On the Measured Performance of Packet Satellite Access Schemes*

L. Kleinrock and M. Gerla
University of California, Computer Science Department
Los Angeles, CA 90024, U.S.A.

ABSTRACT

We report on some extended measurements of the performance of packet communications in a satellite environment. We discuss three multi-access protocols: F-TDMA, R-TDMA and S-ALOHA. After discussing the problem and the SATNET experiment, we describe the protocols, present analytic results describing their behavior, and then finally present the measured performance of these access schemes as compared to theory. This paper generalizes our earlier work in two ways: first, it presents measurements for more than two earth stations (as was reported earlier); and secondly it presents measurements for *controlled* S-ALOHA. We find, as expected, that our approach to controlling S-ALOHA gives performance which is superior to earlier control mechanisms and certainly superior to uncontrolled S-ALOHA.

1. MULTI-ACCESS BROADCAST SATELLITE CHANNELS

In the field of computer communications, there is a growing interest in the use of stationary satellites in geosynchronous orbit as a means for providing long-haul, wide-band, inexpensive communications. This need comes about deep in the backbone of packet networks as well as in spanning oceans to interconnect widely separated networks. While the satellite channel can be used in many different ways, the particular configuration of interest to this paper is that of a multi-access broadcast system whereby all earth terminals have access to the full bandwidth and in which the satellite acts simply as a transponder broadcasting its transmission simultaneously to all earth stations within its broad coverage footprint.

We are thus faced with a communications channel which must be shared among many distributed users in some cost-effective fashion. Since each of the message sources (the earth stations) generates data at unpredictable times, then we are faced with a queueing problem in which the bursty nature of the message sources causes conflicts for access to the satellite channel. In a typical queueing system, these conflicts are resolved by some predetermined schedule (for example, first-come-first-served) at the service station. In our case, however, since the service station (the satellite channel) is remote from the users, we must pay a price for organizing these users into a cooperating queue [1]. Since the only means we have available for passing around scheduling information is the communication channel which we intend to control, this price is represented by a throughput and/or delay degradation of the channel itself.

An ever increasing number of access schemes have recently been analyzed in the published literature [2, 3, 4, 5, 6, 7, 8]. Some of these schemes work well only in an environment where the ratio of the propagation delay to the packet transmission time is much less than one. However in the satellite environment this ratio is more often on the order of from 10 to 20; for example a stationary satellite introduces a propagation delay on the order of 250 to 270 milliseconds and a 1000 bit packet travelling over a 56 KBPS channel takes roughly 18 milliseconds giving a ratio of propagation delay to packet transmission time of roughly 15. Therefore, in evaluating satellite multiple access schemes, we restrict our attention to those which operate well in the long propagation time environment.

Many of these access schemes fall into one of three categories. First we have the *static reservation* systems in which capacity is preassigned to each user once and for all and cannot be shared on a demand basis; F-TDMA is an example of such a scheme (see below). These schemes tend to perform well when the channel is heavily (and evenly) loaded. However, in the case of bursty traffic requirements the assigned capacity is seriously underutilized. In the second class of access schemes, namely, the *demand access* schemes, capacity is dynamically allocated according to user requirements. With such schemes no capacity is assigned to idle sources, thus improving channel efficiency. Clearly, a price must be paid in making the dynamic assignment and this price comes about in the form of control overhead. An example of such a scheme is R-TDMA described below. The third class of schemes, the *random access* schemes, are such that users attempt to grab the channel whenever they need it, hoping they will not "collide" with other users' packets. When such a collision occurs, none of the conflicting users succeed in getting their transmission through and so we have the case of "resource smashing". Slotted ALOHA is an example of such an access scheme as described below. In Table 1 we show this trade-off whereby nature will extract her price in one fashion or another; this price comes about due to the fact that our sources cannot organize themselves into a cooperating queue at no charge [1].

	IDLE SLOTS	CONTROL OVERHEAD	COLLISIONS
STATIC RESERVATION	Yes	No	No
DYNAMIC RESERVATION	No	Yes	No
RANDOM ACCESS	No	No	Yes

Table 1. Cost of organizing distributed resources.

In this paper we present measurement and analytic results for one access scheme from each of these three classes. In a previous paper [9] we presented analogous results for two earth stations. Here we extend those results to three earth stations, each of which is equipped to emulate a large number of other earth stations, thereby giving a rich environment for multiple station experiments. Furthermore, earlier results for S-ALOHA were only for the case of an uncontrolled S-ALOHA channel. In a second paper [10] we discussed the necessity for controlling S-ALOHA channels (they are fundamentally unstable access schemes) and presented a model and some simulation results for a new closed loop control mechanism. In that paper we demonstrated through *simulation* that such a scheme had significant advantages over previously discussed control schemes and certainly over uncontrolled S-ALOHA. Here we present *measurement* results

*This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract MDA 903-77-C-0272.

as derived from the SATNET experiment which support that argument and show the particular behavior of controlled S-ALOHA in a multi-station environment.

The remainder of this paper is organized as follows. First we discuss the SATNET experiment and describe the environment in which these measurements were made. Secondly, we discuss the three protocols of interest with two versions of S-ALOHA, namely uncontrolled and controlled. Then we present analytic results (in some cases approximations) predicting the performance of these protocols. Lastly, we give the measurement results.

2. THE SATNET EXPERIMENT

SATNET is an experimental satellite network sponsored by the Advanced Research Projects Agency (ARPA), the Defense Communications Agency, the British Post Office and the Norwegian Telecommunications Administration.

The development of SATNET was initiated in mid 1975. The network consists of four ground stations (two in the Washington, D.C. area at Etam and Clarksburg; one in Goonhilly, England; and one in Tanum, Sweden) interconnected by a simplex, 64 KBPS Intelsat IV-A SPADE channel. The ground station sites are equipped with satellite message processors, called Satellite IMPs (SIMPs) which are extensions of the ARPANET IMPs and which implement channel access and network access protocols. Gateway computers, implemented with PDP-11 hardware, connect SATNET and ARPANET to permit internetwork communications.

The participants in the SATNET experiment are: Bolt Beranek and Newman, Comsat Corporation, Linkabit Corporation, and UCLA in the U.S.; the University College, London in England; and the Norwegian Defense Research Establishment in Norway.

One of the goals of the experiment is to test the feasibility and the efficiency of different channel access schemes. To this end, three among the most representative and better documented access schemes, namely, Fixed Time Division Multiple Access(F-TDMA), Reservation Time Division Multiple Access (R-TDMA), and S-ALOHA, were implemented in SATNET to gain experience with the implementation, operation and performance evaluation of packet satellite networks [11].

