ICC '79 Boston



# ON PROTOCOLS FOR SATELLITE PACKET SWITCHING*

Simon S. Lam

Department of Computer Sciences
The University of Texas at Austin
Austin, Texas 78712

ABSTRACT: The fundamental concepts of three major classes of packet-oriented multiple access protocols are introduced. Performance implications of the satellite propagation delay and on-board processing capability are discussed. The performance of a specific protocol, R-ALOHA, is examined in detail. Results from extensive simulations and analytic models are shown.

## 1. INTRODUCTION

Over the past decade, the sharply declining cost of computing has made possible the emergence of packet switching as a cost effective technology for the transmission of digital data. In addition to improving the economics of data communications, packet switching networks also provide enhanced reliability and functional flexibility of the communication path over circuit switching networks [1]. Presently the transmission of digital voice in packet networks is also under extensive investigation [2]. The ever increasing importance of packet networks is evidenced by the large number of packet-oriented public data networks in existence or being planned in many countries [3]. Within the same time period, communication satellite system costs have come down drastically [4]. A union of the two technologies appears to be most promising. In addition to potential cost reductions, satellites offer special capabilities that can be used to great advantage in packet networks. A shared broadcast satellite channel provides a fully connected network topology with direct "logical" connections between all earth station pairs. It also enables the traffic loads of a large population of geographically distributed users to be statistically averaged via some suitable "algorithm" that provides for dynamic allocation of the satellite transmission capacity.

The problem of multiple access in the design of satellite systems has been solved in the past with voice communications in mind. The design objective has been to maximize the satellite traffic carrying capacity in terms of the number of (voice grade) channels for given constraints of power, bandwidth, error rate etc. Thus multiple access techniques have traditionally been channel-oriented. The satellite resource available is subdivided into separate channels (with FDMA, TDMA or CDMA). The basic unit for allocation is a channel. Channels can be either (i) fixed assigned, or (ii) demand assigned to users [5,6].

The channel-oriented MA techniques are suitable for voice traffic and may also be suitable for some data traffic. Data communications in general, however, have very diverse requirements ranging from inquiry-response systems with intermittent traffic to file transfers with large volumes of data. In addition, user-specified delay constraints need to be met. In this environment, an appropriate measure of traffic carrying capacity is no longer the number of (voice grade) channels but instead the aggregate throughput rate in number of messages (or packets or bits) that can be transported per unit time while satisfying the specified delay constraints.

The problem of interest in this paper begins where the traditional satellite system designers leave off. The problems of modulation, clock synchronization, coding, random noise etc. are assumed to have been solved already. A satellite channel of C bps is available which may have been derived from a FDMA, TDMA or CDMA system at a higher level of satellite resource allocation. The satellite channel is to be shared by a population of distributed users, within the satellite antenna pattern, for communication among themselves. The users have random traffic demands and delay constraints. Our interest lies in packet-oriented protocols for dynamic allocation of the shared satellite channel; these will be referred to as <u>multiple access (MA) protocols.</u>

### Summary of paper

In the next section, the fundamental concepts of three major classes of multiple access protocols (namely, polling, contention and reservation) are described. Performance implications of the satellite propagation delay and on-board processing capability are discussed. In sections 3 and 4, we study the performance of a specific MA protocol, R-ALOHA, which was originally proposed by Crowther et al. [7]. R-ALOHA has the desirable property of being adaptive to the nature of traffic. It behaves like a contention protocol under a light load; under a heavy load it behaves like a reservation/TDMA protocol. Analytic models were obtained by this author in an earlier paper [8] for characterizing statistics of slot usage, message delay etc. Extensive simulations have since been conducted

*This work was supported by National Science Foundation under Grant No. ENG78-01803

and the analytic results were found to predict the performance of R-ALOHA satisfactorily. The key analytic results are summarized in section 3. Results from both analytical models and simulations are compared for various numerical examples in section 4.

