

Packet-switching in a slotted satellite channel*

by LEONARD KLEINROCK and SIMON S. LAM

University of California
Los Angeles, California

INTRODUCTION

Imagine that two users require the use of a communication channel. The classical approach to satisfying this requirement is to provide a channel for their use so long as that need continues (and to charge them for the full cost of this channel). It has long been recognized that such allocation of scarce communication resources is extremely wasteful as witnessed by their low utilization (see for example the measurements of Jackson & Stubbs).¹ Rather than provide channels on a user-pair basis, we much prefer to provide a single high-speed channel to a large number of users which can be shared in some fashion; this then allows us to take advantage of the powerful "large number laws" which state that with very high probability, the demand at any instant will be approximately equal to the sum of the average demands of that population. In this way the required channel capacity to support the user traffic may be considerably less than in the unshared case of dedicated channels. This approach has been used to great effect for many years now in a number of different contexts: for example, the use of graded channels in the telephone industry,² the introduction of asynchronous time division multiplexing,³ and the packet-switching concepts introduced by Baran et al.,⁴ Davies,⁵ and finally implemented in the ARPA network.⁶ The essential observation is that the full-time allocation of a fraction of the channel to each user is highly inefficient compared to the part-time use of the full capacity of the channel (this is precisely the notion of time-sharing). We gain this efficient sharing when the traffic consists of rapid, but short bursts of data. The classical schemes of synchronous time division multiplexing and frequency division multiplexing are examples of the inefficient partitioning of channels.

As soon as we introduce the notion of a shared channel in a packet-switching mode then we must be prepared to resolve conflicts which arise when more than one demand is simultaneously placed upon the channel. There are two obvious solutions to this problem: the first is to "throw out" or "lose" any demands which are made while the channel is in use; and the second is to form a queue of conflicting demands and serve them in some order as the channel becomes free. The

latter approach is that taken in the ARPA network since storage may be provided economically at the point of conflict. The former approach is taken in the ALOHA system⁷ which uses packet-switching with radio channels; in this system, in fact, *all* simultaneous demands made on the channel are lost.

Of interest to this paper is the consideration of satellite channels for packet-switching. The definition of a packet is merely a package of data which has been prepared by a user for transmission to some other user in the system. The satellite is characterized as a high capacity channel with a fixed propagation delay which is large compared to the packet transmission time (see the next section). The (stationary) satellite acts as a pure transponder repeating whatever it receives and beaming this transmission back down to earth; this broadcasted transmission can be heard by every user of the system and in particular a user can listen to his own transmission on its way back down. Since the satellite is merely transponding, then whenever a portion of one user's transmission reaches the satellite while another user's transmission is being transponded, the two collide and "destroy" each other. The problem we are then faced with is how to control the allocation of time at the satellite in a fashion which produces an acceptable level of performance.

The ideal situation would be for the users to agree collectively when each could transmit. The difficulty is that the means for communication available to these geographically distributed users is the satellite channel itself and we are faced with attempting to control a channel which must carry its own control information. There are essentially three approaches to the solution of this problem. The first has come to be known as a pure "ALOHA" system⁷ in which users transmit any time they desire. If, after one propagation delay, they hear their successful transmission then they assume that no conflict occurred at the satellite; otherwise they know a collision occurred and they must retransmit. If users retransmit immediately upon hearing a conflict, then they are likely to conflict again, and so some scheme must be devised for introducing a random retransmission delay to spread these conflicting packets over time.

The second method for using the satellite channel is to "slot" time into segments whose duration is exactly equal to the transmission time of a single packet (we assume constant length packets). If we now require all packets to begin their transmission only at the beginning of a slot, then we

* This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC-15-69-C-0285.

enjoy a gain in efficiency since collisions are now restricted to a single slot duration; such a scheme is referred to as a "slotted ALOHA" system and is the principal subject of this paper. We consider two models: the first is that of a large population of users, each of which makes a small demand on the channel; the second model consists of this background of users with the addition of one large user acting in a special way to provide an increased utilization of the channel. We concern ourselves with retransmission strategies, delays, and throughput. Abramson⁸ also considers slotted systems and is concerned mainly with the ultimate capacity of these channels with various user mixes. Our results and his have a common meeting point at some limits which will be described below.

The third method for using these channels is to attempt to schedule their use in some direct fashion; this introduces the notion of a reservation system in which time slots are reserved for specific users' transmissions and the manner in which these reservations are made is discussed in the paper by Roberts.⁹ He gives an analysis for the delay and throughput, comparing the performance of slotted and reservation systems.

Thus we are faced with a finite-capacity communication channel subject to unpredictable and conflicting demands. When these demands collide, we "lose" some of the effective capacity of the channel and in this paper we characterize the effect of that conflict. Note that it is possible to use the channel up to its full rated capacity when only a single user is demanding service; this is true since a user will never conflict with himself (he has the capability to schedule his own use). This effect is important in studying the non-uniform traffic case as we show below.