Other important goals of the SATNET experiment are: the efficient integration of speech and data traffic; the development of voice conferencing protocols appropriate for the satellite broadcasting environment; and the development of reliable Host-SATNET protocols which protect the network from congestion and permit its interconnection with other networks. To meet these goals, a more sophisticated channel protocol, the Contention-Priority Oriented Demand Assignment (C-PODA) protocol, was developed to efficiently handle a varied mix of traffic requirements (i.e., interactive, batch, and digitized voice) with diversified delay and priority constraints [6]. Tests of C-PODA have recently begun and are not reported in this paper.

The protocols considered in this paper are all based on a slotted channel structure. The channel is subdivided into uniform time slots, each of 30 msec duration, i.e., sufficient to accommodate the maximum size packet (1008 bits). Slots are grouped into frames of 32 slots each. At the beginning of a frame, slots are reserved (one for each station) to broadcast a "routing" packet. The routing packet is used to broadcast routing information, maintain slot synchronization and return channel acknowledgements. In a 4-station system, the routing frame therefore consists of 4 routing slots and 28 data slots.

The maximum data field length in a packet is 1008 bits. Overhead is 264 bits (header + trailing checksum). Preamble and trailing carrier are 245 bits. Total packet length therefore, is 1517 bits, i.e. approximately 27 msec @ 56 KBPS. (Note: the SPADE channel initial capacity of 64 KBPS becomes effectively 56 KBPS after excluding SPADE overhead.) Considering that the slot size is 30 msec, this leaves a guardband of 3 msec between consecutive packets. A large buffer pool (67 buffers) was dedicated to the buffering of in-flight packets in each SIMP to insure full channel utilization in spite of the long propagation delay (.25 sec, round trip).

SATNET communications are error protected by an explicit acknowledgment scheme which requires the retransmission of a packet if the acknowledgment is not received from the destination before an appropriate timeout. Channel acknowledgments are returned by the destination via routing packets or via regular packets (piggybacked acknowledgments). Sequencing and duplicate packet detection, however, are not provided by the channel protocol, and are the responsibility of higher level protocols. A SIMP also derives an implicit (or echo-) acknowledgment by monitoring its own transmissions after a round-trip time. If no echo-acknowledgment is received, the SIMP assumes that the transmission was corrupted by a collision or by uplink noise, and reschedules the packet for retransmission. This approach may, of course, introduce duplicates in the presence of local downlink noise.

3. THE CHANNEL PROTOCOLS

3.1 Fixed-TDMA (F-TDMA) Protocol [11]

F-TDMA is the simplest and most robust protocol that can be implemented in a multiple access satellite system. It is an example of a static reservation access scheme. Slots are equally subdivided among stations. The assignment is permanent, with no provision for dynamic reallocation of unused slots. In an N-station configuration the slots are grouped into frames and are assigned as follows:

$$...R_1R_2...R_ND_1D_2...D_ND_1D_2...D_N...D_1D_2...D_NR_1R_2...R_N$$

FRAME

where D_n denotes a data slot owned by station n and R_n is a routing slot for station n. For example, if N = 8 then a 32 slot frame would consist of 8 routing slots and 24 data slots (3 for each of the 8 stations).

The permanent slot assignment creates N independent subchannels of fixed bandwidth. This makes the analysis of the F-TDMA scheme easier than for dynamic allocation protocols, but not always trivial because of the interdependence between subchannels due to piggybacked ACK's.

3.2 Reservation-TDMA (R-TDMA) Protocol [11]

The R-TDMA protocol establishes a permanent association (ownership) between slots and stations similar to F-TDMA. Unlike F-TDMA, however, the slots not claimed by the original owner may be reassigned on a round-robin basis to the stations that have traffic to send. Thus we have a dynamic reservation scheme.

Each frame is subdivided into a certain number of reservation subframes, each subframe consisting of N reservation slots (where N is the number of stations), and a number of data slots. Reservation slots are smaller than regular slots, and up to three reservations can be packed into a regular slot. Therefore, only one regular slot is used for reservations in two-station and three-station experiments.

Each station declares its backlog (i.e., the number of packets awaiting transmission) using its reservation slot. Reservations are monitored by all stations and synchronized reservation tables are maintained in all stations showing the outstanding transmission requirements. The reservation table is used by the channel scheduler (a distributed algorithm that runs synchronously and identically in all SIMPs) to assign future slots to users in a demand access round-robin fashion.

If, at the beginning of a reservation subframe, a SIMP does not hear all the reservations correctly (because of channel noise, for example) it declares itself out of synchronization and it switches from R-TDMA to F-TDMA for the duration of the following reservation subframe. Thus, the presence of noise in R-TDMA may cause not only the loss of some packets, but also the use of a less efficient channel assignment, namely F-TDMA.

3.3 S-ALOHA Protocol--Uncontrolled [11]

In the S-ALOHA protocol each station maintains two output queues as shown in Figure 1: the new queue (for new packets); and the retransmit queue (for packets that need to be retransmitted because of a previous conflict). All stations follow the same rules for transmission: at the beginning of a slot the station will transmit a packet from the retransmit queue with probability P_R (retransmit gate). Only if the retransmit queue is empty, will the station then transmit a packet from the new queue with probability P_N (new gate).

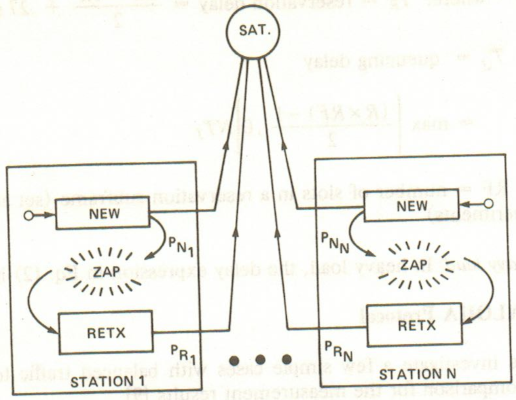


Fig. 1. Slotted ALOHA

Furthermore a packet arriving at an empty station (i.e., both queues empty) is transmitted with probability P = 1. If two or more stations transmit in the same slot, their packets will collide and will mutually destroy each other. The senders detect the conflict after a round trip time by monitoring the channel and promptly return a copy of the collided packet to the retransmit queue.

3.4 S-ALOHA Protocol--Controlled [10]

The uncontrolled version of S-ALOHA discussed above has some serious stability problems. In [10] we surveyed a variety of distributed S-ALOHA controls currently found in the literature and pointed out that most of them are based on one or more restrictive assumptions not valid in the SATNET environment (e.g., only one buffer per station, time invariant traffic load and traffic pattern, large station population, capability to detect collisions, etc.). Therefore we introduced a new control scheme (Closed Loop Control or CLC) which has the ability to adjust to significant traffic pattern changes, a feature which earlier schemes did not enjoy. The CLC scheme, described in [10], is based on the observation made by Abramson [2] that in an uncontrolled S-ALOHA channel with random transmissions, with any number of stations and with any traffic pattern, the maximum throughput is obtained with G = 1, where G is the average sum of transmission and retransmission attempts per slot (including conflicts); that is, it is the average number of packets transmitted per slot. More recently Yemini [12] was able to show that this condition is optimal in a much richer and varied environment.