## 2. MULTIPLE ACCESS PROTOCOLS

Since the downlink of a satellite channel is broadcast in nature, a data packet transmitted successfully by any user i.e. in the absence of errors due to noise or interference from another user, will arrive correctly at all users. The packet will be accepted by the intended receiver(s) and ignored by others. An MA protocol is an algorithm (possibly distributed as well as non-deterministic) for determining the "access rights" of the users for using the uplink of the satellite channel. In some protocols, the access right is not uniquely determined and it is possible for two or more packet transmissions from different users to "collide" in the channel. It is reasonable to assume that in the absence of some special coding technique, none of the packets involved in a collision will arrive correctly at the intended receivers.

The key measure of performance of a multiple access protocol is its channel throughput versus average delay tradeoff characteristic. The throughput of a channel is defined as follows. Let C be the channel transmission rate in bits per second (bps) and let there be on the average P bits in a transmitted block of data. The channel throughput S is defined to be the ratio of the rate of successfully transmitted data blocks to the rate C/P.

The conflict resolution problem is non-trivial since users are geographically distributed. There are two parts to the problem: (i) to identify users with data to send, which we call ready users, and (ii) to schedule usage of the channel by these users. This problem has been solved by a wide variety of MA protocols using either centralized or distributed control. There are in general three major classes of MA protocols [9,10]

### Polling protocols

A major class of MA protocols consists of polling protocols. A central controller is required and users are passive; the users normally keep quiet whether or not they have data to send. They are queried from time to time by the central controller. A user can transmit data only when so queried. Suppose there are N users in the population. Let w be the average overhead in time associated with querying one user; w includes the round-trip propagation delay between controller and user as well as polling message transmission time. To find out who are the ready users, the overhead per polling cycle (querying all users) is Nw, regardless of the number of ready users present. This overhead is an indirect measure of the responsiveness of the protocol; it can be shown that the mean waiting time of a data packet in roll-call polling, in addition to the transmission and propagation delay of the packet, is bounded below by Nw/2[11]. Hayes [12] recently proposed the method of probing (polling a group of users at a time). Probing was found to significantly reduce the number of queries required per cycle when the channel is lightly loaded with few ready users. In any case, the satellite channel propagation delay R, approximately 0.27 second, enters into w twice. As a result, polling protocols are suitable for satellite packet switching only when response time is not a critical performance factor e.g. delay constraints on the order of minutes rather than seconds.

### Contention protocols

Two other classes of MA protocols require ready users to actively seek channel access instead of waiting to be polled. Under contention protocols there is no attempt to coordinate the ready users to avoid collisions entirely. Instead, each user monitors the downlink broadcast and tries to transmit his data packets the best he can without incurring a conflict. Collided packets are retransmitted by users according to control algorithms driven by local information as well as observable outcomes in the broadcast channel. The ALOHA and slotted ALOHA protocols give rise to a maximum channel throughput of 1/(2e) and 1/e respectively, under the assumption of a large population of very bursty users [13]. Various contention-based protocols, such as R-ALOHA, have been proposed to improve the channel throughput beyond that of slotted ALOHA. When the channel is lightly loaded, the mean delay incurred by a packet under contention protocols approaches the minimum value of channel propagation and transmission delay of the packet.

### Reservation protocols

The objective of reservation protocols is to avoid collisions of data packets entirely. To do so, a queue global to all users needs to be maintained for channel access. Each user when he has data to send generates a request to reserve a place in the queue. A fraction of the satellite channel capacity is used to accommodate the reservation request traffic. Thus the maximum channel throughput of data packets is less than one. There are two key problems to be solved for reservation protocols in general: (1) implementation of the reservation channel, and (2) implementation of the global queue. Since users are geographically distributed, the multiple access problem has not disappeared. It exists now in the access of the reservation channel. Any of the previously described multiple access protocols can be used. However, for simplicity, most proposed reservation protocols adopt either a fixed assigned TDMA protocol or some version of the slotted ALOHA protocol.

There are two approaches to implementing a global queue of requests for a population of distributed users. One is to employ a central controller which tells the ready users when to access the channel; an additional subchannel for controller-to-user traffic is typically required. On the other hand, a distributed control implementation is more interesting and probably more desirable. In this approach, each user maintains information on the status of the global queue and makes his own decision on when is his turn to access the channel. This is possible because reservation requests in the downlink broadcast can be received by all users. An important problem here is the synchronization of queue status information of users in the presence of transmission errors as well as for new users to acquire that information.