SLOTTED ALOHA CHANNEL MODELS

Model I. Traffic from many small users

In this model we assume:

- (A1) an infinite number of users* who collectively form an independent source

This source generates M packets per slot from the distribution $v_i = \text{Prob}[M = i]$ with a mean of S_0 packets/slot.

We assume that each packet is of constant length requiring T seconds for transmission; in the numerical studies presented below we assume that the capacity of the channel is 50 kilobits per second and that the packets are each 1125 bits in length yielding $T = 22.5$ msec. Note that $S_0' = S_0/T$ is the average number of packets arriving per second from the source. Let d be the maximum roundtrip propagation delay which we assume each user experiences and let $R = d/T$ be the number of slots which can fit into one roundtrip propagation time; for our numerical results we assume $d = 270$ msec. and so $R = 12$ slots. R slots after a transmission, a user will

either hear that it was successful or know that it was destroyed. In the latter case if he now retransmits during the next slot interval and if all other users behave likewise, then for sure they will collide again; consequently we shall assume that each user transmits a previously collided packet at random during one of the next K slots, (each such slot being chosen with probability $1/K$). Thus, retransmission will take place either $R+1$, $R+2$, ... or $R+K$ slots after the initial transmission. As a result traffic introduced to the channel from our collection of users will now consist of new packets and previously blocked packets, the total number adding up to N packets transmitted per slot where $p_i = \text{Prob}[N = i]$ with a mean traffic of G packets per slot. We assume that each user in the infinite population will have at most one packet requiring transmission at any time (including any previously blocked packets). Of interest to us is a description of the maximum throughput* rate S as a function of the channel traffic G . It is clear that S/G is merely the probability of a successful transmission and G/S is the average number of times a packet must be transmitted until success; assuming

- (A2) the traffic entering the channel is an independent process

We then have,

$$S = Gp_0 \quad (1)$$

If in addition we assume,

- (A3) the channel traffic is Poisson

then $p_0 = e^{-G}$, and so,

$$S = Ge^{-G} \quad (2)$$

Eq. (2) was first obtained by Roberts¹¹ who extended a similar result due to Abramson⁷ in studying the radio ALOHA system. It represents the ultimate throughput in a Model I slotted ALOHA channel without regard to the delay packets experience; we deal extensively with the delay in the next section.

For Model I we adopt assumption A1. We shall also accept a less restrictive form of assumption A2 (namely assumption A4 below) which, as we show, lends validity to assumption A3 which we also require in this model. Assume,

- (A4) the channel traffic is independent over any K consecutive slots

We have conducted simulation experiments which show that this is an excellent assumption so long as $K < R$.

Let,

$$P(z) = \sum_{i=0}^{\infty} p_i z^i \quad (3)$$

$$V(z) = \sum_{i=0}^{\infty} v_i z^i \quad (4)$$

* These will be referred to as the "small" users.

* Note that $S = S_0$ under stable system operation which we assume unless stated otherwise (see below).

Using only assumption A4 and the assumption that M is independent of $N - M$, we find [10] that $P(z)$ may be expressed as

$$\left[\frac{p_1}{K}(1-z) + P\left(1 - \frac{1-z}{K}\right) \right]^K V(z)$$

If, further, the source is an independent process (i.e., assumption A1) and is Poisson distributed then $V(z) = e^{-S(1-z)}$, and then we see immediately that,

$$\lim_{K \uparrow \infty} P(z) = e^{-G(1-z)}.$$

This shows that assumption A3 follows from assumptions A1 and A4 in the limit of large K , under the reasonable condition that the source is Poisson distributed.

We have so far defined the following critical system parameters: S_0 , S , G , K and R . In the ensuing analysis we shall distinguish packets transmitting in a given slot as being either newly generated or ones which have in the past collided with other packets. This leads to an approximation since we do not distinguish how many times a packet has met with a collision. We have examined the validity of this approximation by simulation, and have found that the correlation of traffic in different slots is negligible, except at shifts of $R+1$, $R+2$, \dots , $R+K$; this exactly supports our approximation since we concern ourselves with the most recent collision. We require the following two additional definitions:

$$q = \text{Prob}[\text{newly generated packet is successfully transmitted}]$$

$$q_i = \text{Prob}[\text{previously blocked packet is successfully transmitted}]$$

We also introduce the expected packet delay D :

$$D = \text{average time (in slots) until a packet is successfully received}$$

Our principal concern in this paper is to investigate the trade-off between the average delay D and the throughput S .

