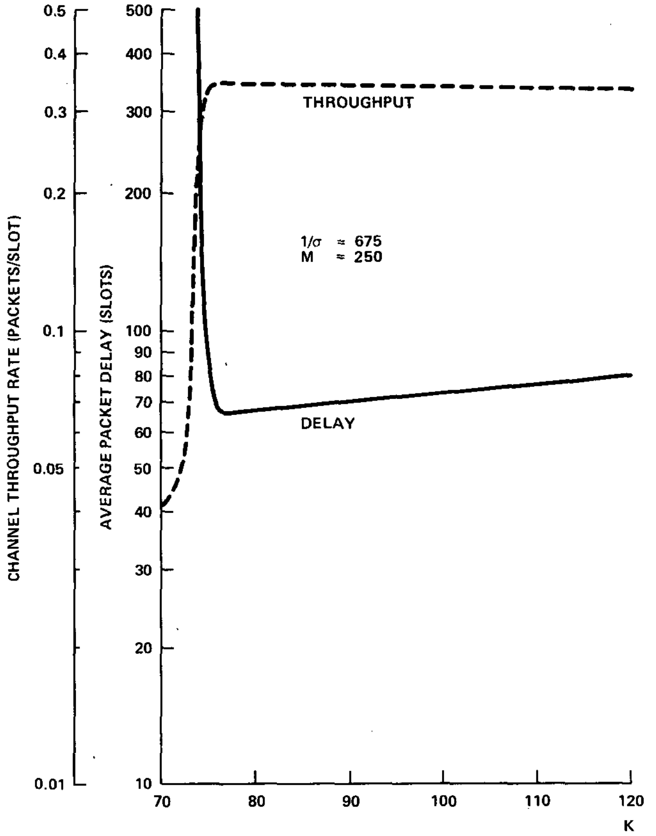


Fig. 11. Channel performance versus K at $M = 250$ and $1/\sigma = 675$.

$n = \infty$ is a stable equilibrium point. In fact, since N^t has an infinite state space and $S > S_{\text{out}}(n,S)$ for $n > n_c$, a stationary probability distribution does not exist for N^t. (See, for example, [29, pp. 543–546] for such a proof in a queueing context.)

The channel load line shown in Fig. 7(d) is stable according to the stability definition. However, the globally stable equilibrium point in this case is the channel saturation point! Thus, this represents an "overloaded" channel as a result of bad system design. To correct this situation, the number of active users M supported by the channel should be reduced. From now on, a stable channel will always refer to the load line depicted in Fig. 7(a) instead of Fig. 7(d).

Let us summarize the major *conclusions* in the above discussion.

1) The steady-state throughput-delay performance of a stable channel is closely approximated by its globally stable equilibrium point and by the equilibrium throughput-delay results for the infinite population model.

2) In an unstable channel, the throughput-delay performance at a locally stable equilibrium point can be achieved only for some finite time period.

A Stability Measure

From the above discussion and referring to Fig. 7(b)' the load line of an unstable channel can be partitioned into two regions. The *safe* region consisting of the channel states $\{0,1,2,\cdots,n_c\}$ and the *unsafe* region consisting of the channel states $\{n_c+1,\cdots,M\}$. A good stability measure (for these unstable channels!) is the average time to exit into the unsafe region starting from a safe channel state. To be exact, we define FET to be the *average first exit time* into the unsafe region starting from an initially empty channel ($N^0 = 0$). Thus, FET gives an approximate measure of the average up time of an unstable channel. Below we derive the probability distributions and expected values of such first exit times. The derivations are based upon well-known results of first entrance times in the theory of Markov chains with stationary transition probabilities [28], [30].