The channel throughput of reservation protocols is generally higher than contention protocols as a result of substantially reducing the volume of traffic vulnerable to collisions; the reduction is from the totality of data messages to just one short reservation request per data message. Part of the price that one pays for the gain in channel throughput is an increase of delay. The mean waiting time incurred by a message, in addition to propagation and transmission delay, is at least the mean time to make a reservation request. This minimum mean waiting time can be reduced, however, if one can anticipate future arrivals and make reservations in advance. This is applicable for specialized traffic such as, for instance, digital speech packet streams.

<u>On-board processing</u>

For a satellite channel with on-board processing capability, the multiaccess problem remains essentially the same since it is the uplink that is being contended for! Thus, contention protocols cannot really take advantage of on-board processing. With polling and reservation protocols, however, on-board processing makes it possible to place a central controller at the satellite instead of at one of the terrestrial sites. The propagation delay in w for polling is reduced from 2R to R; it is still significant enough so that it is unlikely that polling protocols used for a satellite channel can meet the response time requirements of most data networks. The minimum mean waiting time for reservation protocols is also reduced from 2R to R, the same as that for distributed control; however, with distributed control, one needs to tackle the queue synchronization problem among distributed users. A protocol that contains both elements of contention and reservation for on-board processing was proposed and studied by Ng and Mark [14].

## 3. THE R-ALOHA PROTOCOL

The R-ALOHA protocol is next described and some analytic performance results are given [8].

Consider a time slotted channel with slots organized into frames with M slots in each frame. Each time slot is long enough to accommodate the transmission of a packet of data. The duration T of a frame must be greater than the maximum channel propagation delay in the network so that each user is aware of the usage status of time slots one frame ago. The network operates without any central control but requires each user to obey the same set of rules for transmitting packets into time slots depending upon what happened in the previous frame. A time slot in the previous frame may be:

- unused, which means that either (a) it was empty, or (b) two or more packets were transmitted into it (a <u>collision</u>) and thus none could be received correctly;
- used, which means that exactly one packet was transmitted into it and the packet was successfully received. (It is assumed that the channel is error-free except for collisions.)

The transmission rules are:

1) If slot m (say) had a successful transmission by user X (say) in the previous frame, slot m is off limits to everyone except user X in the current frame. Slot m is said to be <u>reserved</u> by user X. (Note that user X has exclusive access to slot m as long as he continues to transmit a packet into it in every frame.)
2) Those slots in the last frame which were unused are available for contention by all users according to a slotted ALOHA protocol (the details of which are not specified).

Two protocols are differentiated depending upon whether an end-of-use flag is included in the last packet before a user gives up his reserved slot:

(P1) End-of-use flag not included, and
(P2) End-of-use flag included.

In the analysis [8], a population of N users is considered with identical behavior and message arrival statistics. Messages arrive to each user according to a stationary Poisson process with rate λ messages/second. Each message consists of a group of h packets, with the first two moments $\overline{h}$ and $\overline{h^2}$.

The analysis requires that each user can reserve at most one time slot in a frame at a time. With this requirement, the problem is interesting only if N > M.

The following user models have been considered:

1) <u>Single-message users</u> - Each user handles one message at a time. (The Poisson source shuts itself off until all packets of the current message have been successfully transmitted.)
2) <u>Queued users</u> - Each user has infinite buffering capacity; a queue is maintained with Poisson arrivals at the constant rate of λ messages/second.

The random variable v is defined to be the total number of packets that a user transmits before he gives up a reserved time slot. For the model of single-message users, v is just the number h of packets in a message. For the model of queued users, v is the number of packets that arrive within a <u>busy period</u> of the user queue. The mean value of v is denoted by $\overline{v}$.

It is well known that a slotted ALOHA channel suffers from instability behavior and needs to be adaptively controlled [15]. For our R-ALOHA analysis, we assume the presence of an effective control algorithm such that the following is true

Assumption (A) A successful packet transmission occurs in each nonreserved time slot with a constant probability, S.