Model II. Background traffic with one large user

In this second model, we refer to the source described above as the "background" source but we also assume that there is an additional single user who constitutes a second independent source and we refer to this source as the "large" user. The background source is the same as that in Model I and for the second source, we assume that the packet arrivals to the large user transmitter are Poisson and independent of other packets over $R+K$ consecutive slots. In order to distinguish variables for these two sources, we let S_1 and G_1 refer to the S and G parameters for the background source and let S_2 and G_2 refer to the S and G parameters for the single large user. We point out that the identity of this large user may

change as time progresses but insist that there be only one such at any given time. We introduce the new variables

$$S = S_1 + S_2 \quad (5)$$

$$G = G_1 + G_2 \quad (6)$$

S represents the total throughput of the system and G represents the traffic which the channel must support (including retransmissions). We have assumed that the small users may have at most one packet outstanding for transmission in the channel; however the single large user may have many packets awaiting transmission. We assume that this large user has storage for queueing his requests and of course it is his responsibility to see that he does not attempt the simultaneous transmission of two packets. We may interpret G_2 as the probability that the single large user is transmitting a packet in a channel slot and so we require $G_2 \leq 1$; no such restriction is placed on G_1 (or on G in Model I).

We now introduce a means by which the large user can control his channel usage enabling him to absorb some of the slack channel capacity; this permits an increase in the total throughput S . The set of packets awaiting transmission by the large user compete among each other for the attention of his local transmitter as follows. Each waiting packet will be scheduled for transmission in some future slot. When a newly generated packet arrives, it immediately attempts transmission in the current slot and will succeed in capturing the transmitter unless some other packet has also been scheduled for this slot; in the case of such a scheduling conflict, the new packet is randomly rescheduled in one of the next L slots, each such slot being chosen equally likely with probability $1/L$. Due to the background traffic, a large user packet may meet with a transmission conflict at the satellite (which is discovered R slots after transmission) in which case, as in Model I, it incurs a random delay (uniformly distributed over K slots) plus the fixed delay of R slots. More than one packet may be scheduled for a future slot and we assume that these scheduling conflicts are resolved by admitting that packet with the longest delay since its previous blocking (due to conflict in transmission or conflict in scheduling) and uniformly rescheduling the others over the next L slots; ties are broken by random selection. We see, therefore, that new packets have the lowest priority in case of a scheduling conflict; however, they seize the channel if it is free upon their arrival. The variable L permits us a certain control of channel use by the large user but does not limit his throughput. We also assume $K, L < R$. Corresponding to q and q_i in Model I, we introduce the success probabilities q_i and q_{i1} ($i=1, 2$) for new and previously blocked packets respectively and where $i=1$ denotes the background source and $i=2$ denotes the single large source. Finally, we choose to distinguish between D_1 and D_2 which are the average number of slots until a packet is successfully transmitted from the background and large user sources respectively.

RESULTS OF ANALYSIS

In this section we present the results of our analysis without proof. The details of proof may be found in Reference 10.

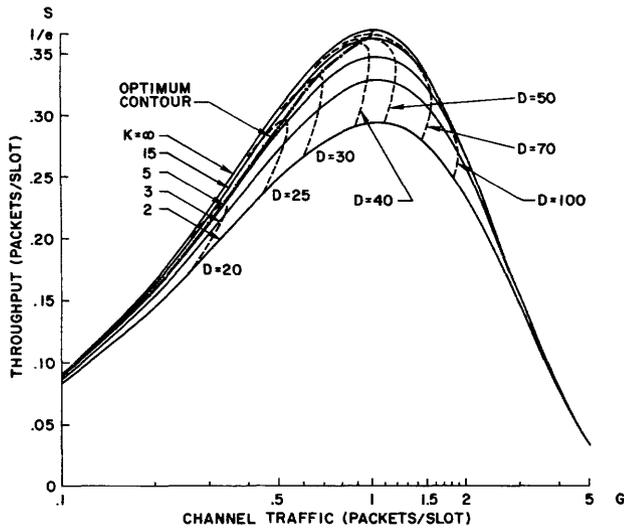


Figure 1—Throughput as a function of channel traffic

Model I. Traffic from many small users

We wish to refine Eq. (2) by accounting for the effect of the random retransmission delay parameter K . Our principal result in this case is

$$S = G \frac{q_t}{q_t + 1 - q} \tag{7}$$

where

$$q = \left[e^{-G/K} + \frac{G}{K} e^{-G} \right]^K e^{-S} \tag{8}$$

and

$$q_t = \left[\frac{1}{1 - e^{-G}} \right] \left[e^{-G/K} - e^{-G} \right] \left[e^{-G/K} + \frac{G}{K} e^{-G} \right]^{K-1} e^{-S} \tag{9}$$

The considerations which led to Eq. (7) were inspired by Roberts¹¹ in which he developed an approximation for Eq. (9) of the form

$$q_t \cong \frac{K-1}{K} e^{-G} \tag{10}$$

We shall see below that this is a reasonably good approximation. Equations (7-9) form a set of non-linear simultaneous equations for S , q and q_t which must be solved to obtain an explicit expression for S in terms of the system parameters G and K . In general, this cannot be accomplished. However, we note that as K approaches infinity these three equations reduce simply to

$$\lim_{K \rightarrow \infty} \frac{S}{G} = \lim_{K \rightarrow \infty} q = \lim_{K \rightarrow \infty} q_t = e^{-G} \tag{11}$$

Thus, we see that Eq. (2) is the correct expression for the throughput S only when K approaches infinity which corresponds to the case of infinite average delay; Abramson⁸ gives this result and numerous others all of which correspond to this limiting case. Note that the large K case avoids

the large delay problem if T is small (very high speed channels).

