Simon S. Lam, *Packet Switching in a Multi-Access Broadcast Channel with Application to Satellite Communication in a Computer Network*, Ph.D. Dissertation, Computer Science Department, University of California at Los Angeles, March 1974; published as Technical Report UCLA-ENG-7429 (ARPA), School of Engineering and Applied Science, UCLA, April 1974, 306 pages.

Available in eleven .pdf files:

- TRcovers.pdf
- Abstract+ToC.pdf
- Chapter1.pdf
- Chapter2.pdf
- Chapter3.pdf
- Chapter4.pdf
- Chapter5.pdf
- Chapter6.pdf
- Chapters7-8.pdf
- Bibliography.pdf
- Appendices.pdf

CHAPTER 3

THROUGHPUT-DELAY PERFORMANCE

3.1 Introduction

In this chapter, analytic models are developed to predict the throughput-delay performance of the slotted ALOHA channel described in the last chapter. A gamut of throughput-delay tradeoffs will be presented corresponding to

- the infinite population model in which the channel supports input from a large number of small users modeled as a Poisson channel input source
- the large user model in which the channel user population consists of a large user (with buffering and scheduling capabilities) in addition to the population of small users above
- the finite population model in which the channel user population consists of a small number of large users

Small and large users may correspond to any physical devices which satisfy their abstract model descriptions given in Section 2.3.2. For example, a small user may represent a teletype console in a ground radio environment or an earth station in satellite communications as long as such a "small" user generates (independently) a new packet for transmission over the multi-access broadcast channel only after its previous packet has been successfully transmitted.

We show below that the slotted ALOHA channel capacity for the infinite population model is less than 37 percent. However, when a major fraction of the channel input is from a single large user which can buffer

and schedule its own conflicting demands, both the channel capacity and throughput-delay performance can be significantly improved. Such improvements are also possible with a channel user population consisting of a small number of large users. However, when the number of large users is ten or more, we show that the channel throughput-delay results already approximate those of an infinite population model.

Throughput-delay results in this chapter are obtained under the assumption of equilibrium conditions. Monte Carlo simulations indicate that often this assumption is valid only for some finite time period beyond which the channel goes into "saturation." This phenomenon will be characterized in Chapter 5. The possibility of unstable channel behavior was first brought up in a private conversation with Martin Graham (University of California, Berkeley).

3.2 The Infinite Population Model

3.2.1 Assumptions

An abstract model for the slotted ALOHA channel is given in Section 2.3.1. We assume here that the user population consists of a very large number of small users such that V^t, the channel input in the t^{th} slot, is an independent process and has a stationary Poisson distribution with an average of S packets/slot.

Suppose X^t is the channel traffic in the t^{th} time slot. We shall assume that during the time period of interest X^t

(1) is an independent process,

(2) is Poisson distributed, and

(3) has a stationary probability distribution.

These assumptions will be referred to as the independence assumption, the Poisson assumption, and the stationarity assumption, respectively.

We define equilibrium solutions (equilibrium points, equilibrium contour) to be those values of the channel input rate S and the channel traffic rate G such that the condition, channel throughput rate equal to the channel input rate, is satisifed. In this chapter, we shall be concerned only with equilibrium solutions. The channel is said to be in equilibrium at an equilibrium point during a period of time if the channel traffic X^t is a stationary process and the average channel traffic and throughput given by the stationary distribution of X^t satisfy the equilibrium point.

We show in Chapter 5, that slotted ALOHA channels supporting input from a large but finite number of small users are either "stable" or "unstable." For stable channels, the equilibrium throughput-delay tradeoffs given in this chapter are achievable over an infinite time horizon. On the other hand, an unstable channel will go into "saturation" after some finite time period.

Both the independence and Poisson assumptions represent approximations in our analytic model. Their accuracy will be examined by comparing analytic results with results from Monte Carlo simulations. Further tests to examine the Poisson assumption are given in Appendix A. It will also be shown in Chapter 4 that the Poisson assumption is actually implied by the independence assumption when the uniform randomization interval K is large.