Several practical control algorithms are considered below. We next define the following equilibrium probability

$$P_i = \text{Prob [i slots in a frame are used]}.$$

Our first major result follows [8].

$$P_i = \binom{M}{i} U^i (1-U)^{M-i} \qquad 0 \le i \le M \tag{1}$$

where

$$U = \frac{S}{S + (1/\bar{v})} \tag{2}$$

under protocol (P1), and

$$U = \frac{S}{S + [(1-S)/\bar{v}]} \tag{3}$$

under protocol (P2).

Given assumption (A), Eq. (1) can be easily proved for the model of single-message users. For the model of queued users, simulation results indicated that Eq. (1) is an excellent approximation. A comparison of some simulation results and Eq. (1) are shown in the next section.

We note that the channel throughput of R-ALOHA is just U which is related to S and $\bar{v}$ by Eq. (2) or Eq. (3). The R-ALOHA channel throughput can be increased by increasing the slotted ALOHA throughput S or the parameter $\bar{v}$.

Message delay analysis

Some results for the model of queued users are summarized below.

Let $d_A$ be the delay, in number of time slots, incurred by a user to successfully transmit a packet into a nonreserved time slot. Consider, for the moment, that the first and second moments of $d_A$ denoted by $\bar{d}_A$ and $\overline{d_A^2}$ are known. The mean message delay, in number of time slots, is given by

$$\bar{d} = \frac{\bar{x}_0}{1-\lambda(\bar{x}-\bar{x}_0)} + \frac{\lambda(\overline{x_0^2} - \overline{x^2})}{2[1-\lambda(\bar{x}-\bar{x}_0)]} + \frac{\lambda\overline{x^2}}{2(1-\lambda\bar{x})} \tag{4}$$

where

$$\bar{x}_0 = \bar{d}_A + (\bar{h}-1)M$$

$$\bar{x} = \bar{h}M$$

$$\overline{x_0^2} = \overline{d_A^2} + 2\bar{d}_A(\bar{h}-1)M + (\overline{h^2} - 2\bar{h} + 1)M^2$$

$$\overline{x^2} = \overline{h^2}\, M^2$$

Also $\bar{v}$ is just the mean number of packets served within a busy period and is given by

$$\bar{v} = \left(1 + \frac{\lambda T[\bar{d}_A + (\bar{h}-1)M]}{M(1-\lambda\bar{h}T)}\right)\bar{h} \tag{5}$$

To calculate $\bar{d}$ and $\bar{v}$ using Eqs. (4) and (5) we need $\bar{d}_A$ and $\overline{d_A^2}$. However, $d_A$ depends upon the slotted ALOHA channel throughput S which in turn depends upon U and $\bar{v}$ through Eq. (2) or Eq. (3); from now on we shall assume protocol (P1) and the use of Eq. (2). Once the traffic statistics are specified, U is known and is given by

$$U = N\lambda\bar{h}\,T/M \tag{6}$$

Therefore, if we know the slotted ALOHA throughput-delay relationship for the nonreserved time slots (i.e., $\bar{d}_A$ as a function of S), $\bar{v}$ and S can be solved numerically using Eqs. (2) and (5). This we shall do in the next section.

## 4. NUMERICAL RESULTS

In this section, we compare the above analytic results with experimental results from simulation. In the simulation program, we let the number of slots in a frame M=10 and the number of users N=40. For simplicity, a Bernoulli process is used to approximate the Poisson arrival process of each user; in each time slot, a message arrives to each user with probability $\sigma = \lambda(T/M)$. Each message consists of a group of h packets with the following distribution

$$\text{Prob}[h=i] = \begin{cases} 0.2 & i = 1 \\ 0.1 & i = 2,3,4,5,6,7,8,16 \\ 0 & \text{otherwise} \end{cases}$$

which has a mean of 5.3 packets.

Recall that the R-ALOHA protocol is applicable as long as the channel propagation delay is less than the frame duration T. The channel propagation delay was assumed to be zero in both our simulation and analysis results presented below (without any loss of generality).