The numerical solution to Eqs. (7-9) is given in Figure 1 where we plot the throughput S as a function of the channel traffic G for various values of K . We note that the maximum throughput at a given K occurs when $G=1$. The throughput improves as K increases, finally yielding a maximum value of $S=1/e=.368$ for $G=1, K=\infty$. Thus we have the unfortunate situation that the ultimate capacity of this channel supporting a large number of small users is less than 37 percent of its theoretical maximum (of 1). We note that the efficiency rapidly approaches this limiting value (of $1/e$) as K increases and that for $K=15$ we are almost there. The figure also shows some delay contours which we discuss below. In Figure 2, we show the variation of q and q_t with K for various values of G . We note how rapidly these functions approach their limiting values as given in Eq. (11). Also on this curve, we have shown Roberts' approximation in Eq. (10) which converges to the exact value very rapidly as K increases and also as G decreases.

Our next significant result is for packet delay as given by

$$D = R + 1 + \frac{1-q}{q_t} \left[R + 1 + \frac{K-1}{2} \right] \tag{12}$$

We note from this equation that for large K , the average delay grows linearly with K at a slope

$$\lim_{K \rightarrow \infty} \frac{\partial D}{\partial K} = \frac{1 - e^{-G}}{2e^{-G}}$$

Using Eq. (11), we see that this slope may be expressed as $G - S/2S$ which is merely the ratio of that portion of transmitted traffic which meets with a conflict to twice the throughput of the channel; since $G - S/2S = \frac{1}{2}(G/S - 1)$, we see that the limiting slope is equal to $\frac{1}{2}$ times the average number of times a packet is retransmitted. Little's well-known result¹² expresses the average number (\bar{n}) of units (packets in our case) in a queueing system as the product of the average arrival rate ($S_0 = S$ in our case) and the average time in system (D). If we use this along with Eqs. (7) and (12), we get

$$\bar{n} = SD = G \left[R + 1 + \frac{K-1}{2} \right] - S \left[\frac{K-1}{2} \right] \tag{13}$$

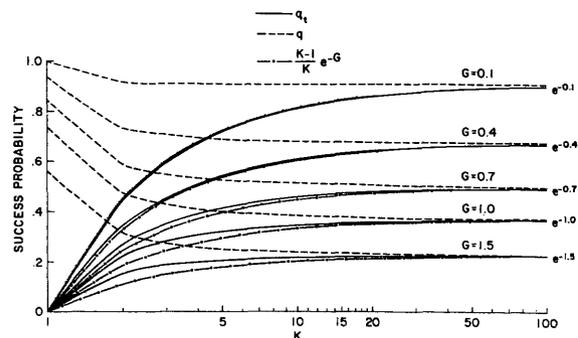


Figure 2—Success probabilities as a function of retransmission delay

In Figure 1 we plot the loci of constant delay in the S, G plane. Note the way these loci bend over sharply as K increases defining a maximum throughput $S_{max}(D)$ for any given value of D ; we note the cost in throughput if we wish to limit the average delay. This effect is clearly seen in Figure 3 which is the fundamental display of the tradeoff between delay and throughput for Model I; this figure shows the delay-throughput contours for constant values of K . We also give the minimum envelope of these contours which defines the optimum performance curve for this system (a similar optimum curve is also shown in Figure 1). Note how sharply the delay increases near the maximum throughput $S=0.368$; it is clear that an extreme price in delay must be paid if one wishes to push the channel throughput much above 0.360 and the incremental gain in throughput here is infinitesimal. On the other hand, as S approaches zero, D approaches $R+1$. Also shown here are the constant G contours. Thus this figure and Figure 1 are two alternate ways of displaying the relationship among the four critical system quantities S, G, K , and D .

From Figure 3 we observe the following effect. Consider any given value of S (say at $S=0.20$), and some given value of K (say $K=2$). We note that there are two possible values of D which satisfy these conditions ($D=21.8, D=161$). How do we explain this? It is clear that the lower value is a stable

