Packet Broadcast Networks—A Performance Analysis of the R-ALOHA Protocol

SIMON S. LAM, MEMBER, IEEE

Abstract—**In packet broadcast networks, users are interconnected via a broadcast channel. The key problem is multiple access of the shared broadcast channel. The performance of the R-ALOHA protocol for multiple access is studied in this paper. Two user models with Poisson message arrivals are analyzed; each message consists of a group of packets with a general probability distribution for group size. In the first model, each user handles one message at a time. In the second model, each user has infinite buffering capacity for queueing. Analytic models are developed for characterizing message delay and channel utilization. Bounds on channel throughput are established for two slightly different protocols. Numerical results from both analysis and simulation are presented to illustrate the accuracy of the analytic models as well as performance characteristics of the R-ALOHA protocol.**

Index Terms—**Broadcast channel, broadcast networks, contention algorithms, multiple access protocols, packet broadcasting, performance analysis, queueing, R-ALOHA, satellite networks.**

I. INTRODUCTION

PACKET broadcast networks may be defined to be packet switching networks in which the connectivity requirements of a population of distributed users are furnished by a broadcast medium. Two obvious examples of such broadcast media are satellite and ground radio channels [1], [2]. However, they may also be multipoint cable networks [3], [4]. In recent years, multipoint cable networks have been gaining increasing importance for local area network interconnection [3]–[7].

The basic operation of a packet broadcast network can be explained as follows. A single broadcast channel is shared among a population of distributed users. It is assumed that each user is capable of sending and receiving data at the channel transmission rate of C bits/s. Data messages are segmented into fixed length packets for transmission. Each packet contains its destination address(es) as well as parity bits for error detection. A 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. When packets transmitted by different users "collide" in the channel, it is assumed that (in the absence of some special coding technique) none of the packets involved in a collision will arrive correctly at the intended receivers; such collisions are detected as transmission errors.

Manuscript received May 8, 1979; revised December 10, 1979. This work was supported by the National Science Foundation under Grant ENG78-01803.

The author is with the Department of Computer Sciences, University of Texas at Austin, Austin, TX 78712.

The central problem of a packet broadcast network is *conflict resolution* among the population of users sharing use of the broadcast channel. The problem is nontrivial since the users are typically geographically distributed. The distances involved range from thousands of miles for a satellite network to, perhaps, tens of feet for an in-house cable network.

Many multiple access protocols for conflict resolution have been proposed and studied [8]. They can be classified into three general categories: polling protocols, contention protocols, and reservation protocols. Under polling protocols, a central controller is required and users are passive, i.e., they normally keep quiet whether or not they desire access of the channel. They are queried from time to time by the central controller; a user can transmit data only when so queried.

Both contention and reservation protocols require users who have data to send ("ready" users) to actively seek channel access. Under contention protocols, there is no attempt to coordinate the ready users to avoid collisions entirely. Instead, each user monitors the broadcast channel 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 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 channel capacity is used to accommodate the reservation request traffic. Since users are geographically distributed, the multiple access problem has not disappeared. It exists now in the access of the reservation channel. Synchronization of the distributed queue is also a nontrivial problem.

The three classes of protocols are suitable for different traffic environments. For some traffic environments which are a fixed or time-varying combination of the above, various "mixed" or adaptive protocols have also been proposed [8]–[12]. R-ALOHA is a protocol that contains both elements of contention and reservation. It was originally proposed by Crowther *et al.* [9] to improve the throughput of a satellite channel beyond that of slotted ALOHA which is a pure contention protocol [13]–[15]. Although it was invented for a satellite channel, R-ALOHA can be used for any of the other broadcast media. We shall next describe the R-ALOHA protocol. Assumptions and results of a performance analysis are then presented.

II. The R-ALOHA Protocol

The broadcast channel is assumed to be slotted in time, and the slots are organized into frames with M slots in each frame, just as in traditional TDMA (see Fig. 1). Each time slot is long enough for the transmission of a packet of data. The duration T of a frame is assumed to be greater than the maximum channel propagation delay in the broadcast network. Consequently, 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: 1) it was empty, or 2) two or more packets were transmitted into it (a *collision*) 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 as follows.