Consider the Markovian model with constant M and σ, where M may be infinite. N^t is a Markov process (chain) with stationary transition probabilities $\{p_{ij}\}$ given by (1) or (2). Define the random variable T_{ij} to be the number of transitions which N^t goes through until it enters state j for the first time starting from state i. The probability distribution of T_{ij} (called the *first entrance probabilities* from state i to state j) may be defined as

$$f_{ij}(m) = \text{Prob}\,[T_{ij} = m] = \begin{cases} 0 & m = 0 \\ p_{ij} & m = 1 \\ \text{Prob}\,[N^{t+m} = j,\, N^{t+h} \neq j,\, h = 1,\cdots,m-1 \mid N^t = i] & m \geq 2. \end{cases} \tag{9}$$

The state space $\mathcal{S}$ for N^t consists of the set of nonnegative integers $\{0,1,2,\cdots,n_c,\ n_c+1,\cdots,M\}$ which is partitioned into the safe region $\{0,1,2,\cdots,n_c\}$ and the unsafe region $\{n_c+1,\cdots,M\}$. Now consider the modified state space $\mathcal{S}' = \{0,1,2,\cdots,n_c,n_u\}$ where n_u is an absorbing state such that N^t is now characterized by the transition probabilities

$$p_{ij}' = \begin{cases} p_{ij} & i,j = 0,1,\cdots,n_c \\ \sum_{l=n_c+1}^{M} p_{il} & i = 0,1,\cdots,n_c;\, j = n_u \\ 0 & i = n_u;\, j = 0,\cdots,n_c \\ 1 & i = j = n_u. \end{cases} \tag{10}$$

Define the random variable T_i to be the number of transitions which N^t goes through before it enters the unsafe region for the first time starting from state i in the safe region. T_i is called the *first exit time from state i*. The probability distribution of T_i is defined to be $\{f_i(m)\}_{m=1}^{\infty}$ which are called the *first exit probabilities.* It is trivial to show that starting from state i $(0 \leq i \leq n_c)$, the first entrance probabilities into the absorbing state n_u in the modified state space $\mathcal{S}'$ are the same as the first exit probabilities into the unsafe region of $\mathcal{S}$. Using (9), such probabilities are given by the following recursive equation [30],

$$f_{in_u}(m) = p_{in_u}'\delta(m-1) + \sum_{j=0}^{n_c} p_{ij}'f_{jn_u}(m-1) \qquad m \geq 1;\, i \neq n_u$$

where

$$\delta(m) = \begin{cases} 1 & m = 1 \\ 0 & \text{otherwise.} \end{cases}$$

The above equation can be rewritten in terms of the first exit probabilities as

$$f_i(m) = \sum_{j=n_c+1}^{M} p_{ij}\delta(m-1) + \sum_{j=0}^{n_c} p_{ij}f_j(m-1) \qquad m \geq 1;\, 0 \leq i \leq n_c \tag{11}$$

where $f_i(m)$ can be solved recursively for $m \geq 1$ starting with $f_i(0) = 0$ for all i.

The probability distribution $\{f_i(m)\}_{m=1}^{\infty}$ for the random variable T_i typically has a very long tail and cannot be easily computed. We had defined earlier FET as a stability measure for an unstable channel. By our definition, FET is the same as the expected value of the random variable T_0. Let $\bar{T}_i$ be the expected value and $\overline{T_i^2}$ be the second moment of T_i. These moments can be obtained by solving a set of linear simultaneous equations. It can easily be shown[30] that

$$T_i = \begin{cases} 1 & \text{with probability } p_{in_u}' \\ 1 + T_j & \text{with probability } p_{ij} \end{cases}$$

from which we obtain [28], [30]

$$\bar{T}_i = 1 + \sum_{j=0}^{n_c} p_{ij}\bar{T}_j \qquad i = 0, 1,\cdots,n_c \tag{12}$$

$$\overline{T_i^2} = 2\bar{T}_i - 1 + \sum_{j=0}^{n_c} p_{ij}\overline{T_j^2} \qquad i = 0, 1,\cdots,n_c. \tag{13}$$

Equation (12) forms a set of $n_c + 1$ linear simultaneous equations from which $\{\bar{T}_i\}_{i=0}^{n_c}$ can be solved and the stability measure FET $(= \bar{T}_0)$ determined. After $\{\bar{T}_i\}_{i=0}^{n_c}$ have been found, (13) can then be solved in a similar manner for $\{\overline{T_i^2}\}_{i=0}^{n_c}$.