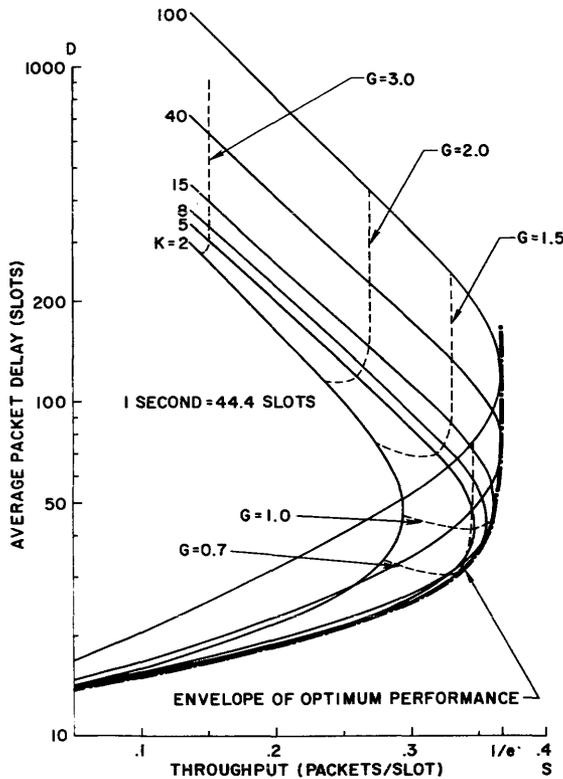


Figure 3—Delay-throughput tradeoff

* This question was raised in a private conversation with Martin Graham (University of California, Berkeley). A simulation of this situation is reported upon in Reference 13.

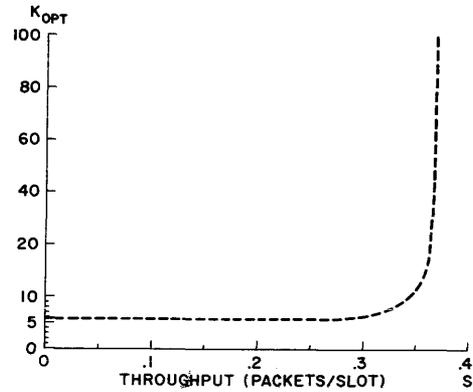


Figure 4—Optimum K

operating point since the system has sufficient capacity to absorb any fluctuation in the rate S_0 . Suppose that we now slowly increase S_0 (the source rate); so long as we do not exceed the maximum value of the system throughput rate for this K (say, $S_{max}(K)$), then we see that $S=S_0$ and the system will follow the input. Note that $S_{max}(K)$ always occurs at the intersection of the $G=1$ curve as noted earlier. However, if we attempt to set $S_0 > S_{max}(K)$, then the system will go unstable! In fact, the throughput S will drop from $S_{max}(K)$ toward zero as the system accelerates up the constant K contour toward infinite delay! The system will remain in that unfortunate circumstance so long as $S_0 > S$ (where now S is approaching zero). All during its demise, the rate at which new packets are being trapped by the system is $S_0 - S$. To recover from this situation, one can set $S_0=0$; then the delay will proceed down the K contour, round the bend at $S_{max}(K)$ and race down to $S=0$. All this while, the backlogged packets are being flushed out of the system. The warning is clear: one must avoid the knee of the K contour. Fortunately, the optimum performance curve does avoid the knee everywhere except when one attempts to squeeze out the last few percent of throughput. In Figure 4, we show the optimum values of K as a function of S . Thus, we have characterized the tradeoff between throughput and delay for Model I.

Model II. Background traffic with one large user

In this model the throughput equation is similar to that given in Eq. (7), namely,

$$S_i = G_i \frac{q_{it}}{q_{it} + 1 - q_i} \quad i = 1, 2 \quad (14)$$

the quantities q_{it} and q_i are given in the appendix. Similarly the average delays for the two classes of user are given by

$$D_1 = R + 1 + \frac{1 - q_1}{q_{1t}} \left[R + 1 + \frac{K - 1}{2} \right] \quad (15)$$

$$D_2 = R + 1 + \frac{1 - q_2}{q_{2t}} \left[R + 1 + \frac{K - 1}{2} \right] + \frac{L + 1}{2} \left[E_n + \frac{1 - q_2}{q_{2t}} E_t \right] \quad (16)$$

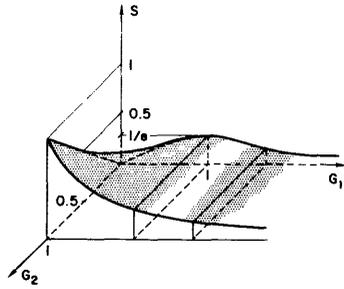


Figure 5—Throughput surface

where E_n and E_t are given in the appendix. It is easy to show that as K, L approach infinity,

$$q_1 = q_{1t} = e^{-G_1}(1 - G_2) \quad (17)$$

$$S_1 = G_1 e^{-G_1}(1 - G_2) \quad (18)$$

$$q_2 = q_{2t} = e^{-G_1} \quad (19)$$

$$S_2 = G_2 e^{-G_1} \quad (20)$$

$$S = (G - G_1 G_2) e^{-G_1} \quad (21)$$

where we recall $G = G_1 + G_2$ and $S = S_1 + S_2$. From these last equations or as given by direct arguments in an unpublished note by Roberts, one may easily show that at a constant background user throughput S_1 , the large user throughput S_2 will be maximized when

$$G = G_1 + G_2 = 1 \quad (22)$$

This last is a special case of results obtained by Abramson in Reference 8 and he discusses these limiting cases at length for various mixes of users. We note that,

$$\frac{\partial S}{\partial G_2} = e^{-G_1}(1 - G_1) \quad (23)$$

$$\frac{\partial S}{\partial G_1} = -e^{-G_1}(G - G_1 G_2 - 1 + G_2) \quad (24)$$