3.2.2 The Analysis

Let E be the average number of retransmissions a packet incurs. Consider the time interval $[t_0,t_1]$ during which

$$\sum_{t=t_0}^{t_1} X^t = \text{total number of packet transmissions in } [t_0,t_1]$$

and

$$\sum_{t=t_0}^{t_1} \Delta(X^t) = \text{total number of successful packet transmissions in } [t_0,t_1]$$

where

$$\Delta(y) = \begin{cases} 1 & y = 1 \\ 0 & \text{otherwise} \end{cases}$$

Under the independence and stationarity assumptions, the average number of transmissions required for a packet is

$$1 + E = \lim_{(t_1 - t_0)\to\infty} \frac{\sum_{t=t_0}^{t_1} X^t/(t_1 - t_0 + 1)}{\sum_{t=t_0}^{t_1} \Delta(X^t)/(t_1 - t_0 + 1)} = \frac{G}{S_{out}}$$

For an equilibrium solution, the channel throughput rate S_{out} is equal to the channel input rate. Thus,

$$1 + E = \frac{G}{S} \tag{3.1}$$

We next define q to be the probability of success given that a packet transmission has occurred. By similar arguments to the above, we have

$$q = \frac{S}{G} = \frac{1}{1 + E} \tag{3.2}$$

The slotted ALOHA channel capacity for the infinite population model can be obtained by the following zeroth order approximation approach similar to Abramson's analysis of an unslotted ALOHA channel [ABRA 70, ROBE 72B]. Consider a test packet transmission in a channel time slot. Its probability of success is the probability that no other packet is transmitted in the same channel slot. Applying the Poisson assumption and Eq. (3.2), we have

$$q = e^{-G} \tag{3.3}$$

and

$$S = Ge^{-G} \tag{3.4}$$

Now if we differentiate Eq. (3.4) with respect to G, it can easily be shown that the maximum channel throughput rate (channel capacity) is

$$S_{max} = \frac{1}{e} \cong 0.368$$

The zeroth order approximation above disregards both the time history of the test packet and the uniform randomization interval K for retransmissions. In order to compute the average packet delay D, we shall take the following approach (to be referred to as the first order approximation).

Given a test packet, two states are distinguished depending upon its immediate history: new or previously collided. We then define

q_n = Prob[success/transmission of a new test packet]

q_t = Prob[success/transmission of a previously collided test packet]

Hence,

Prob[a packet is retransmitted exactly i times before success]

$$= (1 - q_n)(1 - q_t)^{i-1} q_t \qquad i \geq 1$$

$$E = \sum_{i=1}^{\infty} i(1 - q_n)(1 - q_t)^{i-1} q_t$$

$$= \frac{1 - q_n}{q_t} \tag{3.5}$$

and

$$q = \frac{1}{1 + E} = \frac{q_t}{q_t + 1 - q_n} \tag{3.6}$$

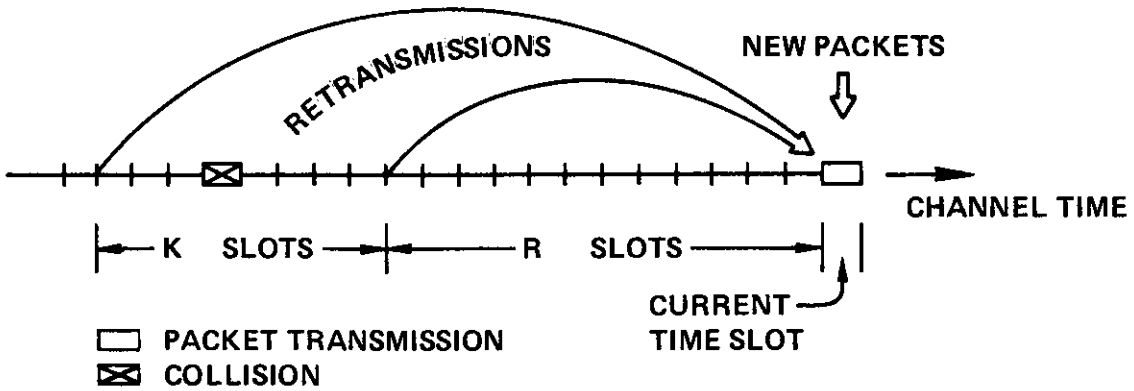

Figure 3-1. Channel Traffic Into a Time Slot.

We now condition on a test packet transmission in the current time slot. This transmission may be unsuccessful due to interference by new or previously collided packets transmitting also in the current slot (see Fig. 3-1). Suppose the test packet had a previous collision in one of the K slots (say the j^{th}) indicated in the figure. We define q_c to be the conditional probability that no packet from the j^{th} slot other than the test packet retransmits into the current slot. Using the Poisson assumption for channel traffic in each of the K slots,

$$q_c = \frac{1}{1 - e^{-G}} \left[\sum_{n=1}^{\infty} \left(\frac{K - 1}{K} \right)^n \frac{G^n}{n!} e^{-G} \right]$$

$$= \frac{e^{-\frac{G}{K}} - e^{-G}}{1 - e^{-G}}$$

Let q_0 be the probability that no packet which collided in one of the $(K - 1)$ slots (other than the j^{th} slot) retransmits in the current slot. We have

$$q_0 = \sum_{n=2}^{\infty} \left(\frac{K - 1}{K} \right)^n \frac{G^n}{n!} e^{-G} + Ge^{-G} + e^{-G}$$

$$= e^{-\frac{G}{K}} + \frac{G}{K} e^{-G}$$

Invoking the independence assumption, we then obtain

$$q_t = q_c \, q_0^{K-1} \, e^{-S}$$

$$= \frac{e^{-\frac{G}{K}} - e^{-G}}{1 - e^{-G}} \left[e^{-\frac{G}{K}} + \frac{G}{K} e^{-G} \right]^{K-1} e^{-S} \qquad (3.7)$$