The CLC control scheme consists of measuring G over a proper history (a window into the past) and then adjusting the ALOHA transmission gates (see Figure 1) so as to nullify the error ϵ = G - 1, using a CLC feedback mechanism.

Because of the distributed implementation of the closed-loop control procedures, it is required that each station broadcast its current gate value, P_N. This is done by stamping the value P_N in the packet header at the time of transmission (or retransmission). Gate value broadcasting is necessary in order to equalize P_N's in all stations and to guarantee fairness by avoiding capture by stations whose P_N is higher than average.

In heavy traffic, the controls maintain the equilibrium condition G = 1, thus achieving optimal channel utilization for any traffic pattern. In light traffic conditions, the value G = 1 obviously cannot be reached; in this case the effect of the closed-loop controls will be to open the gates completely ($P_N = 1$) so as to minimize delay. The system is protected from congestion since a sudden channel overload causes prompt reduction of gate values and therefore a reduction of the load.

In [10], three closed-loop algorithms based on different assumptions regarding channel load information were considered. In this paper, we are concerned with only one of these, namely, the Closed-Loop Control, Collision Non-Detect Algorithm which does not require the detection of collisions. In fact, the SATNET hardware installed at the time of this experiment did not permit us to distinguish between empty and collision slots. (Note: each station will still detect its own collisions, which are identified by the failure to receive the broadcast ACK correctly.)

The algorithm consists of three routines: (a) estimation of collisions, (b) estimation of total channel load G; (c) ALOHA gate value update. A formal definition of the algorithm is provided below:

Assumptions: Over a window of W slots, each station monitors the total success rate S (= total number of successful transmissions divided by W), its own success rate S_i and collision rate C_i, and calculates $\bar{P}$, the average gate value over all transmitting stations.

Algorithm: Every W slots, station *i* updates its parameters with the following steps:

(a) Estimate total collision rate C':

If $C_i = S_i = 0$, then:
If S = 0, let G = 0, go to (c)
Otherwise, let G = 1, go to (c)

If $C_i > 0$ and $S_i = 0$,
let $G = G_{MAX}$, $\bar{P} = \min(\bar{P}, P_N^{(i)})$
and go to (c)

If $S_i > 0$, let $C' = (C_i/S_i)S$
and go to (b)

Note: C_i/S_i is the ratio of collisions versus successes at station i. By multiplying this ratio by S (total success rate), we obtain the estimate of the total collision rate C'.

(b) Estimate total channel load G:

$$G = S + C'$$

where $0 \leqslant G \leqslant G_{MAX}$

(c) Update gate values:

$$P_N^{(i)} = \bar{P} - (G-1)DP$$

$$P_R^{(i)} = min(P_N^{(i)},\ PRMAX)$$

where $0 \leqslant P_R^{(i)} \leqslant 1;\ 0 \leqslant P_N^{(i)} \leqslant 1$.

Parameters: *W*: History window

DP: Probability increment

G_{MAX}: Ceiling value of *G* estimate

(Set at 2 in our experiments)

$P_{NO}^{(i)}, P_{RO}^{(i)}$: Initial probability gate values

(Set at 1 in our experiments)

PRMAX: Ceiling value for P_R

(Set at .5 in our experiments)

4. ANALYTICAL MODELS

In parallel with the experimental activity a modeling activity was carried out in order to validate the experiments (i.e., to verify the correctness of the protocol implementation and the measurement software implementation) and to assist the experimenters in the preparation of the measurement plan and the selection of the appropriate traffic patterns and experimental parameter ranges [9]. In this section, we present some of the delay and throughput models that we developed for the SATNET protocols.

The *delay* is the "one-way" delay measured from the time the packet is accepted in the source SIMP buffer pool, to the time the packet is received without errors at the destination SIMP. (A packet arriving to a full buffer pool is rejected.) The *throughput* from SIMP A to SIMP B is defined as the number of successful packets from A to B divided by the "available" slots (i.e., elapsed slots minus routing slots). Thus, the maximum throughput achievable on the channel is S = 1 pkt/slot.

4.1 F-TDMA Protocol

We develop an analytical model of the F-TDMA scheme for two traffic cases:

A. *Light and medium load:* the offered load is less than channel capacity so that buffer overflow probability is negligible. Let:

$$T = T_p + T_T + T_Q$$

where: T = average one-way delay

T_p = propagation delay = .25 sec

T_T = transmission delay = $.03 \times (\frac{32}{28}) = .034$ sec

T_Q = queueing and latency delay

By approximating the geometric packet generation process with a Poisson process we obtain the following approximate expression [13]:

$$T_Q = T_T \left[\frac{\rho}{2(1-\rho)} + \frac{N}{2} \right] \quad (1)$$

where ρ = utilization of a subchannel = $NR \frac{32}{28}$

R = packet input rate for the SIMP under consideration (pkts/slot)

N = number of SIMPs (in our experiment, $N \leqslant 4$).

B. *Heavy Load:* the load exceeds channel capacity and drives the channel to saturation. For this case, we use Little's result [14] to obtain the following value of T (in seconds):

$$T = .034 \frac{\bar{q}}{S_i} - T_{ACK} \quad (2)$$

where: S_i = one-way throughput of the station under consideration

$\bar{q}$ = average queue length (new queue + retransmit queue + ACK wait queue)

T_{ACK} = average time between the arrival of a packet at its destination and the return of the ACK to the origin.

The value of $\bar{q}$ is calculated as a function of the traffic pattern and the input rates. Typical values range between 52 and 67 (recall: buffer pool size = 67). The value of T_{ACK} varies between .25 and .75 sec, depending on the traffic in the opposite direction. Eq. (2) is very general (as general as Little's result!) and applies to any channel protocol scheme operating at saturation. In particular, it applies to the S-ALOHA and R-TDMA protocols and we use it below for that purpose. This general relationship is very useful in the experimentation of complex channel protocols since it provides a simple check in the absence of more sophisticated analytical or simulation models.

4.2 R-TDMA Protocol

A few approximate results for 3-station configurations and specific load situations are offered here to permit the validation of selected measurement data points [9].