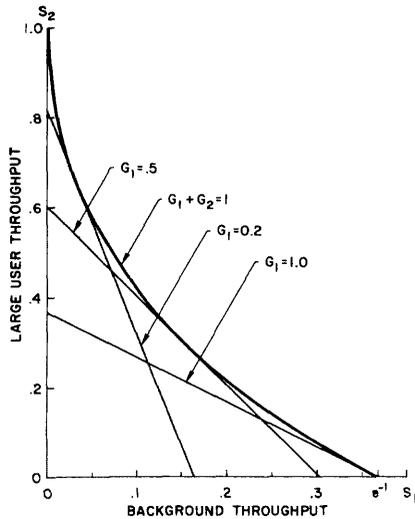


Figure 6—Throughput tradeoff

In Figure 5 we give a qualitative diagram of the 3-dimensional contour for S as a function of G_1 and G_2 . We remind the reader that this function is shown for the limiting case K, L approaching infinity only. From our results we see that for constant $G_1 < 1$, S increases linearly with G_2 ($G_2 < 1$). For constant $G_1 > 1$, S decreases linearly as G_2 increases. In addition, for constant $G_2 < 1/2$, S has a maximum value at $G_1 = 1 - 2G_2 / (1 - G_2)$. Furthermore, for constant $G_2 > 1/2$, S decreases as G_1 increases and therefore the maximum throughput S must occur at $S = G_2$ in the $G_1 = 0$ plane.

The optimum curve given in Eq. (22) is shown in the S_1, S_2 plane in Figure 6 along with the performance loci at constant G_1 . We note in these last two figures that a channel throughput equal to 1 is achievable whenever the background traffic drops to zero thereby enabling $S = S_2 = G_2 = 1$; this corresponds to the case of a single user utilizing the satellite channel at its maximum throughput of 1. Abramson [8] dis-

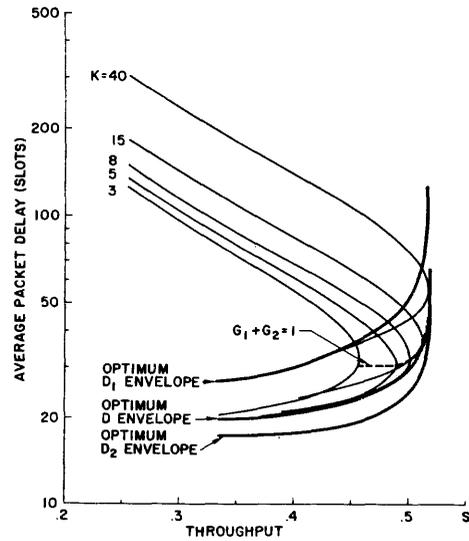


Figure 7—Delay-throughput tradeoff at $S_1 = 0.1$

usses a variety of curves such as those in Figure 6; he considers the generalization where there may be an arbitrary number of background and large users.

In the next three figures, we give numerical results for the finite K case; in all of these computations, we consider only the simplified situation in which $K = L$ thereby eliminating one parameter. In Figure 7 we show the tradeoff between delay and throughput similar to Figure 3. (Note that Figure 5 is similar to Figure 1.) Here we show the optimum performance of the average delay $D = S_1 D_1 + S_2 D_2 / S$ along with the behavior of D at constant values of K and $S_1 = 0.1$ (note the instability once again for overloaded conditions). Also shown are minimum curves for D_1 and D_2 , which are obtained by using the optimum K as a function of S . If we are willing to reduce the background throughput from its maximum at $S_1 = 0.368$, then we can drive the total throughput up to approximately $S = 0.52$ by introducing additional traffic from the large user. Note that the minimum D_1 curve is much higher than the minimum D_2 curve. Thus our net gain in

channel throughput is also at the expense of longer packet delays for the small users. Once again, we see the sharp rise near saturation.

In Figure 8, we display a family of optimum D curves for various choices of S_1 as a function of the total throughput S . We also show the behavior of Model I as given in Figure 3. Note as we reduce the background traffic, the system capacity increases slowly; however, when S_1 falls below 0.1, we begin to pick up significant gains for S_2 . Also observe that each of the constant curves "peels off" from the Model I curve at a value of $S = S_1$. At $S_1 = 0$, we have only the large user operating with no collisions and at this point, the optimal value of L is 1. This reduces to the classical queueing system with Poisson input and constant service time (denoted $M/D/1$) and represents the *absolute optimum performance* contour for any method of using the satellite channel when the input is Poisson; for other input distributions we may use the $G/D/1$ queueing results to calculate this absolute optimum performance contour.

In Figure 9, we finally show the throughput tradeoffs between the background and large users. The upper curve shows the absolute maximum S at each value of S_1 ; this is a clear display of the significant gain in S_2 which we can achieve if we are willing to reduce the background throughput. The middle curve (also shown in Figure 6 and in Reference 8) shows the absolute maximum value for S_2 at each value of S_1 . The lowest curve shows the net gain in system capacity as S_1 is reduced from its maximum possible value of $1/e$.