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 *reserved* 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 an adaptive algorithm (the details of which will be considered below).

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.

With (P1), a time slot is always wasted whenever a user gives up his reserved slot. With (P2), such time slots are made available for contention according to the second transmission rule. (P2) therefore gives rise to a higher channel throughput than (P1), but (P1) is easier to implement; specifically, users do not have to examine the contents of each transmitted packet and look for the end-of-use flag.

III. The Analysis

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/s. Each message consists of a group of h packets, with the first two moments $\bar{h}$ and $\overline{h^2}$ and probability generating function $H(z)$.

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 $N > M$. Another consequence is that the number of users who may access a nonreserved slot is $N - m$ where m is the number of users holding reserved slots.

The following user models will be considered.

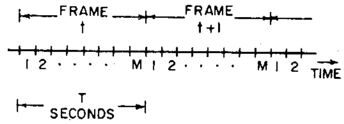

Fig. 1. Frame structure.

1) *Single-Message Users*—Each user handles one message at a time, i.e., the Poisson source shuts itself off until all packets of the current message have been successfully transmitted.

2) *Queued Users*—Each user has infinite buffering capacity; a queue is maintained with Poisson arrivals at the constant rate of λ messages/s.

Note that both user models are more general than user models previously considered for contention-based protocols [10], [14], [15].

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 *busy period* of the user queue; the mean value of v is denoted by $\bar{v}$.

It is well known that adaptive control algorithms are needed for proper operation of contention-based protocols. Such algorithms have been proposed and studied extensively in the past for slotted ALOHA channels based upon pure contention [12], [15], [16], [17]. A mathematically tractable exact analysis of the R-ALOHA protocol using a realistic adaptive control algorithm for contention is not currently available. However, given an effective control algorithm, the following assumption can be made for an approximate analysis.

Constant Throughput Assumption: A successful packet transmission occurs in each nonreserved time slot with a constant probability S.

The above assumption decouples the analysis of the R-ALOHA protocol from the specific details of the contention protocol as long as it has an effective control algorithm for stable operation; in the analysis, S is taken to be the steady-state slotted ALOHA throughput rate. The accuracy of the analytic results will be demonstrated below by comparing them with simulation results. Several practical control algorithms will also be discussed in conjunction with the simulation experiments.

We next define the equilibrium channel utilization probabilities

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

Proposition: For the model of single-message users and given the constant throughput assumption,

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

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).

The key of the proof is to consider each of the M time division multiplexed (TDM) subchannels in Fig. 1 separately. (The ith subchannel is made up of the ith slot of each frame.) Given the constant throughput assumption and the assumption that users have independent identical arrival statistics, outcomes (used or unused time slots) in the M subchannels are statistically independent events. Also, each subchannel has alternating idle and busy periods which are statistically independent and constitute an alternating renewal process. (By definition, an idle period consists of unused time slots. A busy period consists of used time slots.) Let t_{idle} and t_{busy} be the idle and busy period duration, respectively. The probability that a subchannel is busy is [18]

$$U = \frac{E[t_{\text{busy}}]}{E[t_{\text{idle}}] + E[t_{\text{busy}}]}.$$

Since the M subchannels are statistically independent, (1) follows.

Given the constant throughput assumption, we have

$$\text{Prob}\,[t_{\text{idle}} = k \text{ slots}] = S(1 - S)^{k-1} \qquad k = 1, 2, \cdots$$

under (P1) and

$$\text{Prob}\,[t_{\text{idle}} = k \text{ slots}] = S(1 - S)^{k} \qquad k = 0, 1, \cdots$$

under (P2). Hence,

$$E[t_{\text{idle}}] = \begin{cases} 1/S & \text{under (P1)} \\ (1/S) - 1 & \text{under (P2).} \end{cases}$$

Under both (P1) and (P2),

$$E[t_{\text{busy}}] = \bar{v}.$$

Hence,

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

under (P1) and

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

under (P2).