Now suppose the test packet is a new packet. By similar arguments to the above, we can express its probability of success as

$$q_n = q_0^K \, e^{-S}$$

$$= \left[e^{-\frac{G}{K}} + \frac{G}{K} e^{-G} \right]^K e^{-S} \qquad (3.8)$$

From Eqs. (3.2) and (3.6), we then have

$$S = G \frac{q_t}{q_t + 1 - q_n} \qquad (3.9)$$

The <u>average delay D</u> incurred by a packet at the channel includes the channel propagation time, the packet transmission time and retransmission delays and is given below (in number of time slots) by

$$D = R + 1 + E \left(R + \frac{K + 1}{2} \right) \qquad (3.10)$$

where $R + (K + 1)/2$ is equal to the average retransmission delay (see Fig. 2-1).

Equations (3.7)-(3.9) form a set of nonlinear implicit equations which must be solved numerically for the equilibrium relationships between S and G. The average packet delay can then be obtained from Eqs. (3.5) and (3.10). Numerical solutions will be given in the next section. Below we examine some limiting cases in which explicit solutions are available and consider their implications.

Limiting results as $K \to \infty$

It can easily be shown from Eqs. (3.7)-(3.9) that in the limit as $K \to \infty$,

$$\lim_{K\to\infty} \frac{S}{G} = \lim_{K\to\infty} q_n = \lim_{K\to\infty} q_t = e^{-G} \tag{3.11}$$

These limiting results are consistent with the Poisson assumption we made. In fact, in the next chapter we show, given only the independence assumption, that in the limit as $K \to \infty$, the channel traffic in a time slot must be Poisson distributed.

Observe that Eqs. (3.11) are the same as the zeroth order approximation results. Thus, the first order approximation reduces to the zeroth order approximation in the limit as $K \to \infty$ (which corresponds to infinite average packet delay!).

Limiting results as $S \to 0$

In the limit as the channel input rate S decreases to zero, Eqs. (3.7)-(3.10) reduce to

$$\lim_{S\downarrow 0} \frac{S}{G} = \lim_{S\downarrow 0} q_n = 1 \tag{3.12}$$

$$\lim_{S\downarrow 0} q_t = \frac{K-1}{K} \tag{3.13}$$

and

$$\lim_{S \downarrow 0} D = R + 1 \tag{3.14}$$

Define K_{opt} to be the value of K minimizing the average packet delay D for a fixed channel input (throughput) rate S.

Proposition 3.1 In the limit as $S \downarrow 0$, D is convex in K and K_{opt} is given by the largest integer K such that

$$K^2 - 3K - 2R \leq 0 \tag{3.15}$$

Proof With $K = 1$, any channel collision will propagate indefinitely. Thus, $K = 1$ cannot be optimal. We shall consider $K \geq 2$. For an arbitrarily small S, Eqs. (3.7) and (3.8) become

$$q_t \cong \frac{K - 1}{K} (1 - S)$$

$$q_n \cong (1 - S)$$

and from Eqs. (3.5) and (3.10),

$$E = \frac{1 - q_n}{q_t} \cong \frac{S}{\frac{K - 1}{K} (1 - S)} \cong \frac{S}{\frac{K - 1}{K}}$$

$$D \cong R + 1 + \frac{S}{\frac{K - 1}{K}} \left[R + 1 + \frac{K - 1}{2} \right]$$

Since $S > 0$, D is minimized by minimizing the function

$$f(K) = \frac{K}{K - 1}(R + 1) + \frac{K}{2}$$

Consider

$$f(K) - f(K - 1) = -\frac{(R + 1)}{(K - 1)(K - 2)} + \frac{1}{2} \qquad K > 2$$

which is less than or equal to zero if $K^2 - 3K - 2R \leq 0$. Now consider

$$[f(K + 2) - f(K + 1)] - [f(K + 1) - f(K)]$$

$$= \frac{2(R + 1)}{(K + 1)\,K(K - 1)} > 0 \qquad K \geq 2$$

which implies that $f(K)$ is convex in K .

From the above results, D is convex in K and minimized by the largest integer K such that $K^2 - 3K - 2R \leq 0$.

Q.E.D.