Numerical Results

With the stability measure defined above, we are now in a position to examine quantitatively the tradeoff among channel stability, throughput and delay for *unstable* channels. Below we first give a computational procedure to solve for $\bar{T}_i$ and hence, FET. We then compute these quantities for various values of K, S_o, and M (corresponding to different channel load lines). The trading relations among channel stability, throughput, and delay are then illustrated.

The solution of the set of simultaneous equations in either (12) or (13) requires inverting the $(n_c + 1)$ by $(n_c + 1)$ matrix of p_{ij} for $i, j = 0, 1,\cdots,n_c$. When n_c is large, this becomes a nontrivial task because of the large number of computational steps and large computer storage requirement for the $[p_{ij}]$ matrix. The fact that $p_{ij} = 0$ for $j \leq i - 2$ in (1) and (2) enables us to use an algorithm given in the Appendix which is very efficient in terms of both computer time and space requirements. For our purposes, this algorithm is superior to conventional methods such as Gauss elimination [31] for solving linear simultaneous equations. In this algorithm, each p_{ij} is used exactly once and can be computed using (1) or (2) only when it is needed in the algorithm. This eliminates the need for storing the $[p_{ij}]$ matrix and practically eliminates any computer storage constraint on the dimensionality of the problem. The number of arithmetic operations $(+ - \times \div)$ required by the above algorithm is in the order of $2n_c^2$ which is comparable to that of Gauss elimination.

In Fig. 12, we show FET as a function of K for the infinite population model and for fixed values of the channel throughput rate S_o (at the channel operating point). We see that FET can be improved by either decreasing the channel throughput rate S_o or by increasing K (which in turn increases the average packet delay). The infinite population model results give the worst case estimates for channel stability as demonstrated in Fig. 13 in which we show FET as a function of M for $K = 10$ and four values of S_o. Note that FET increases as M decreases and there is a critical value of M below which the channel is always stable in the sense of Fig. 7(a). As M increases to infinity, FET reaches a limiting value corresponding to the infinite population model with a Poisson channel input. Fig. 14 is similar to Fig. 12 except now the number of users

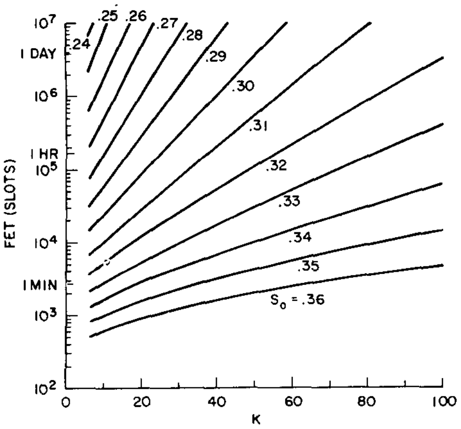


Fig. 12. FET values for the infinite population model.

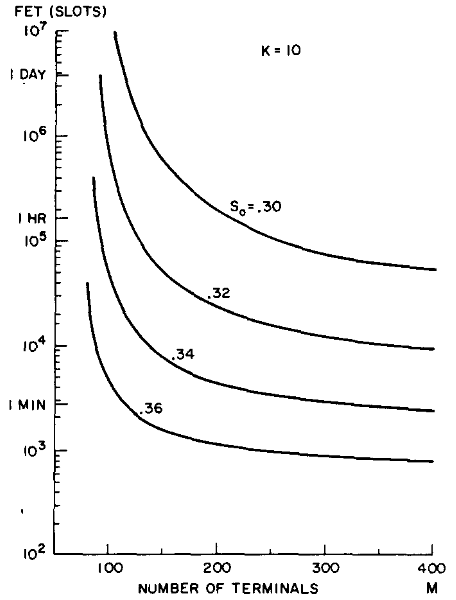


Fig. 13. FET versus M.