To obtain numerical results for either analysis or simulation, it is necessary to specify the slotted ALOHA protocol, in particular, the adaptive control algorithm. Many adaptive control algorithms have been proposed and studied in the past; see [16] for instance. Since this is not the primary concern of the present study, we have considered mainly algorithms which are easy to implement.

Specifically, the following class of algorithms that depend only upon local information was considered. Each ready user who does not have a reserved slot transmits the packet at the head of his queue into each nonreserved slot with probability $p_k$ where k = 0,1,2,..... is the number of previous transmissions attempted for the same packet. These algorithms were referred to as heuristic RCP policies in [16]. The following algorithms have been tested in our simulator.

(1)
$$p_k = \begin{cases} 1 & k = 0 \\ 0.2 & k = 1,2,3 \\ 0 & \text{otherwise} \end{cases}$$

(2)
$$p_k = \begin{cases} 1 & k = 0 \\ 0.1 & k = 1,2,3 \\ 0 & \text{otherwise} \end{cases}$$

(Users who have incurred more than 3 collisions "lose" their messages.)

(3) $p_k = (0.5)^k \quad k = 0,1,2,\ldots$

(4) $p_k = 1/(k+1) \quad k = 0,1,2,\ldots$

Both algorithms (1) and (2) were found to give rise to stable channel operation. Stability was achieved at the expense of some "lost" messages when the channel was heavily loaded. Of course, in a real system, messages are not actually lost but rather some users experience temporarily a busy condition and cannot generate new messages. Both algorithms (3) and (4) were found to be far inferior to (1) and (2). In particular, they failed to prevent excessive collisions in nonreserved slots when the channel is heavily loaded. They also give rise to much longer message delays than (1) and (2) when the channel is moderately loaded.

We have also considered the special case when global information is available to individual users. In particular, the instantaneous number n of users competing for a nonreserved slot is known to each such user. The optimal (symmetric) strategy in this case is for each such user to transmit into the nonreserved slot with probability 1/n. This particular algorithm is difficult to implement in practice. However, they give rise to throughput-delay results which are useful as performance bounds.

In Fig. 1 we have shown both experimental results and theoretical results of $P_i$ given by Eq. (1) at four different values of channel throughput U = 0.05, 0.19, 0.60 and 0.86. Note that under both light load (U = 0.05 or 0.19) and heavy load (U = 0.86), experimental and theoretical results agree almost exactly. At U = 0.60, there is some minor discrepancy. The simulation results shown were obtained when the optimal control strategy was used. The good agreement between experimental and theoretical results in Fig. 1, however, was representative of all effective control algorithms considered.

To calculate $\bar{d}$ and $\bar{v}$ using Eqs. (4) and (5), the following formulas for the moments of $d_A$ were used.

$$\bar{d}_A = \left(1 + \frac{1-q}{pq}\right)/(1-U) \qquad (7)$$

$$\overline{d_A^{\,2}} = \frac{2}{(1-U)^2}\left[1+\frac{1-q}{pq}+\frac{1-q}{p^2q^2}\right] - \frac{1}{1-U}\left(1+\frac{1-q}{pq}\right) \qquad (8)$$

where p is equal to 0.2, 0.1 respectively for control algorithms (1) and (2), and q is obtained as a function of S from

$$q = e^{-S/q} \qquad (9)$$

Eqs. (7) - (9) were derived under very strong assumptions of independence using the approach in [17]. For more accurate results, the Markov chain technique in [15] may be used instead. Eqs. (7) - (9) were adopted mainly for their simplicity; despite their inaccuracies, the analytic results of $\bar{d}$, $\bar{v}$ and S for R-ALOHA compare very well with experimental results. These are illustrated in Figs. (2) and (3) for control algorithm (1).

## 5. CONCLUSION

The basic ideas of polling, contention and reservation protocols were surveyed. Performance implications of the satellite channel propagation delay and on-board processing were discussed. The R-ALOHA protocol contains both elements of contention and reservation and its performance was examined. An analytic model was given which was found to give good performance predictions of R-ALOHA.

Acknowledgement

The author thanks Brenda Brasher for her typing assistance.

REFERENCES

[1] Roberts, L., "The evolution of packet switching," Proc. IEEE, Vol. 66, Nov. 1978.