CONCLUSIONS

In this paper we have analyzed the performance of a slotted satellite system for packet-switching. In our first model, we have displayed the trade-off between average delay and average throughput and have shown that in the case of traffic consisting of a large number of small users, the limiting

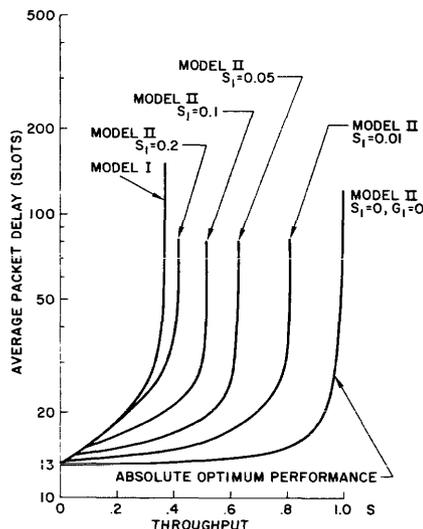


Figure 8—Optimum delay-throughput tradeoffs

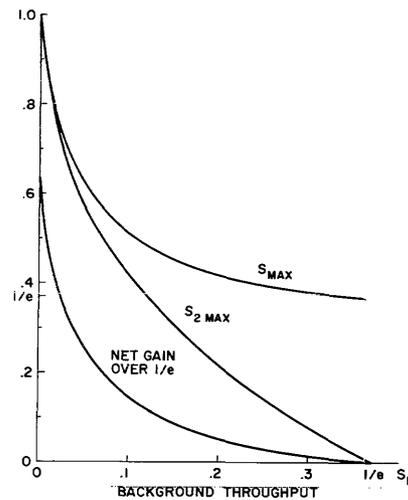


Figure 9—Throughput contours

throughput of the channel ($1/e$) can be approached fairly closely without an excessive delay. This performance can be achieved at relatively small values of K which is the random retransmission delay parameter. However, if one attempts to approach this limiting capacity, not only does one encounter large delays, but one also flirts with the hazards of unstable behavior.

In the case of a single large user mixed with the background traffic, we have shown that it is possible to increase the throughput rather significantly. The qualitative behavior for this multidimensional trade-off was shown and the numerical calculations for a given set of parameters were also displayed. The optimum mix of channel traffic was given in Eq. (22) and is commented on at length in Abramson's paper.⁸ We have been able to show in this paper the relationship between delay and throughput which is an essential trade-off in these slotted packet-switching systems.

In Roberts' paper⁹ he discusses an effective way to reserve slots in a satellite system so as to predict and prevent conflicts. It is worthwhile noting that another scheme is currently being investigated for packet-switching systems in which the propagation delay is small compared to the slot time, that is, $R = d/T \ll 1$. In such systems it may be advantageous for a user to "listen before transmitting" in order to determine if the channel is in use by some other user; such systems are referred to as "carrier sense" systems and seem to offer some interesting possibilities regarding their control. For satellite communications this case may be found when the capacity of the channel is rather small (for example, with a stationary satellite, the capacity should be in the range of 1200 bps for the packet sizes we have discussed in this paper). On the other hand, a 50 kilobit channel operating in a ground radio environment with packets on the order of 100 or 1000 bits lend themselves nicely to carrier sense techniques.

In all of these schemes one must trade off complexity of implementation with suitable performance. This performance must be effective at all ranges of traffic intensity in that no unnecessary delays or loss of throughput should occur due to

complicated operational procedures. We feel that the slotted satellite packet-switching methods described in this paper and the reservation systems for these channels described in the paper by Roberts do in fact meet these criteria.

REFERENCES

1. Jackson, P. E., Stubbs, C. D., "A Study of Multi-access Computer Communications," Spring Joint Computer Conf., *AFIPS Conf. Proc.*, Vol. 34, 1969, pp. 491-504.
2. Syski, R., *Introduction to Congestion in Telephone Systems*, Oliver & Boyd, Edinburgh and London, 1960.
3. Chu, W. W., "A Study of Asynchronous Time Division Multiplexing for Time-Sharing Computer Systems," Spring Joint Computer Conf., *AFIPS Conf. Proc.*, Vol. 35, 1969, pp. 669-678.
4. Baran, P., Boehm, S., and Smith, P., "On Distributed Communications," series of 11 reports by Rand Corp., Santa Monica, Calif., 1964.
5. Davies, D. W., "The Principles of a Data Communication Network for Computers and Remote Peripherals," *Proc. IFIP Hardware*, Edinburgh, 1968, D11.
6. Roberts, L. G., "Multiple Computer Networks and Inter-Computer Communications," *ACM Symposium on Operating Systems*, Gatlinburg, Tenn., 1967.
7. Abramson, N., "The ALOHA System—Another Alternative for Computer Communications," Fall Joint Computer Conf., *AFIPS Conf. Proc.*, Vol. 37, 1970, pp. 281-285.
8. Abramson, N., "Packet Switching with Satellites," these proceedings.
9. Roberts, L. G., "Dynamic Allocation of Satellite Capacity through Packet Reservation," these proceedings.
10. Kleinrock, L., Lam, S. S., Arpanet Satellite System Notes 12 (NIC Document #11294); 17 (NIC Document #11862); 25 (NIC Document #12734); and 27 (NIC Document #12756), available from the ARPA Network Information Center, Stanford Research Institute, Menlo Park, California.
11. Roberts, L., Arpanet Satellite System Notes 8 (NIC Document #11290) and 9 (NIC Document #11291), available from the ARPA Network Information Center, Stanford Research Institute, Menlo Park, California.
12. Kleinrock, L., *Queueing Systems: Theory and Applications*, to be published by Wiley Interscience, New York, 1973.
13. Rettberg, R., Arpanet Satellite System Note 11 (NIC Document #11293), available from the ARPA Network Information Center, Stanford Research Institute, Menlo Park, California.