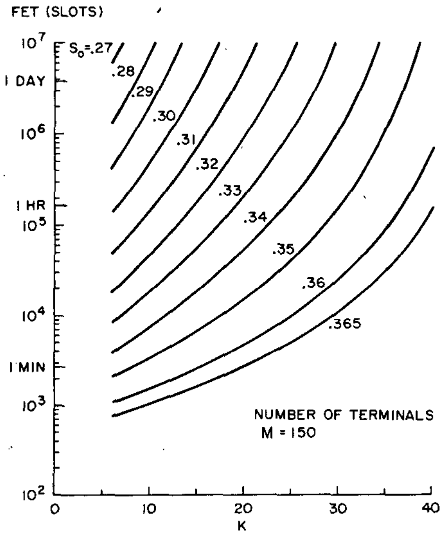


Fig. 14. FET values for a finite user population ($M = 150$).

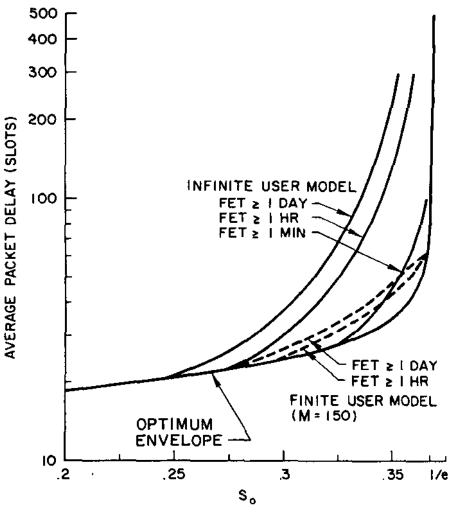


Fig. 15. Stability-throughput-delay tradeoff.

M is 150. Recall that if M is finite, the channel will become stable when K is sufficiently large.

As an example, we see that in Fig. 14 for $M = 150$, if the channel throughput rate S_o is kept at approximately 0.28 and $K = 10$ is used, the channel is estimated to fail once every two days on the average. If this is an acceptable level of channel reliability, then no other channel control procedure is necessary except to restart the channel whenever it goes into saturation. However, if absolute channel reliability is required at the same throughput-delay performance, then dynamic channel control strategies should be adopted. Channel control schemes have been studied [10] and the results will be published in a forthcoming paper [1].

In Fig. 15, we show the optimum performance envelope in Fig. 2 as a lower bound for the throughput-delay tradeoff of the infinite population model. This corresponds to the performance of the channel at the channel operating point. However, from Fig. 7, we see that *the channel operating point* (n_o,S_o) *provides no information regarding the stability behavior of the channel.* The equilibrium performance given by (n_o,S_o) is achievable in the long run if M is small enough such that the channel is stable; elsewhere it is achievable only for some random time period whose average is estimated by our stability measure FET.

In addition to the infinite population model optimum envelope, we also show in Fig. 15 two sets of equilibrium throughput-delay performance curves with guaranteed FET values. The first set consists of three solid curves corresponding to an infinite population model with the stability measure FET $\geq$ 1 day, 1 hour, and 1 minute. Again, these results represent worst case estimates if M is actually finite. The second set consists of two dashed curves corresponding to $M = 150$ with FET $\geq$ 1 day and 1 hour. These results were obtained by looking up the values of K and S_o in Fig. 12 or Fig. 14 corresponding to a fixed FET. The average packet delay was then obtained from Fig. 2. This figure illustrates the *fundamental tradeoff* among channel stability, throughput and delay. In [1], [10], control strategies are devised to dynamically regulate the channel usage to achieve truly stable throughput-delay performance close to the optimum performance envelope.