From (1), the probability generating function of P_i is

$$Q(z) = (1 - U + Uz)^M \tag{4}$$

with mean

$$\overline{m} = MU, \tag{5}$$

variance

$$\sigma_m^2 = MU(1 - U),$$

and coefficient of variation

$$C_m = \sigma_m/\overline{m} = \sqrt{(1 - U)/(MU)}.$$

Note that in the above results, v can have a general probability distribution. However, if we restrict v to be geometrically distributed, i.e.,

$$\text{Prob}\,[v = i] = r(1 - r)^{i-1} \qquad i = 1, 2, \cdots$$

for some parameter r, the following equations can be derived for $Q(z)$ using a Markov chain approach (see Appendix):

$$Q(z) = (1 - S + Sz)^M \, Q\left(\frac{z + r(1 - z)}{1 - S + Sz}\right) \tag{6}$$

under (P1) and

$$Q(z) = (1 - S + Sz)^M \, Q\left(\frac{(1 - r)z}{1 - S + Sz} + r\right) \tag{7}$$

under (P2). These equations are the result of a different solution approach to our problem (and under the more restrictive assumption of v being geometrically distributed). It can be easily verified that $Q(z)$ given by (4) with the appropriate expression for U from either (2) or (3) is a solution to (6) or (7) above (respectively), as it should be.

Assuming that v is geometrically distributed, Kanehira [19] derived an expression equivalent to (2) for U under protocol (P1). Our model and results in this paper are more general than his. In particular, v can have an arbitrary probability distribution.

In the model of queued users, v corresponds to the number of packets served in a busy period. In this case, the M subchannels are not statistically independent. Nevertheless, simulation results indicated that (1) is still an excellent approximation.

R-ALOHA Channel Throughput

At this point, let us investigate the maximum possible throughput of a channel that employs the R-ALOHA protocol. The throughput of a channel is defined to be the fraction of time slots in which data packets are successfully transmitted and is equal to U above for R-ALOHA. Let C_{RA} and C_{SA} denote the maximum channel throughput of R-ALOHA and slotted ALOHA, respectively. With $\bar{v}$ fixed in (2) or (3), it is easy to see that U is maximized when S is maximized. We then have under protocol (P1)

$$U \le C_{RA} = \frac{C_{SA}}{C_{SA} + (1/\bar{v})}. \tag{8}$$

Hence,

$$\frac{C_{SA}}{1 + C_{SA}} \le C_{RA} \le 1 \tag{9}$$

for $\bar{v}$ ranging from 1 to ∞.

Similarly, under protocol (P2)

$$U \le C_{RA} = \frac{C_{SA}}{C_{SA} + [(1 - C_{SA})/\bar{v}]}. \tag{10}$$

Hence,

$$C_{SA} \le C_{RA} \le 1 \tag{11}$$

for $\bar{v}$ ranging from 1 to ∞.

For a large population of users, we know that [13], [14]

$$C_{SA} = 1/e.$$

The maximum channel throughput of R-ALOHA is shown in Fig. 2 as a function of $\bar{v}$ for a large population of users for both protocols (P1) and (P2).

Equations (2) and (3) are useful analytic relationships among U, S, and $\bar{v}$. We know that $\bar{v} = \bar{h}$ for the model of single-message users. However, $\bar{v}$ is still unknown for the model of queued users and S is unknown for both user models. Their derivations will be given below. The R-ALOHA channel throughput U, however, is easily obtained using the following argument without first determining S and $\bar{v}$.

Since the system is assumed to be in equilibrium, the channel throughput rate must be equal to the channel input rate. We therefore have for the model of single-message users

$$\begin{aligned} U &= \text{channel input rate in packets/slot} \\ &= (N - \bar{m})\lambda\bar{h}T/M \\ &= (N - MU)\lambda\bar{h}T/M \end{aligned}$$

from which we get

$$U = \frac{(N\lambda\bar{h}T)}{M(1 + \lambda\bar{h}T)}. \tag{12}$$

For the model of queued users, we simply have

$$U = N\lambda\bar{h}T/M. \tag{13}$$

Note that the channel utilization probabilities P_i depend only upon U and are independent of S and $\bar{v}$; they are thus independent of the specific contention protocol. The latter does, however, affect the message delay characteristic of R-ALOHA as shown below.