[2] Gitman, I. and H. Frank, "Economic analysis of integrated voice and data networks," Proc. IEEE, Vol. 66, Nov. 1978.

[3] Kelly, P.T.F., "Public packet switched data networks, international plans and standards," Proc. IEEE, Vol. 66, Nov. 1978.

[4] Abramson, N. and E.R. Cacciamani, "Satellites: not just a big cable in the sky," IEEE Spectrum, Vol. 12, Sept. 1975.

[5] Pritchard, W.L., "Satellite communication - An overview of the problems and programs," Proc. IEEE, Vol. 65, March 1977.

[6] Puente, J.G., et al., "Multiple-access techniques for commercial satellites," Proc. IEEE, Vol. 59, Feb. 1971.

[7] Crowther, W. et al., "A system for broadcast communication: Reservation-ALOHA," Proc. 6th HICSS, University of Hawaii, Honolulu, Jan. 1973.

[8] Lam, S.S., "An analysis of the R-ALOHA protocol for satellite packet switching," Conf. Rec. ICC'78, Toronto, June 1978.

[9] Lam, S.S., "Satellite multiaccess schemes for data traffic," Conf. Rec. ICC'77, Chicago, June 1977.

[10] Lam, S.S. "Multiple access protocols," to appear in Computer Communication : State of the Art and Direction for the Future, W. Chou (ed.), Prentice-Hall.

[11] Konheim, A.G. and B. Meister, "Waiting lines and times in a system with polling," J.ACM, Vol. 21, July 1974.

[12] Hayes, J.F., "An adaptive technique for local distribution," IEEE Trans. on Commun., Vol. COM-26, August 1978.

[13] Abramson, N. "Packet switching with satellites," AFIPS Conf. Proc., Vol. 42, 1973.

[14] Ng, S.F.W. and J.W. Mark, "A multiaccess model for packet switching with a satellite having some processing capability," IEEE Trans. on Commun., Vol. COM-25, Jan. 1977.

[15] Kleinrock, L. and S.S. Lam, "Packet switching in a multiaccess broadcast channel: performance evaluation," IEEE Trans. on Commun., Vol. COM-23, April 1975.

[16] Lam, S.S. and L. Kleinrock, "Packet switching in a multiaccess broadcast channel: dynamic control procedures," IEEE Trans. on Commun., Vol. COM-23, Sept. 1975.

[17] Kleinrock, L., and S. S. Lam, "Packet-switching in a slotted satellite channel," AFIPS Conf. Proc., Vol. 42, 1973.