A Design Example

The designer of a slotted ALOHA channel is faced with the problem of deciding whether he wants 1) a stable channel by limiting its use to a small population of users and sacrificing channel utilization, or 2) an unstable channel which supports a large population of users operating at a certain level of reliability (some value of FET). For example, suppose K is chosen to be 10. (Note in Fig. 2 that $K = 10$ gives close to optimum equilibrium throughput-delay performance over a wide range of channel throughput rate.) Also, suppose that the channel users have an average think time of 20 s which, for our channel numerical constants, correspond to 888 time slots. Now if we draw channel load lines in Fig. 4 with a slope equal to -888, the channel is stable up to approximately 110 channel users. For $M = 110$, the channel throughput rate S_o is about 0.125 packet/slot. From Fig. 2, the average packet delay is roughly 16.5 time slots ($=0.37$ s). The same channel can be used (in an unstable mode) to support 220 users at a channel throughput rate of $S_o = 0.25$ packet/slot. The average packet delay is 21 time slots ($=0.47$ s). From Fig. 12, for $K = 10$ and $S_o = 0.25$, the average up time (FET) of the channel is approximately two days for the infinite population model. Note that this value represents a lower bound for the FET of $M = 220$. Thus, we see that if a channel failure rate of once every two days on the average is an acceptable level of reliability, the second channel design is much more attractive than the first since the number of channel users is more than doubled at a modest increase in delay.

CONCLUSIONS

In this paper, the rationale and some advantages for broadcast packet communication have been discussed. A mathematical model was then formulated for a slotted ALOHA random access system. Using this model, a theory was put forth which gives a coherent qualitative interpretation of the system stability behavior. Quantitative estimates for the relative instability of unstable channels were obtained through definition of the stability measure FET. Numerical results were shown illustrating the trading relations among channel stability, throughput and average packet delay. These results establish tools for the performance evaluation and design of an uncontrolled slotted ALOHA system. Further improvement in the system performance may be accomplished through adaptive control techniques studied in [1], [10].

APPENDIX

The algorithm below solves for the variables $\{t_i\}_{i=0}^{I}$ in the following set of $(I + 1)$ linear simultaneous equations,

$$t_0 = h_0 + \sum_{j=0}^{I} p_{0j}t_j \tag{A1}$$

$$t_i = h_i + \sum_{j=i-1}^{I} p_{ij}t_j \qquad i = 1, 2,\cdots,I. \tag{A2}$$

The Algorithm

1) Define

$$e_I = 1$$

$$f_I = 0$$

$$e_{I-1} = \frac{1 - p_{II}}{p_{I,I-1}}$$

$$f_{I-1} = -\frac{h_I}{p_{I,I-1}}.$$

2) For $i = I - 1, I - 2,\cdots,1$ solve recursively

$$e_{i-1} = \frac{1}{p_{i,i-1}}\left[e_i - \sum_{j=i}^{I} p_{ij}e_j\right]$$

$$f_{i-1} = \frac{1}{p_{i,i-1}}\left[f_i - h_i - \sum_{j=i}^{I} p_{ij}f_j\right].$$

3) Let

$$t_I = \frac{f_0 - h_0 - \sum_{j=0}^{I} p_{0j}f_j}{\sum_{j=0}^{I} p_{0j}e_j - e_0}$$

$$t_i = e_it_I + f_i \qquad i = 0, 1, 2,\cdots,I - 1.$$

Derivation of the Algorithm

Define

$$t_i = e_it_I + f_i \qquad i = 0, 1, 2,\cdots,I - 1 \tag{A3}$$

and

$$e_I = 1$$

$$f_I = 0. \tag{A4}$$

The last equation in (A2) is

$$t_I = h_I + p_{I,I-1}t_{I-1} + p_{II}t_I.$$

Substituting $t_{I-1} = e_{I-1}t_I + f_{I-1}$ into the above equation, we get

$$t_I = h_I + p_{I,I-1}e_{I-1}t_I + p_{I,I-1}f_{I-1} + p_{II}t_I.$$

Equating the coefficients of t_I and the constant terms, we have

$$e_{I-1} = \frac{1 - p_{II}}{p_{I,I-1}}$$

$$f_{I-1} = -\frac{h_I}{p_{I,I-1}}. \qquad \text{(A5)}$$

Equation (A2) can be rewritten as follows,

$$t_{i-1} = \frac{1}{p_{i,i-1}} \left[t_i - h_i - \sum_{j=i}^{I} p_{ij} t_j \right]. \qquad \text{(A6)}$$