Message Delay Analysis

For the moment, assume that the delay d_A incurred by a user to successfully transmit a packet into a nonreserved time slot has a known probability density function (pdf) with the Laplace transform $D_A^*(s)$, mean value $\bar{d}_A$, and second moment $\overline{d_A^2}$.

Note, however, that although the delay incurred by a packet in a slotted ALOHA channel has known results [14], [15], such results must be modified for our use here. The nonreserved time slots imbedded within an R-ALOHA channel belong to currently unused TDM subchannels and are thus not contiguous. The number of time slots in between nonreserved slots is a random variable J with

$$\text{Prob}\,[J = j] = U^j(1 - U) \qquad j = 0, 1, 2, \cdots$$

under the constant throughput assumption.

For the model of single-message users, the overall delay d of a message consists of d_A and the transmission delay of the rest of the message (if more than one packet long). The pdf of message delay has the Laplace transform

$$D^*(s) = D_A^*(s)H(e^{-sT})e^{sT} \tag{14}$$

with mean

$$\bar{d} = \bar{d}_A + (\bar{h} - 1)T. \tag{15}$$

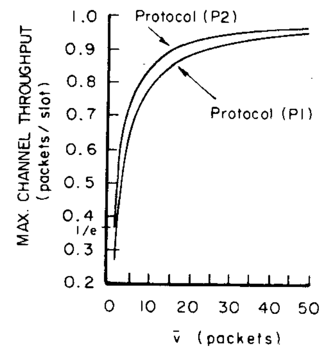

Fig. 2. Maximum channel throughput versus $\bar{v}$.

For the model of queued users, the overall delay d of a message can be obtained by considering each user queue as a generalized $M|G|1$ queue in which the first customer of each busy period receives exceptional service [20]. In that context, the service time pdf of customers who initiate busy periods has the Laplace transform

$$B_0^*(s) = D_A^*(s)H(e^{-sT})e^{sT} \tag{16}$$

with mean

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

and second moment

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

The service time pdf of customers who arrive to find the queue busy has the Laplace transform

$$B^*(s) = H(e^{-sT}) \tag{17}$$

with mean

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

and second moment

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

The pdf of message delay has the Laplace transform [20]

$$D^*(s) = \frac{P_0[(\lambda - s)B_0^*(s) - \lambda B^*(s)]}{(\lambda - s) - \lambda B^*(s)} \tag{18}$$

where

$$\begin{aligned} P_0 &= \frac{1 - \lambda\bar{x}}{1 - \lambda(\bar{x} - \bar{x}_0)} \\ &= \frac{1 - \lambda\bar{h}T}{1 + \lambda(\bar{d}_A - T)}. \end{aligned} \tag{19}$$

The average message delay is

$$\bar{d} = \frac{\bar{x}_0}{1 - \lambda(\bar{x} - \bar{x}_0)} + \frac{\lambda(\overline{x_0^2} - \overline{x^2})}{2[1 - \lambda(\bar{x} - \bar{x}_0)]} + \frac{\lambda\overline{x^2}}{2(1 - \lambda\bar{x})}. \tag{20}$$

Finally, we note that the message delay analysis here can be easily extended (following [21]) to a nonpreemptive priority queue discipline with a finite number of message priority classes.

Solution for S and $\bar{v}$

To solve for the average message delay $\bar{d}$, we need the first and second moments of d_A which depend upon S and $\bar{v}$. For the model of single-message users, we have

$$S = \frac{(N - \bar{m})\lambda T}{(1 - U)M} = \frac{\lambda T(N - MU)}{(1 - U)M} \tag{21}$$

and

$$\bar{v} = \bar{h}. \tag{22}$$

For the model of queued users, the Laplace transform $G^*(s)$ of the busy period pdf can be obtained using a delay cycle analysis [22]

$$G^*(s) = B_0^*(s + \lambda - \lambda Y^*(s)) \tag{23}$$

where

$$Y^*(s) = B^*(s + \lambda - \lambda Y^*(s)). \tag{24}$$

Let $\bar{g}$ and $\bar{y}$ be the mean values obtained from $G^*(s)$ and $Y^*(s)$, respectively. We have

$$\bar{g} = \bar{x}_0(1 + \lambda \bar{y}) \tag{25}$$

where

$$\bar{y} = \frac{\bar{x}}{1 - \lambda \bar{x}} = \frac{\bar{h}T}{1 - \lambda \bar{h}T}. \tag{26}$$