A. *Light Load (with light opposite traffic):* We recall that a station may use its own slots without prior reservation, provided that such slots were not assigned to other backlogged stations [11]. Since we assume light load, the backlog is negligible, and stations may transmit packets without any reservations. The expression for the delay T is therefore given by:

$$T = 2T_S + T_P \quad (3)$$

where: T_S = effective slot time = $\frac{32}{32-RR-RS}$ $(.03)\,sec$

RR = number of routing slots in a frame = 4

RS = number of slots used for reservations in a frame

B. *Light load (heavy opposite traffic):* In this case, the opposite station has a heavy backlog, thus requiring reservations at all times. The delay expression becomes:

$$T = T_R + T_Q + T_T + T_P \quad (4)$$

where: T_R = reservation delay = $\frac{.03 \times RF}{2} + .27$ sec.

T_Q = queueing delay

$= \max \left\{ \frac{(R \times RF) - 1}{2}, 0 \right\} NT_T$

RF = number of slots in a reservation subframe (set at 10 in our experiments)

C. *Heavy load:* In heavy load, the delay expression in Eq. (2) holds.

4.3 S-ALOHA Protocol

Here we investigate a few simple cases with balanced traffic to provide a comparison for the measurement results [9].

A. *Light and Medium Load, balanced traffic,* $P_N = P_R = P$: The delay T (in seconds) at station i is given by:

$$T = .25 + T_T \Big[1 - \rho + \rho/P + \rho/P(1-\rho)(1-\rho_r) \Big] + \frac{G_i - S_i}{S_i} (.25 + T_T/P) \quad (5)$$

where: $S_i = G_i(1-G_i)^{N-1}$

$\rho = G_i/P$

$\rho_R = (G_i - S_i)/P$

S_i = total successful transmissions per slot from station i

G_i = total transmissions and retransmissions per slot from station i

In order to evaluate T in Eq. (5) we first determine G_i (note that R_i, and therefore S_i, is given); then we calculate ρ and ρ_R.

B. *Heavy Load, balanced traffic,* $P_N = P_R = P$: The delay T (in seconds) is given by Eq. (2), where S_i = throughput (pkts/slot) = $P(1-P)^{N-1}$. As in the F-TDMA and R-TDMA cases, the values of $\bar{q}$ and T_{ACK} depend on the traffic pattern.

C. *Heavy Load, balanced traffic,* $P_N \neq P_R$: The asymmetric gate case is of practical importance since it models most of the S-ALOHA schemes found in the published literature. In those schemes, in fact, a new packet is transmitted immediately (i.e., $P_N = 1$), while a packet that has previously collided is retransmitted after a sufficiently long interval to minimize further conflicts (i.e., $P_R << 1$).

In the asymmetric gate case, the delay T is given by the same expression as in case B; however, the determination of throughput S_i (which is needed to evaluate T in Eq. (2)) is not as straightforward as for case B. We first present the solution for a simple case (N = 2; zero propagation delay) and then discuss the approach for the general case.

For N = 2 and for propagation delay = 0, the system is represented by the Markov chain in Figure 2.

The solution for this model is

$$S = 2S_i = 2\Big[\rho_0 p_N (1-P_N) + (\rho_1/2)(P_R(1-P_N) + P_N(1-P_R)) + \rho_2 P_R (1-P_R) \Big] \quad (6)$$

where:

$\rho_2 = P_N^2/2(1-P_R)P_N^2 + P_N^2 + 2(1-P_R)(1-P_N)P_R$

$\rho_1 = 2(1-P_R)\rho_2$

$\rho_0 = 1 - \rho_1 - \rho_2$

It can be shown that S achieves its maximum value S_{MAX} = 2/3 for $P_N \rightarrow 1$ and $P_R \rightarrow 0$. This is quite interesting if we consider that the maximum value for symmetric gates $P_N = P_R = .5$ is S = .5. Unfortunately, the operation with $P_R << P_N$ leads to undesirable capture conditions. Namely, the stations take turns using the channel and, at any one time, one station will transmit at a high rate P_N while the other station is "idling" in the low transmission rate (P_R) state. This behavior introduces a large variance in delay, as observed experimentally (see section 5).

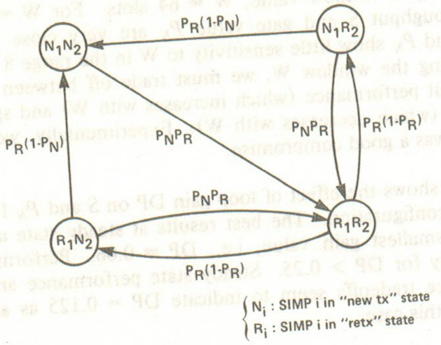


Fig. 2. 2-SIMP Markov Chain

For $N>2$ and propagation delay >0 the Markov chain increases in size and generally can be solved only by using numerical techniques. For the limiting case $P_N \to 1$ and $P_R \to 0$, however, we can still establish some closed form results. In particular, for propagation delay $=0$, $P_N \to 1$, $P_R \to 0$, we have:

$$S = \frac{N}{2N-1}$$

For $N \to \infty$, $S \to .5$. This is better than the result S = 1/e obtained with symmetric gates! (But, of course, we must then live with "capture" side effects.)

For N = 2, $P_N \to 1$, $P_R \to 0$ and propagation delay $\geqslant 0$, we can show that $S \leqslant 2/3$. The upper bound is attained for propagation delay = 0 (as previously shown).

5. MEASURED PERFORMANCE

The following performance measures were used in the evaluation of the SATNET channel access schemes: (a) channel efficiency (b) delay performance (c) fairness (d) stability and (e) robustness to noise. Throughput efficiency and delay performance are generally a function of load distribution. Therefore a variety of traffic patterns were investigated in the SATNET experiments, taking advantage of the very flexible features offered by the traffic generators [9].

Fairness is an issue distinct from channel efficiency. A protocol may be very efficient and yet unfair. For a better appraisal of fairness, more elaborate performance parameters (such as ratios of throughput versus offered rate for each SIMP) must be investigated in carefully designed experiments. An issue related to fairness is the "capture" effect described earlier in which one or a few SIMPs may seize the channel for a prolonged interval of time, while the remaining SIMPs are barred from access.

Some access schemes exhibit unstable behavior; namely, under unfavorable traffic conditions the system is driven into a degraded mode of operation, characterized by low throughput and high delay which persists even after the removal of the cause that produced the degradation. The degree of stability of a protocol may be adequately measured by exposing the system to properly selected time-varying traffic patterns (e.g., pulse patterns). Robustness to noise is a property of great importance for all access protocols that rely on past channel observations and measurement for the scheduling of future transmissions (e.g., R-TDMA and C-PODA) or for the updating of transmission parameters (e.g., S-ALOHA). Incorrect reception of past information due to noise may lead to wrong scheduling and updating decisions and, in extreme conditions, to severe throughput loss until synchronization is reacquired. To test the robustness of the protocols, artificial noise gates were introduced in the SIMPs. To simulate the effect of downlink noise, a certain fraction of the received packets is discarded before processing. This fraction is specified during experiment set-up and may vary from SIMP to SIMP.