In each of the above equations, use (A3) to substitute for t_i. We then have

$$e_{i-1} t_I + f_{i-1} = \frac{1}{p_{i,i-1}} \left[e_i t_I + f_i - h_i - \left(\sum_{j=i}^{I} p_{ij} e_j \right) t_I - \sum_{j=i}^{I} p_{ij} f_j \right].$$

Equating the coefficients of t_I and the constant terms, we get

$$e_{i-1} = \frac{1}{p_{i,i-1}} \left[e_i - \sum_{j=i}^{I} p_{ij} e_j \right]$$

$$f_{i-1} = \frac{1}{p_{i,i-1}} \left[f_i - h_i - \sum_{j=i}^{I} p_{ij} f_j \right]. \qquad \text{(A7)}$$

From (A4), (A5), and (A7), e_i and f_i $(i = I - 2, I - 3, \cdots, 1, 0)$ can then be determined recursively.

We next solve for t_I. Equation (A3) is used to substitute for t_i in (A1), which then becomes

$$e_0 t_I + f_0 = h_0 + \left(\sum_{j=0}^{I} p_{0j} e_j \right) t_I + \sum_{j=0}^{I} p_{0j} f_j.$$

Solving for t_I in the above equation, we have

$$t_I = \frac{f_0 - h_0 - \sum_{j=0}^{I} p_{0j} f_j}{\sum_{j=0}^{I} p_{0j} e_j - e_0}. \qquad \text{(A8)}$$

Finally, $t_i (i = 0, 1, 2, \cdots, I - 1)$ can be obtained from (A3), since e_i, f_i, and t_I are all known. The derivation of the algorithm is now complete.

REFERENCES

[1] S. S. Lam and L. Kleinrock, "Packet switching in a multiaccess broadcast channel: dynamic control procedures," *IEEE Trans. Commun.*, to be published; also in IBM Corp., Yorktown Heights, N. Y., Res. Rep. RC-5062, Oct. 1974.

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

[3] W. Crowther, R. Rettberg, D. Walden, S. Ornstein, and F. Heart, "A system for broadcast communication: reservation—ALOHA," in *Proc. 6th Hawaii Int. Conf. System Sciences*, Univ. Hawaii, Honolulu, Jan. 1973.

[4] R. M. Metcalfe, "Steady-state analysis of a slotted and controlled ALOHA system with blocking," in *Proc. 6th Hawaii Int. Conf. System Sciences*, Univ. Hawaii, Honolulu, Jan. 1973.

[5] N. Abramson, "Packet switching with satellites," in *1973 Nat. Comput. Conf., AFIPS Conf. Proc.*, vol. 42. New York: AFIPS Press, 1973, pp. 695–702.

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

[7] L. G. Roberts, "Dynamic allocation of satellite capacity through packet reservation," in *1973 Nat. Comput. Conf., AFIPS Conf. Proc.*, vol. 42. New York: AFIPS Press, 1973, pp. 711–716.

[8] L. Kleinrock and S. S. Lam, "On stability of packet switching in a random multi-access broadcast channel," in *Proc. 7th Hawaii Int. Conf. System Sciences (Special Subconf. Computer Nets)*, Univ. Hawaii, Honolulu, Jan. 8–10, 1974.

[9] S. Butterfield, R. Rettberg, and D. Walden, "The satellite IMP for the ARPA network," in *Proc. 7th Hawaii Int. Conf. System Sciences (Special Subconf. Computer Nets)*, Univ. Hawaii, Honolulu, Jan. 8–10, 1974.

[10] 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. Calif., Los Angeles, Mar. 1974; also in Univ. of Calif., Los Angeles, Tech. Rep. UCLA-ENG-7429, Apr. 1974.

[11] L. G. Roberts, "Data by the packet," *IEEE Spectrum*, vol. 11, pp. 46–51, Feb. 1974.

[12] L. G. Roberts and B. D. Wessler, "Computer network development to achieve resource sharing," in *1970 Spring Joint Comput. Conf., AFIPS Conf. Proc.*, vol. 36. Montvale, N. J.: AFIPS Press, 1970, pp. 543–549.