In this user model, $\bar{v}$ is the total number of packets that arrive within a busy period (including those of the initial message). We then have

$$\bar{v} = (1 + \lambda \bar{x}_0 \bar{k})\bar{h} \tag{27}$$

where

$$\bar{k} = 1/(1 - \lambda \bar{x}). \tag{28}$$

Thus,

$$\bar{v} = \left(1 + \frac{\lambda \overline{x_0}}{1 - \lambda \bar{x}}\right) \bar{h} = \left(1 + \frac{\lambda[\bar{d}_A + (h - 1)T]}{1 - \lambda \bar{h}T}\right) \bar{h}. \tag{29}$$

At this point, if we know the 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 (29) together with the equation

$$U = \frac{S}{S + (1/\bar{v})} = N\lambda \bar{h}T/M \tag{30}$$

obtained from (2) and (13) for protocol (P1) or

$$U = \frac{S}{S + [(1 - S)/\bar{v}]} = N\lambda \bar{h}T/M \tag{31}$$

obtained from (3) and (13) for protocol (P2).

IV. 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 $N = 40$. The model of queued users is more general than the model of single-message users and is the only one considered below. 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 contention protocol, in particular, the adaptive control algorithm. Many adaptive control algorithms have been proposed and studied in the past. 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 $k = 0, 1, 2, \cdots$ is the accumulated number of collisions incurred by the same packet. These algorithms were referred to as heuristic RCP policies in [16]. The following algorithms have been tested in our simulator.

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

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

$$3)\ p_k = (0.5)^k \qquad k = 0, 1, 2, \cdots$$

$$4)\ p_k = 1/(k + 1) \qquad k = 0, 1, 2, \cdots.$$

In both 1) and 2) above, users who have incurred more than three collisions for the same packet "lose" all their messages.

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

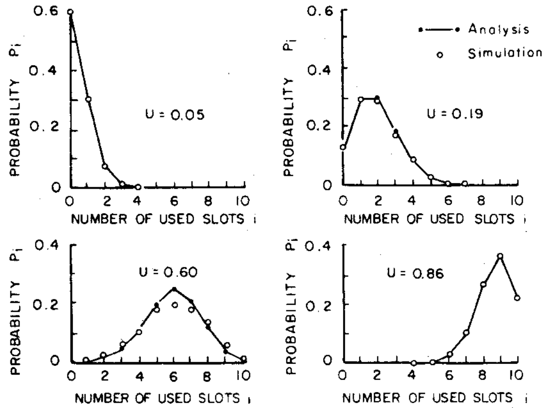

Fig. 3. Analysis and simulation results for P_i.

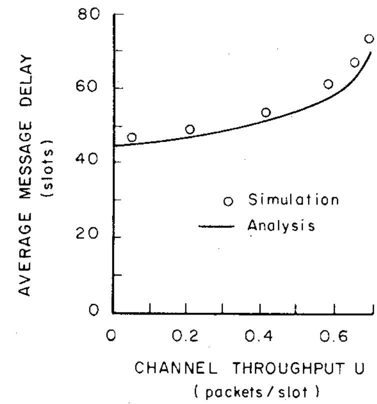

Fig. 4. Average message delay versus channel throughput for control algorithm 1).

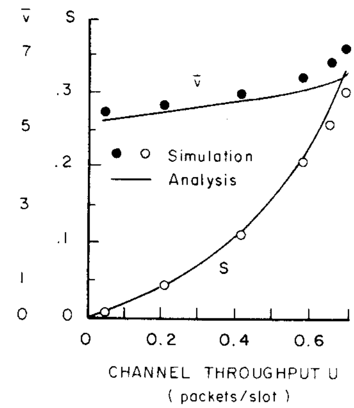

Fig. 5. $\bar{v}$ and S versus channel throughput for control algorithm 1).

message delays than 1) and 2) when the channel is moderately loaded. (However, algorithm 3) has been found to perform well in CSMA protocols [23].)