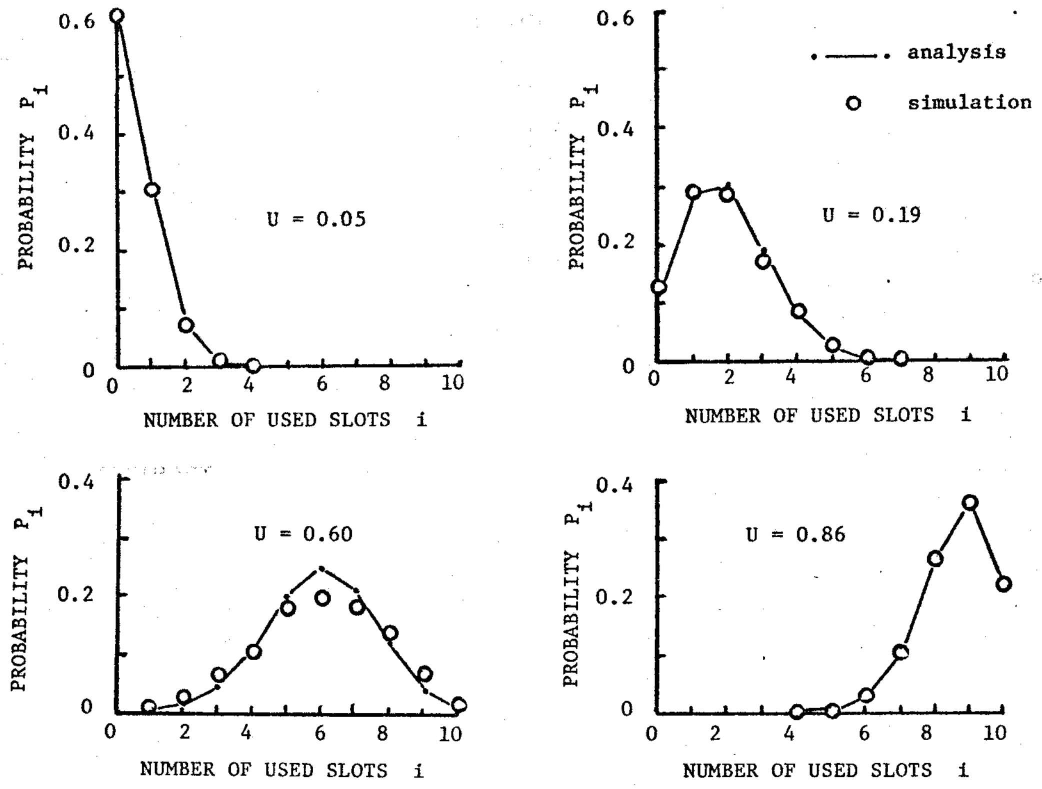


Fig. 1. Analysis and simulation results for $P_i$ .

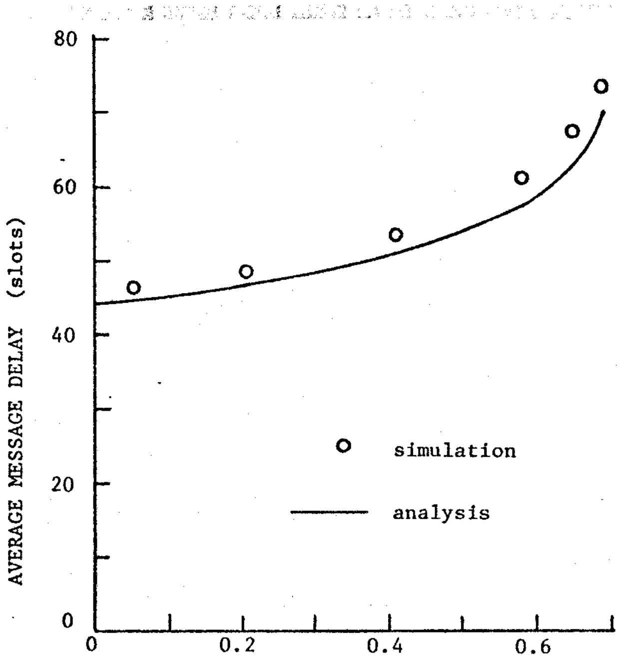


Fig. 2. Ave. message delay vs. channel throughput for control algorithm (1).

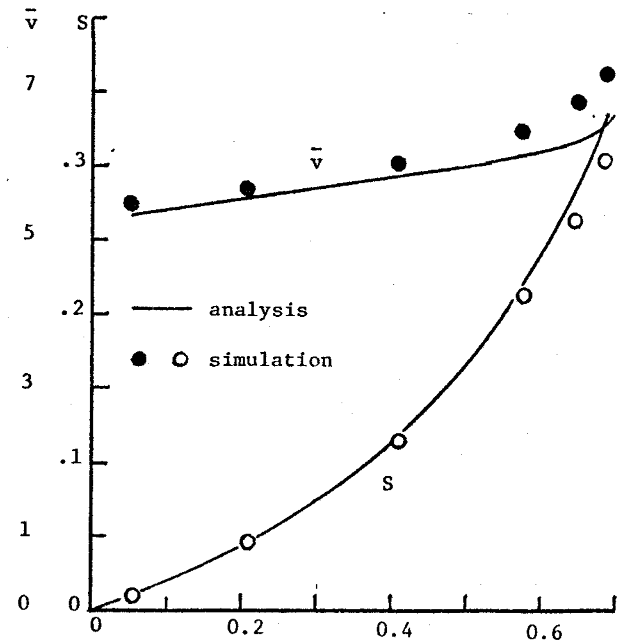


Fig. 3. $\bar{v}$ and S versus channel throughput for control algorithm (1).