[13] P. E. Jackson and C. D. Stubbs, "A study of multiaccess computer communications," in *1969 Spring Joint Comput. Conf., AFIPS Conf. Proc.*, vol. 34. Montvale, N. J.: AFIPS Press, 1969, pp. 491–504.

[14] J. Martin, *Systems Analysis for Data Transmission.* Englewood Cliffs, N. J.: Prentice-Hall, 1972.

[15] J. R. Pierce, "Network for block switching of data," in *IEEE Conv. Rec.*, New York, Mar. 1971.

[16] W. W. Chu, "A study of asynchronous time division multiplexing for time-sharing computer systems," in *1969 Fall Joint Comput. Conf., AFIPS Conf. Proc.*, vol. 35. Montvale, N. J.: AFIPS Press, 1969, pp. 669–678.

[17] P. Baran, "On distributed communications XI. Summary overview," Rand Corp., Santa Monica, Calif., Memo. RM-3767-PR, Aug. 1964.

[18] L. Kleinrock, *Communication Nets: Stochastic Message Flow and Delay.* New York: McGraw-Hill, 1964 (out of print); reprinted by New York: Dover, 1972.

[19] D. W. Davies, "The principles of a data communication network for computers and remote peripherals," in *Proc. Int. Fed. Information Processing Congr.*, Edinburgh, Scotland, 1968, pp. D11–D15.

[20] P. Wright, "Facing a booming demand for networks," *Datamation*, vol. 19, pp. 138–139, Nov. 1973.

[21] H. Frank, M. Gerla, and W. Chou, "Issues in the design of large distributed computer communication networks," in *Proc. Nat. Telecommunications Conf.*, Atlanta, Ga., Nov. 26–28, 1973.

[22] L. G. Roberts, "Extensions of packet communication technology to a hand held personal terminal.," in *1972 Spring Joint Comput. Conf., AFIPS Conf. Proc.*, vol. 40. Montvale, N. J.: AFIPS Press, 1972, pp. 295–298.

[23] In *Inst. Elec. Eng. (London) Proc. Int. Conf. Satellite Systems for Mobile Communications and Surveillance*, Mar. 13–15, 1973.

[24] N.T . Gaarder, "ARPANET satellite system," ARPA Network Inform. Center, Stanford Res. Inst., Menlo Park, Calif., ASS Note 3 (NIC 11285), Apr. 1972.

[25] L. G. Roberts, "ALOHA packet system with and without slots and capture," ARPA Network Inform. Center, Stanford Res. Inst., Menlo Park, Calif., ASS Note 8-(NIC 11290), June 1972.

[26] L. Kleinrock and F. A. Tobagi, "Carrier-sense multiple access for packet switched radio channels," in *Proc. Int. Conf. Communications*, Minneapolis, Minn., June 1974.

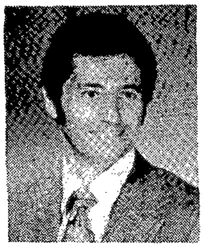
[27] L. Kleinrock, *Queueing Systems, Vol. I, Theory, Vol. II, Computer Applications.* New York: Wiley-Interscience, 1975.

[28] E. Parzen, *Stochastic Processes.* San Francisco, Calif.: Holden-Day, 1962.

[29] J. W. Cohen, *The Single Server Queue.* New York: Wiley, 1969.

[30] R. Howard, *Dynamic Probabilistic Systems, Vol. 1: Markov Models and Vol. 2: Semi-Markov and Decision Processes.* New York: Wiley, 1971.

[31] E. J. Craig, *Laplace and Fourier Transforms for Electrical Engineers.* New York: Holt, Rinehart, and Winston, 1964.