With these performance measures in mind, a series of experiments were carried out with various access protocols and with different SATNET configurations (2 and 3 real stations, and up to 30 emulated stations (for the S-ALOHA protocol only)). A selection of the most significant measurement results is reported below and shown in Figures 3-14. In these figures, all data points shown are measured values which were found to agree extremely well with our theoretical predictions; in some figures, we have included these theory points and have denoted them with the symbol ○.

5.1 F-TDMA Measurements

The main purpose of the F-TDMA experiment in SATNET was the calibration of the measurement tools and the establishment of a well understood term of comparison for more complex protocols. Clearly, no issues of fairness, stability and robustness are raised for the F-TDMA protocol.

The experimental configuration consisted of a 3-station system with balanced traffic load. The artificial traffic generator at each station was generating R packets per slot, where R was varied from 0 to 1. During these experiments, throughput S and delay T were measured for each station (recall that throughput S is defined as the number of successful packets per *available slot*, whereas offered rate R represents the average number of arriving packets in *any slot*, including routing slots, and therefore S may exceed R).

Figures 3 and 4 show the throughput S_E and the delay T_E for the station at Etam as a function of the offered rate R_E (in the sequel, the subscript E, T, or G will refer to the station, namely, Etam, Tanum or Goonhilly, respectively, at which the measurement has been taken). Note the excellent agreement between measurements and theory.

5.2 R-TDMA Measurements

For the R-TDMA protocol, the most interesting property we wish to investigate is the high channel efficiency achieved with dynamic assignment. To demonstrate the operation of dynamic assignment, we consider a 3-station experiment with unbalanced input rates, namely, $R_G = 1$, $R_T = 0.1$ and R_E variable between 0 and 1.

Theory predicts that the entire channel capacity available after removing the reservation slots, i.e., S = 0.9, is allocated to the stations based on their respective demands. For $R_E = 0$, we expect $S_G \approx 0.8$ and $S_T \approx 0.1$; and for $R_E = 1$ we expect $S_G \approx S_E \approx 0.4$ and $S_T \approx 0.1$. Throughput measurement results in Figure 5 confirm the theoretical predictions. Delay measurements in Figure 6 also show a reasonable agreement with theory.

The results in Figures 5 and 6 may also be used to investigate the fairness of the R-TDMA protocol. For fairness, the following conditions must be satisfied: (a) all backlogged stations (i.e., the stations with very large queues) get an equal share of the channel; (b) the nonbacklogged stations get a share of the channel equal to their input traffic requirements; and (c) the throughput of the nonbacklogged stations does not exceed the throughput of the backlogged stations. From Figure 5 it is easy to verify that the fairness conditions are satisfied for any value of R_E.

As for the robustness of the R-TDMA protocol, a comprehensive set of experiments with artificial noise gates was carried out for a 2-station configuration. The results of this study, reported in [9], show that for noisy packet fractions as high as 0.3, and for the most unfavorable traffic pattern, the throughput degradation of the R-TDMA protocol in the presence of noise is still superior to that of the F-TDMA protocol, in spite of synchronization loss effects. These results can be extended to the 3-station environment [15] and lead us to conclude that the R-TDMA is extremely robust to noise.

5.3 S-ALOHA (Uncontrolled Measurements)

Among the S-ALOHA uncontrolled experiments, we first show in Figures 7 and 8 the results of a balanced load, symmetric gates ($P_N = P_R = 0.33$), 3-station experiment. The measured performance is in good agreement with the theoretical predictions. Maximum throughput per station is $S = 0.33 \quad (1-.33)^2 = 0.148$ (pkts/slot). Thus, in this balanced case, the maximum S-ALOHA throughput is less than half the maximum throughput measured in the F-TDMA mode, confirming the well-known fact that F-TDMA has a larger capacity than S-ALOHA in a balanced traffic environment.

The uncontrolled S-ALOHA scheme is fair if symmetric gates are used (i.e., $P_N = P_R = P$, where P is the same for all stations). For non-symmetric gate selections (i.e., $P_N \neq P_R$, where P_N and P_R are the same for all stations), however, the system shows symptoms of unfairness. In particular, for $P_R << P_N$ and a heavy offered load, a "capture" situation manifests itself in which one station holds the channel and transmits at a high rate P_N, while the other stations are idling in the retransmit state characterized by very low retransmit rate P_R. Stations randomly take turns in capturing the channel. Simulation and measurement results show that one station may capture the channel for as long as a few minutes [10]. This behavior was predicted by our models as discussed in section 4.3.

The most interesting aspect of the S-ALOHA experiments is, of course, stability. However in order to observe some appreciable symptoms of unstable behavior, we must create an environment with a large number of stations. This can be achieved in SATNET by activating a number of emulated ("fake") stations in each SIMP. From the point of view of packet generation and channel scheduling, each fake station acts as an independent station. Some simplifications were introduced in the emulation in order to keep SIMP processing requirements within reasonable limits (e.g., explicit ACK'S were eliminated for the fake stations); these simplifications, however, do not compromise the validity of the experiments. In all, up to 10 fake stations can be turned on in each SIMP, thus enabling experiments with up to 33 stations, with 3 real SIMP's.

Using real and fake stations, a variety of stability experiments were carried out with artificially enlarged station populations, for different traffic patterns and ALOHA gate parameters P_N and P_R. Figure 12 shows the measured throughput performance S(pkts/slot) as a function of time t (in slots) for a 10-station configuration, with $P_N = P_R = 0.5$. Letting $\mathbf{R} = (R_1, R_2, ..., R_{10})$ be the vector of input rates for the various stations, the traffic pattern for this experiment consists of a steady load $\mathbf{R} = (1,1,0,0,...,0)$ to which a pulse pattern $\mathbf{R}' = (0,0,1,...,1)$ is superimposed of duration = 20 slots starting approximately at t = 1300 slots. According to the above pattern, we have 2 stations generating at rate R = 1 for $0 \leqslant t \leqslant 1300$; then, we have 10 stations generating R = 1 simultaneously in $1300 \leqslant t \leqslant 1320$ (pulse); finally, we return to the 2 active station case for $t \geqslant 1320$.

From Figure 12, we note that the average total throughput before the pulse is S = .5, as expected in a 2 station system with $P_N = P_R = P = .5$ since $S = 2P(1-P)^{N-1}$. The introduction of the pulse, however, creates an environment with 10 backlogged stations. In this situation, the average total throughput $S' = 5(1-.5)^9 = .01$. Since we have a total backlog of $10 \times 20 = 200$ packets to clear, it will take at least $200/.01 = 2 \times 10^4$ slots to recover from the pulse! Figure 12 shows the dramatic decrease of throughput which persists through the end of the experiment.