We have also considered the special case when global information is available to individual users. In particular, the instantanous 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. 3 we have shown both experimental results and theoretical results of P_i given by (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. 3, however, was representative of all effective control algorithms considered.

To calculate $\bar{d}$ and $\bar{v}$ using (20) and (29), the following formulas for the moments of d_A were used.

$$\bar{d}_A = \left(1 + \frac{1-q}{pq}\right) \Big/ (1-U) \tag{32}$$

$$\overline{d_A^2} = \frac{2}{(1-U)^2}\left[1 + \frac{1-q}{pq} + \frac{1-q}{p^2q^2}\right] - \frac{1}{1-U}\left(1 + \frac{1-q}{pq}\right) \tag{33}$$

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

$$q = e^{-S/q}. \tag{34}$$

Equations (32)–(34) were derived under very strong assumptions of independence using the approach in [14]. For more accurate results, the Markov chain technique in [15] may be used instead. Equations (32)–(34) 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. 4 and 5 for control algorithm 1) and in Figs. 6 and 7 for control algorithm 2).

V. Conclusions

Following a brief overview of three classes of multiple access protocols for packet broadcast networks, R-ALOHA was introduced as a protocol that contains both elements of contention and reservation. We found that R-ALOHA is a protocol that adapts itself to the nature of the input traffic. Fig. 2 shows that the R-ALOHA maximum channel throughput ranges from that of slotted ALOHA at one extreme ($\bar{v} = 1$) to that of fixed assigned TDMA channels at the other extreme ($\bar{v} = \infty$).

Two user models with Poisson arrivals were considered. Each arrival is a message consisting of a group of packets. In the first model, each user handles one message at a time. In the second model, each user has infinite buffering capacity for queueing. We make two observations. First, our user models (which permit buffering and queueing) are more general than user models previously considered for contention-based protocols. Second, our performance results on message delay and channel utilization depend upon the constant throughput assumption, which decouples the analysis of the R-ALOHA protocol from specific details of the contention protocol for nonreserved slots. We found that the approximations thus introduced in the analytic results are acceptable as long as an effective control algorithm is implemented for stable channel operation.

In this paper, both our R-ALOHA protocol and user models are such that a tractable analysis is possible. In a practical implementation, some other issues will have to be addressed and the protocol should probably be significantly enriched.

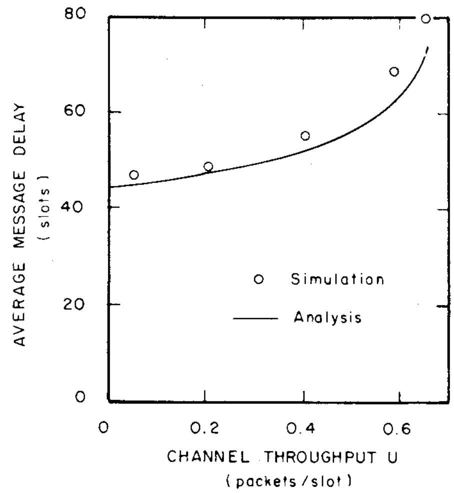

Fig. 6. Average message delay versus channel throughput for control algorithm 2).

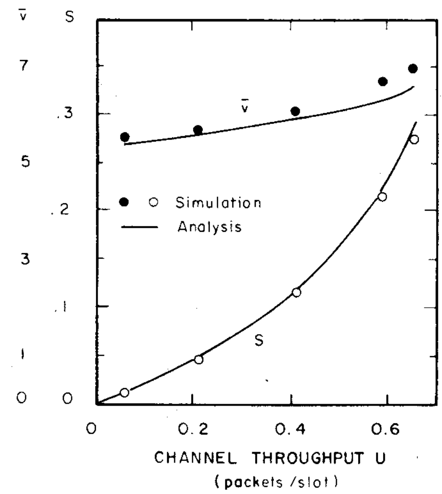

Fig. 7. $\bar{v}$ and S versus channel throughput for control algorithm 2).