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

From 1951 to 1957, he was employed at the Photobell Company, Inc., New York, N. Y., an industrial electronics firm. He spent the summers from 1957 to 1961 at the M.I.T. Lincoln Laboratory, Lexington, Mass., first in the Digital Computer Group and later in the Systems Analysis Group. At M.I.T. he was a Research Assistant, initially with the Electronic Systems Laboratory, and later with the Research Laboratory for Electronics, where he worked on communication nets in the Information Processing and Transmission Group. After completing his graduate work at the end of 1962, he worked at Lincoln Laboratory on communication nets and on signal detection. In 1963 he accepted a position on the faculty at the University of California, Los Angeles, where he is now Professor of Computer Science. He is a referee for numerous scholarly publications, book reviewer for several publishers, and a consultant for various aerospace, research, and governmental organizations. He is principal investigator of a large contract with the Advanced Research Projects Agency (ARPA) of the Department of Defense. He has published over 60 papers and is the author of *Communication Nets; Stochastic Message Flow and Delay* (New York: McGraw-Hill, 1964), *Queueing Systems, Vol. 1: Theory* and *Vol. 2: Computer Applications* (New York: Wiley-Interscience, 1975). His main interests are in communication nets, computer nets, data compression, priority queueing theory, and theoretical studies of time-shared systems.

Dr. Kleinrock is a member of Tau Beta Pi, Eta Kappa Nu, Sigma Xi, the Operations Research Society of America, and the Association for Computing Machinery. He was awarded a Guggenheim Fellowship in 1971.

★

Simon S. Lam (S'69–M'74) was born in Macao on July 31, 1947. He received the B.S.E.E. degree in electrical engineering from Washington State University, Pullman, in 1969, and the M.S. and Ph.D. degrees in engineering from the University of California, Los Angeles, in 1970 and 1974, respectively.

At the University of California, Los Angeles, he held a Phi Kappa Phi Fellowship from 1969 to 1970, and a Chancellor's Teaching Fellowship from 1969 to 1973. He also participated in the ARPA Network project at UCLA as a postgraduate research engineer from 1972 to 1974 and did research on satellite packet communication. Since June 1974 he has been a research staff member with the IBM Thomas J. Watson Research Center, Yorktown Heights, N. Y. His current research interests include computer-communication networks and queueing theory.

Dr. Lam is a member of Tau Beta Pi, Sigma Tau, Phi Kappa Phi, Pi Mu Epsilon, and the Association for Computing Machinery.

Quantization Error in Predictive Coders

DONALD S. ARNSTEIN, MEMBER, IEEE

***Abstract*—Predictive coders have been suggested for use as analog data compression devices. Exact expressions for reconstructed signal error have been rare in the literature. In fact most results reported in the literature are based on the assumption of Gaussian statistics for prediction error. Predictive coding of first-order Gaussian Markov sequences are considered in this paper. A numerical iteration technique is used to solve for the prediction error statistics expressed as an infinite series in terms of Hermite polynomials. Several interesting properties of predictive coding are thereby demonstrated. First, prediction error is in fact close to Gaussian, even for the binary quantizer. Sencond, quantizer levels may be optimized at each iteration according to the calculated density. Finally, the existence of correlation between successive quantizer outputs is shown. Using the series solutions described above, performance in terms of mean-square reconstruction error versus bit rate can be shown to parallel the theoretical *rate distortion* function for the first-order Markov process by about 0.6 bits/sample at low bit rates.**

I. INTRODUCTION

THE PREDICTIVE coder shown in Fig. 1 has been suggested for video and voice coding applications. The usefulness of predictive coding for data compression and digitization of analog signals is well known, yet due to its nonlinear nature, few exact solutions for quantization error can be found. Let us note that the signal y_k can be considered a sample function of a Gaussian Markov sequence generated according to the recursion equation

$$y_k = w_k + \sum_{n=1}^{N} a_n y_{k-n} \tag{1}$$

and w_k is a sequence of independent unit variance (zero mean) Gaussian variables. Such functions are known as Gaussian autoregressive sequences and may be used to model signals whose spectra contain no zeros.

Paper approved by the Associate Editor for Communication Theory of the IEEE Communications Society for publication after presentation at the 1972 Information Theory Symposium, Asilomar, Calif. Manuscript received August 6, 1974; revised October 9, 1974. This work was supported by the U.S. Army under Grant DA-ARO-D-31-124-71-G89 through a research assistantship at the Department of System Science, University of California, Los Angeles, Calif.

The author is with the Computer Sciences Corporation, Falls Church, Va. 22046.