The above experiment confirms the well-known tendency of S-ALOHA uncontrolled systems to become unstable. Of course, we could have avoided this instability by conservatively setting $P_N = P_R = 1/N_{MAX}$, where N_{MAX} = max number of stations simultaneously active. This, however, would have severely reduced channel efficiency at steady state. As a compromise solution, while still preserving stability, we could have chosen $P_N = .5$ and $P_R = 1/N_{MAX}$; for $N_{MAX} >> 1$ however, this system would exhibit undesirable "capture" effects.

From the above results we conclude that the only viable solution for S-ALOHA stable operation is the introduction of control mechanisms. The next section discusses the experimental findings with controlled S-ALOHA schemes.

5.4 S-ALOHA (Controlled) Measurements

The first set of S-ALOHA (controlled) experiments investigates steady state performance and sensitivity to changes in control parameter values. As explained in section 3.4, the critical control parameters are the window size W and the closed loop gain DP. Intuitively, we know that performance at steady state is improved by increasing W and reducing DP. We also know that excessively large values of W and small values of DP yield a sluggish response to changes. We must therefore seek the values W and DP that optimize the tradeoff between steady state performance and dynamic convergence.

Figure 9 shows the effect of window size W on total throughput S and gate value P_N for a 5 station (3 real + 2 fake) system with balanced load ($R_i = 1$, for all i). As expected, the best results are obtained for the largest window value, W = 64 slots. For W = 64, in fact, both throughput S and gate value P_N are very close to optimum. Both S and P_N show little sensitivity to W in the range $8 \leqslant W \leqslant 64$. In choosing the window W, we must trade off between steady state throughput performance (which increases with W) and speed of convergence (which decreases with W). Experimentally, we found that W = 32 was a good compromise.

Figure 10 shows the effect of loop gain DP on S and P_N for the same 5-station configuration. The best results at steady state are obtained with the smallest gain value, i.e., DP = 0.06. Performance drops significantly for DP > 0.25. Steady state performance and speed of convergence tradeoffs seem to indicate DP = 0.125 as an excellent choice for this case.

Figure 11 shows the effect of loop gain DP on performance for a 20 station configuration. Again, the optimum is achieved for the smallest value DP = 0.06. Performance drops rapidly for larger values of DP, suggesting that perhaps values of DP < 0.06 should be explored. Comparing these results with the results in Figure 10, we note that the sensitivity to changes in DP grows with the number of active stations N. This trend is explained by observing that performance is dependent on the relative step change DP/P_N (rather than absolute step size DP) and by recalling that $P_N = 1/N$ for heavy, balanced loads. Thus, for fixed DP, the ratio DP/P_N increases with N. Sensitivity to changes in window size W, on the other hand, was approximately the same for N = 20 as for N = 5, as suggested by our intuition.

Fairness in the controlled system is guaranteed by the fact that all stations tend to have the same gate value P_N at steady state. Furthermore, the algorithm maintains $P_R \approx P_N$ in heavy load, thus precluding capture situations similar to those observed in uncontrolled systems for $P_R << P_N$.

Stability, the main benefit of the controlled scheme, was carefully investigated with a series of experiments. Figure 13 shows the throughput of a 10 station controlled system as a function of time, subject to a traffic pulse. Traffic pattern conditions are the same as those assumed in Figure 12. Recovery from pulse effects is completed in the controlled system after 1000 slots, while the uncontrolled system required 20,000 slots! The 1000 slot recovery time may be analytically verified by observing that at least 550 slots are required to clear the 200 packets introduced during the pulse (plus the packets generated by the two active stations during the recovery interval), assuming optimal gate values $P_N = P_R = 0.1$. In addition, 150 slots are required for the transition from $P_N = 0.5$ to $P_N = 0.1$, and 300 slots for the transition from $P_N = 0.1$ to $P_N = 0.5$. Figure 14 shows the correlation between gate value $P_N^{(1)}$ and throughput $S^{(1)}$ for station 1. (In a previous paper [10] the performance of the Closed Loop Control (CLC) algorithm was compared (via simulation) with that of another popular algorithm, namely the Control Limit Policy (CLP) [16]. CLC outperforms CLP, in spite of the a priori traffic pattern knowledge advantage assigned to CLP.)

The final set of experiments investigated robustness to noise. Analytically, it can be shown that S-ALOHA controls are insensitive to down link noise (and indeed this is what is required for optimality!). In fact, letting α = fraction of noisy packets and $G' = S + C_i S/S_i$ = channel load estimate (from section 3.4), and recalling that noisy packets are not distinguished from collisions, we have:

$$G'(\alpha > 0) = S(1-\alpha) + (C_i + \alpha S_i)\,\frac{S(1-\alpha)}{S_i(1-\alpha)} = G'(\alpha{=}0)$$

Q.E.D.

The experimental results confirm the robustness of S-ALOHA controls to noise. As an example, for a 20 station experiment with a noisy packet fraction $\alpha = 0.5$, the throughput performance of the controlled system was S = 0.17, only 10% lower than the optimum value S = 0.188! A noise gate $\alpha = 0.5$ would probably destroy reservation schemes based on channel monitoring!

6. CONCLUSION

In this paper we studied three protocols implemented in the SATNET experimental network, namely F-TDMA, R-TDMA and S-ALOHA. These protocols were chosen as representative examples of the three most common types of access schemes used in packet satellite networks, namely: (1) fixed allocation schemes; (2) dynamic allocation schemes; and (3) random access. The above protocols were evaluated in a 3-station configuration, in part extending the 2-station results reported in a previous paper. The evaluation criteria included: (1) delay performance; (2) channel efficiency; (3) fairness; (4) stability; and (5) robustness.

For the R-TDMA protocol, the measurement results confirm the efficiency of channel utilization (due to dynamic allocation), the fairness, and the robustness to noise already observed in the 2-station configuration.

For the S-ALOHA protocol, without controls, we verified fairness for symmetric gate values, but identified the possibility of capture situations for non-symmetric gate values. For the latter case we showed analytically that the total throughput may exceed the well known S-ALOHA channel limits established for symmetric gates. Using carefully designed experiments we then showed that the uncontrolled S-ALOHA scheme may become unstable, thus justifying the need for stability control mechanisms.

Finally, we tested the S-ALOHA protocol equipped with the CLC (closed loop control) stability control mechanism and found some extremely encouraging results. In particular, the controlled scheme is very stable, is fair, is robust to noise and shows a steady state performance very close to the theoretical optimum.