APPENDIX

Define $G_s \triangleq$ Poisson arrival rate of packets to the transmitter of the large user

$$= S_2[1 + E_n + E_2(1 + E_t)] \quad (\text{A.1})$$

The variables q_i , q_{it} ($i=1, 2$) in Eqs. (14-16) are then given as follows (see Reference 10 for details of the derivations):

$$q_1 = (q_0)^K (q_h)^L e^{-S} \quad (\text{A.2})$$

$$q_{1t} = (q_0)^{K-1} q_{1c} (q_h)^L e^{-S} \quad (\text{A.3})$$

where

$$q_0 = e^{-G_1/K} + \frac{1}{K} [(1 - e^{-G_s}) (e^{-G_1} - e^{-G_1/K}) + G_1 e^{-(G_1+G_s)}] \quad (\text{A.4})$$

$$q_h = \begin{cases} (G_s + 1)e^{-G_s} & L = 1 \\ \frac{1}{L-1} (Le^{-G_s/L} - e^{-G_s}) & L \geq 2 \end{cases} \quad (\text{A.5})$$

$$q_{1c} = \frac{1}{1 - e^{-(G_1+G_s)}} \left[e^{-G_1/K} \left(1 - \frac{1 - e^{-G_s}}{K} \right) - e^{-(G_1+G_s)} \right] \quad (\text{A.6})$$

Let us introduce the following notation for events at the large user:

SS = scheduling success (capture of the transmitter)
 SC = scheduling conflict (failure to capture transmitter)
 TS = transmission success (capture of a satellite slot)
 TC = transmission conflict (conflict at the satellite)
 NP = newly generated packet

Then,

$$q_2 = \frac{r_n + r_s E_n}{1 + E_n} \quad (\text{A.7})$$

$$q_{2t} = \frac{r_t + r_s E_t}{1 + E_t} \quad (\text{A.8})$$

where

$$E_n \triangleq \text{average number of } SC \text{ events before an } SS \text{ event conditioning on } NP = \frac{1 - a_n}{a_s} \quad (\text{A.9})$$

$$E_t \triangleq \text{average number of } SC \text{ events before an } SS \text{ event conditioning on } TC = \frac{1 - a_t}{a_s} \quad (\text{A.10})$$

The variables a_i , r_i ($i=n, t, s$) are defined and given below:

$$a_n \triangleq \text{Prob } [SS/NP] = \left(\frac{q_0}{q} \right)^K (q_h)^L \left(\frac{1 - e^{-S_2}}{S_2} \right) \quad (\text{A.11})$$

$$r_n \triangleq \text{Prob } [TS/SS, NP] = q^K e^{-S_1} \quad (\text{A.12})$$

$$a_t \triangleq \text{Prob } [SS/TC] = \frac{1}{K} \frac{1 - (q_0/q)^K}{1 - q_0/q} \quad (\text{A.13})$$

$$r_t \triangleq \text{Prob } [TS/SS, TC] = q^{K-1} q_{2c} e^{-S_1} \quad (\text{A.14})$$

$$a_s \triangleq \text{Prob } [SS/SC] = \left(\frac{q_0}{q} \right)^K \frac{q_{sc}}{L} \frac{1 - (q_h)^L}{1 - q_h} \quad (\text{A.15})$$

$$r_s \triangleq \text{Prob } [TS/SS, SC] = q^K e^{-S_1} \quad (\text{A.16})$$

where

$$q = e^{-G_1/K} + \frac{G_1}{K} e^{-(G_1+G_s)} \quad (\text{A.17})$$

$$q_{2c} = \frac{e^{-G_1/K} - e^{-G_1}}{1 - e^{-G_1}} \quad (\text{A.18})$$

$$q_{sc} = \frac{1}{G_s - 1 + e^{-G_s}} \left(\frac{L}{L-1} \right)^2 \left[G_s \left(1 - \frac{1}{L} \right) e^{-G_s/L} - e^{-G_s/L} + e^{-G_s} \right] \quad (\text{A.19})$$