One issue is fairness. How do we prevent some users from being locked out of the channel for a long time? One idea is to allocate a minimum number of slots in each frame which cannot be reserved. (Our analysis can be trivially extended to include this.) Some fairness algorithm may also be designed into the contention protocol. The fairness algorithm will affect our analytic model only to the extent that it affects the accuracy of the constant throughput assumption.

Another issue is that of a nonhomogeneous user population. In addition, some users may be more important than others. It will be desirable to enrich the R-ALOHA protocol in several ways. A user may be permitted to reserve multiple slots in a frame. (What algorithm should be used to increase or decrease his reserved slots?) A user may use dummy packets to hold on to a reserved slot when his queue is empty. (This may be a fair and acceptable practice if users can be charged according to the number of slots used.) A user may also be forced to give up his reserved slots even if his queue is nonempty. Such implementation considerations will give rise to rather formidable analysis problems. One might be forced to rely more heavily upon simulation for performance evaluation.

Appendix

Given that v is geometrically distributed with parameter r, we derive below (6) and (7) using a Markov chain approach. Define

q_n = number of used slots in the nth frame

a_n = number of nonreserved slots used (successfully) in the nth frame

d_n = number of reserved slots given up in the nth frame.

In general, we have

$$q_{n+1} = q_n - d_{n+1} + a_{n+1}$$

and thus

$$Q_{n+1}(z) \triangleq E[z^{q_{n+1}}] = E[z^{q_n - d_{n+1} + a_{n+1}}].$$

Under protocol (P1), d_{n+1} and a_{n+1} are dependent upon q_n, but independent of each other. Each has a binomial distribution. We then have

$$\begin{aligned}Q_{n+1}(z) &= \sum_{i=0}^{M} P[q_n = i] z^i E[z^{-d_{n+1}}/q_n = i] E[z^{a_{n+1}}/q_n = i]\\ &= \sum_{i=0}^{M} P[q_n = i] z^i \left(1 - r + \frac{r}{z}\right)^i (1 - S + Sz)^{M-i}\\ &= (1 - S + Sz)^M \sum_{i=0}^{M} P[q_n = i] \left(\frac{z + r(1-z)}{1 - S + Sz}\right)^i\\ &= (1 - S + Sz)^M Q_n \left(\frac{z + r(1-z)}{1 - S + Sz}\right).\end{aligned}$$

Equation (6) is obtained in the limit as $n \to \infty$.

Under protocol (P2), d_{n+1} is dependent upon q_n, while a_{n+1} is dependent upon both q_n and d_{n+1}. We then have

$$\begin{aligned}Q_{n+1}(z) &= \sum_{i=0}^{M} P[q_n = i] z^i E[z^{-d_{n+1}+a_{n+1}}/q_n = i]\\ &= \sum_{i=0}^{M} P[q_n = i] z^i \sum_{j=0}^{i} \binom{i}{j} r^j (1-r)^{i-j} z^{-j} E[z^{a_{n+1}}/q_n = i, d_{n+1} = j]\\ &= \sum_{i=0}^{M} P[q_n = i] z^i \sum_{j=0}^{i} \binom{i}{j} r^j (1-r)^{i-j} z^{-j} (1 - S + Sz)^{M-i+j}\\ &= (1 - S + Sz)^M \sum_{i=0}^{M} P[q_n = i] z^i \left(\frac{1-r}{1 - S + Sz} + \frac{r}{z}\right)^i\\ &= (1 - S + Sz)^M Q_n \left(\frac{(1-r)z}{1 - S + Sz} + r\right).\end{aligned}$$

Equation (7) is obtained in the limit as $n \to \infty$.

REFERENCES

[1] N. Abramson and E. R. Cacciamani, "Satellites: Not just a big cable in the sky," *IEEE Spectrum*, vol. 12, Sept. 1975.

[2] R. E. Kahn, "The organization of computer resources into a packet radio network," *IEEE Trans. Commun.*, vol. COM-25, Jan. 1977.