ACKNOWLEDGMENT

The authors are much indebted to L. Nelson, W. Chu, and W. Naylor for their valuable support in the conduction and interpretation of the experiments and to Lou Nelson, George Ann Hornor, Trudy Cook and Carolyn Inouye for their invaluable support in the typing, editing, formatting and preparation of this beautiful phototypeset manuscript.

REFERENCES

[1] Kleinrock, L., "Performance of Distributed Multi-Access Computer Communication Systems", *Information Processing 77,* Proc. of IFIP Congress, Toronto, Canada, August 1977, pp. 547-552.

[2] Abramson, N., "Packet Switching with Satellites", *AFIPS Conference Proceedings,* 1973 NCC, vol. 42, pp. 695-702.

[3] Kleinrock, L. and S. Lam, "Packet Switching in a Slotted Satellite Channel", *AFIPS Conference Proceedings,* 1973 NCC, vol. 42, pp. 703-710.

[4] Roberts, L. G., "Dynamic Allocation of Satellite Capacity through Packet Reservation", *AFIPS Conference Proceedings,* 1973 NCC, Vol. 42, pp. 711-716.

[5] Binder, R., "A Dynamic Packet-Switching System for Satellite Broadcast Channels", *Conference Record, International Conference on Communications,* San Francisco, California, June 1975, vol. 3, pp. 41-1 to 41-5.

[6] Jacobs, I., L. Lee, A. Viterbi, R. Binder, R. Bressler, N. Hsu and R. Weissler, "CPODA - A Demand Assignment Protocol for SATNET", *Proc. of Fifth Data Communications Symposium,* Snowbird, Utah, September 1977, pp. 2-5 to 2-9.

[7] Kleinrock, L. and F. A. Tobagi, "Random Access Techniques for Data Transmission over Packet Switched Radio Channels", *AFIPS Conference Proceedings,* 1975 NCC, vol. 44, pp. 187-201.

[8] Kleinrock, L. and M. Scholl, "Packet Switching in Radio Channels: New Conflict-Free Multiple Access Schemes for a Small Number of Data Users", *Conference Record, International Conference on Communications,* Chicago, Illinois, June 1977, vol. 2, pp. 22.1-105 to 22.1-111.

[9] Gerla, M., L. Nelson, and L. Kleinrock, "Packet Satellite Multiple Access: Models and Measurements", *Proc. of the National Telecommunications Conference,* Los Angeles, California, December 1977, pp. 12.2-1 to 12.2-8.

[10] Gerla, M. and L. Kleinrock, "Closed Loop Stability Controls for S-ALOHA Satellite Communications", *Proc. of Fifth Data Communications Symposium,* Snowbird, Utah, September 1977, pp. 2-10 to 2-19.

[11] Binder, R., "Current SIMP Protocols", Packet Satellite Project Working Note No. 11, Bolt Beranek and Newman, April 1976.

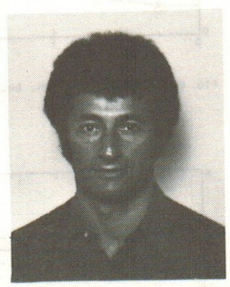
[12] Yemini, Y., "On Channel Sharing in Discrete-Time Packet Switched Multi-Access Broadcast Communications", Ph.D. Dissertation, Computer Science Department, University of California, Los Angeles, 1978.

[13] Scholl, M. O., "Multiplexing Techniques for Data Transmission over Packet Switched Radio Channels", Computer Science Department, University of California, Los Angeles, UCLA-ENG-76123, December 1976.

[14] Kleinrock, L., *Queueing Systems, Vol. I: Theory,* Wiley Interscience (New York), 1975.

[15] Spilling, P., "R-TDMA Measurements Performed on the Satellite Channel", Packet Satellite Project Working Note No. 93, Norwegian Defense Research Establishment, February 1978.

[16] Lam, S. and L. Kleinrock, "Dynamic Control Schemes for a Packet Switched Multi-Access Broadcast Channel", *AFIPS Conference Proceedings,* 1975 NCC, Vol. 44, pp. 143-153.



Leonard Kleinrock is Professor of Computer Science at UCLA. He received his Ph.D. from MIT in 1963. He serves as consultant for many domestic and foreign corporations and governments. He is the author of three major books in the field of computer networks: *Communication Nets: Stochastic Message Flow and Delay* (1964); *Queueing Systems, Vol. 1: Theory* (1975); *Queueing Systems, Vol. 2: Computer Applications* (1976) and has published over 85 articles. He has received various outstanding teacher and best paper awards, the most recent being the 1976 Lanchester prize for the outstanding paper in operations research. At UCLA, he directs a large group in advanced teleprocessing systems and computer networks.

Mario Gerla received a graduate degree in Electrical Engineering from the Politecnico of Milan, Italy in 1966, and the Ph.D. degree in Engineering, specialization computer science, from the University of California, Los Angeles, in 1973. He came to UCLA with a Fulbright scholarship in 1969, and worked on the ARPA network project as a Research Assistant. In January 1973 he joined Network Analysis Corporation, Glen Cove, New York, where as Manager of Computer Networking, he was involved in several computer network design projects for both government and industry. Since July 1977 he has been on the faculty of the Computer Science Department at UCLA, where he is associated with the ARPA packet satellite access schemes, and computer networks in general.

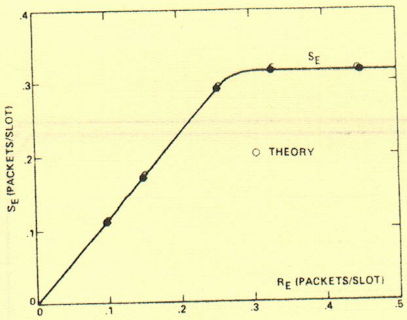


FIG. 3 F-TDMA, 3 STATION, BALANCED LOAD: S_E VS. R_E

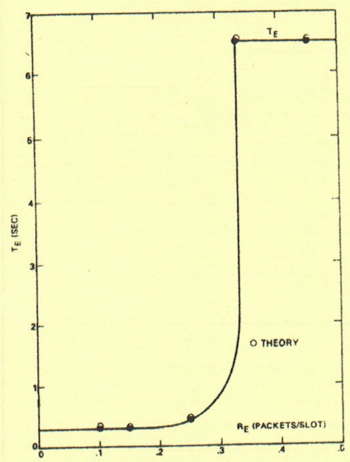


FIG. 4 F-TDMA, 3 STATION, BALANCED LOAD: T_E VS. R_E

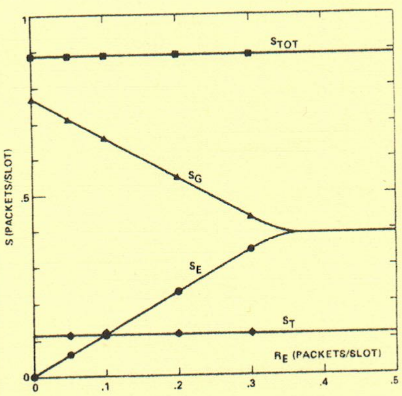


FIG. 5 R-TDMA, 3 STATION, UNBALANCED LOAD: S VS. R_E

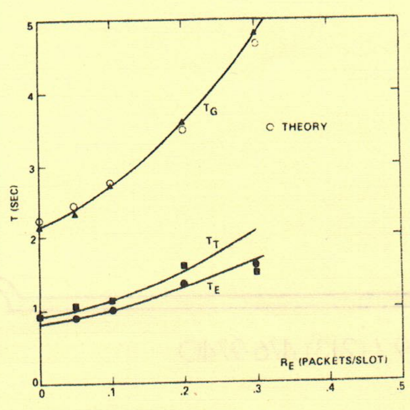


FIG 6 R-TDMA, 3 STATION, UNBALANCED LOAD: T VS. R_E

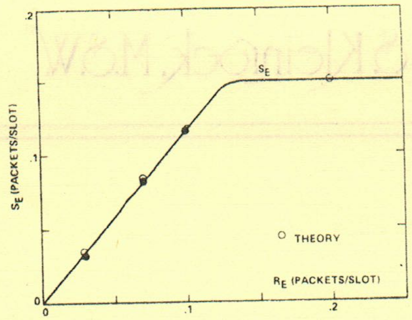


FIG. 7 S-ALOHA, 3 STATION, BALANCED LOAD: S_E VS. R_E

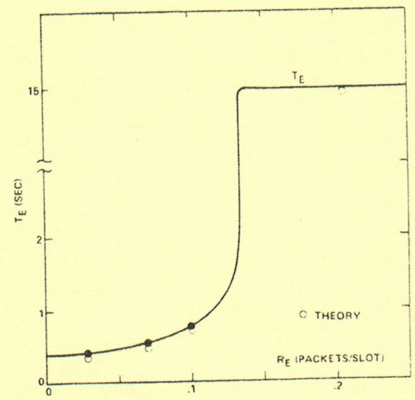


FIG. 8 S-ALOHA, 3 STATION, BALANCED LOAD: T_E VS. R_E

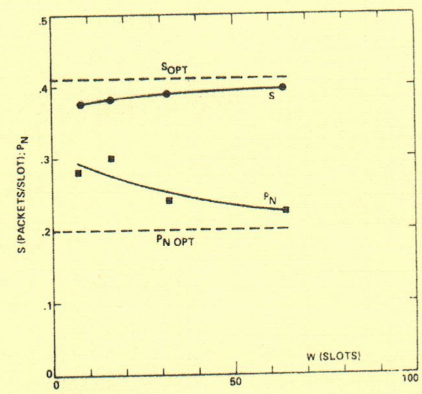


FIG. 9 S-ALOHA CONTROLLED, 5 STATION: W-SENSITIVITY

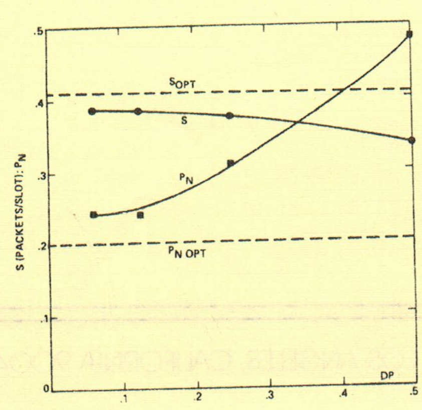


FIG. 10 S-ALOHA CONTROLLED, 5 STATION: DP-SENSITIVITY

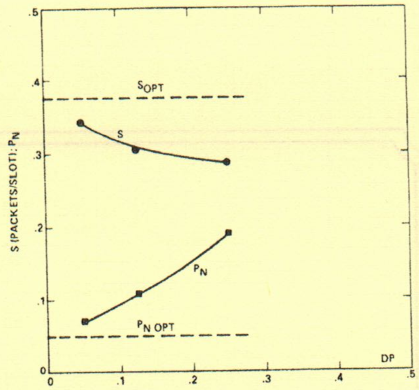


FIG 11 S-ALOHA CONTROLLED, 20 STATION: DP-SENSITIVITY

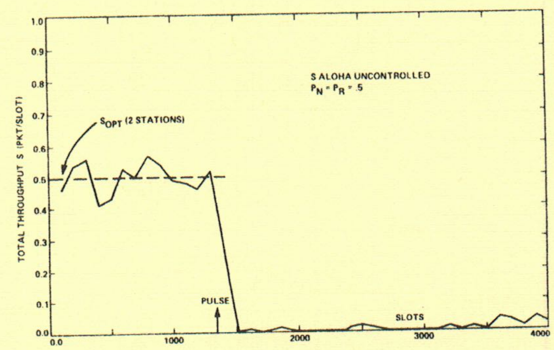


FIG. 12 S-ALOHA UNCONTROLLED, 10 STATION: PULSE EXPERIMENT

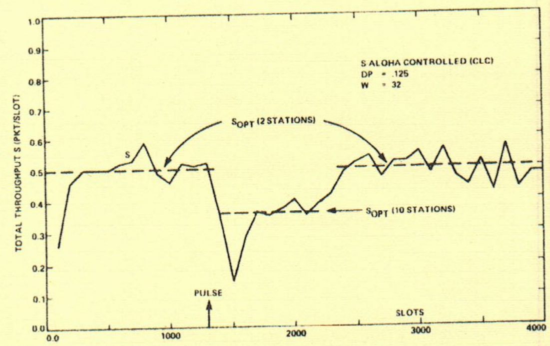


FIG. 13 S-ALOHA CONTROLLED, 10 STATION: PULSE EXPERIMENT

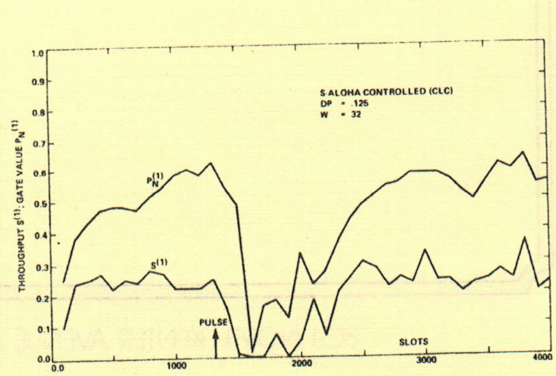


FIG. 14 S-ALOHA CONTROLLED, 10 STATION: PULSE EXPERIMENT