[3] R. M. Metcalfe and D. R. Boggs, "Ethernet: Distributed packet switching for local computer networks," *Commun. Ass. Comput. Mach.*, vol. 19, July 1976.

[4] A. West and A. Davison, "CNET—A cheap network for distributed computing," Dep. Comput. Sci. and Statistics, Queen Mary College, Univ. London, London, England, Rep. TR120, Mar. 1978.

[5] D. G. Willard, "Mitrix: A sophisticated digital cable communications systems," in *Conf. Rec., Nat. Telecommun. Conf.*, Nov. 1973.

[6] V. A. DeMarines and L. W. Hill, "The cable bus in data communications," *Datamation*, Aug. 1976.

[7] J. E. Thornton *et al.*, "A new approach to network storage management," *Comput. Design*, Nov. 1975.

[8] S. S. Lam, "Multiple access protocols," Dep. Comput. Sci., Univ. Texas at Austin, Tech. Rep. TR-88, Jan. 1979; to appear in *Computer Communication: State of the Art and Direction for the Future*, W. Chou, Ed. Englewood Cliffs, NJ: Prentice-Hall.

[9] W. Crowther, R. Rettberg, D. Walden, S. Ornstein, and F. Heart, "A system for broadcast communication: Reservation-ALOHA," in *Proc. 6th HICSS*, Univ. Hawaii, Honolulu, Jan. 1973.

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

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

[12] L. Kleinrock and Y. Yemini, "An optimal adaptive scheme for multiple access broadcast communication," in *Conf. Rec., ICC'78*, Toronto, Ont., Canada, June 1978.

[13] N. Abramson, "Packet switching with satellites," in *AFIPS Conf. Proc.*, vol. 42. Montvale, NJ: AFIPS Press, 1973, pp. 695–702.

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

[15] ——, "Packet switching in a multiaccess broadcast channel: Performance evaluation," *IEEE Trans. Commun.*, vol. COM-23, Apr. 1975.

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

[17] M. Gerla and L. Kleinrock, "Closed loop stability controls for S-ALOHA satellite communication," in *Proc. 5th Data Commun. Symp.*, Snowbird, UT, Sept. 1977.

[18] D. R. Cox, *Renewal Theory*. London: Methuan, 1962, p. 84.

[19] E. M. Kanehira, "Transient and steady-state analysis of reservation ALOHA," ARPA Network Inform. Cent., Stanford Res. Inst., Menlo Park, CA, ASS NOTE 35 (NIC 14180), Feb. 1973.

[20] P. D. Welch, "On a generalized $M/G/1$ queueing process in which the first customer of each busy period receives exceptional service," *Oper. Res.*, vol. 12, pp. 736–752, 1964.

[21] S. S. Lam, "Delay analysis of a time division multiple access (TDMA) channel," *IEEE Trans. Commun.*, vol. COM-25, pp. 1489–1494, Dec. 1977.

[22] L. Kleinrock, *Queueing Systems Vol. 2: Computer Applications*. New York: Wiley-Interscience, 1976, p. 113.

[23] S. S. Lam, "A carrier sense multiple access protocol for local networks," *Computer Networks*, to be published.

Simon S. Lam (S'69–M'74) was born in Macao on July 31, 1947. He received the B.S.E.E. degree (with Distinction) 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.

From 1972 to 1974 he was with the ARPA Network Project at UCLA and did research on satellite packet communication. From 1974 to 1977 he was a Research Staff member with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, where he worked on performance analysis problems of pakcet switching networks, SNA, and satellite networks. Since September 1977 he has been with the University of Texas at Austin where he is now an Associate Professor of Computer Sciences. His current research interests include computer systems modeling and analysis, computer-communication networks, and distributed systems.

At the University of California, Los Angeles, Dr. Lam held a Phi Kappa Phi Fellowship from 1969 to 1970, and a Chancellor's Teaching Fellowship from 1969 to 1973. In 1975 he received the Leonard G. Abraham Award for the best paper of the year in the field of communication systems published in IEEE TRANSACTIONS ON COMMUNICATIONS. He is a member of Tau Beta Pi, Sigma Tau, Phi Kappa Phi, Pi Mu Epsilon, and the Association for Computing Machinery.