For $R = 12$, $\lim_{S \downarrow 0} K_{opt} = 6$ which, as we show below, represents a lower bound on the optimum value of K for any channel input rate S .

3.2.3 Throughput-Delay Results

Numerical results

Equations (3.7)-(3.9) were solved numerically and the results plotted in Figs. 3-2 and 3-3. In Fig. 3-2, whe show the probability of success, q , as a function of K at a different channel traffic rates. For a fixed G , q increases with K and rapidly approaches its limiting value of e^{-G} as predicted by Eq. (3.11). q also increases as G decreases for a fixed K .

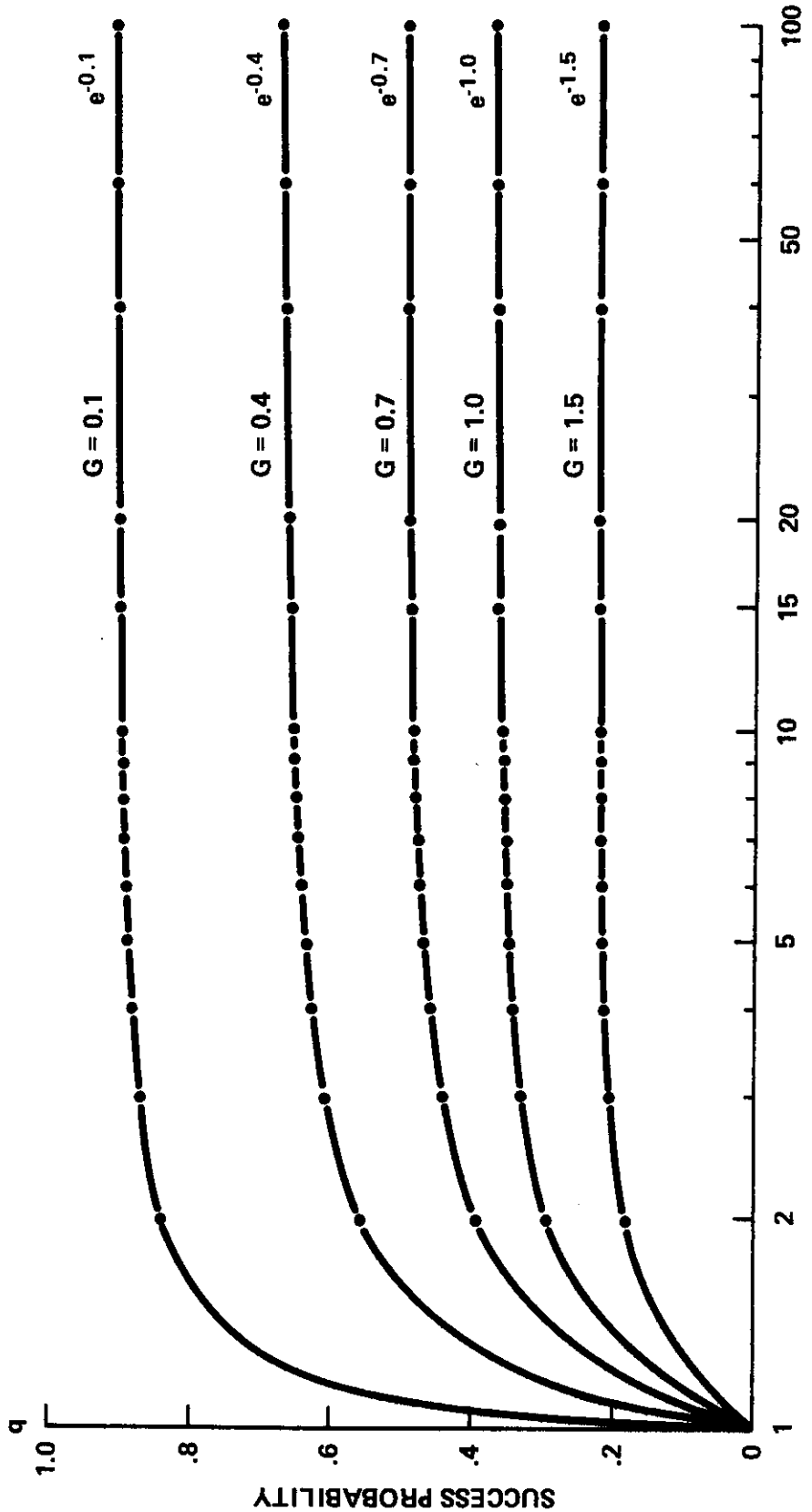

Figure 3-2. Probability of Success as a Function of K.

In Fig. 3-3, the channel throughput rate (same as the channel input rate S in an equilibrium solution) is shown as a function of G for fixed values of K. For a fixed G, the channel throughput rate increases rapidly to its limiting value of Ge^{-G} as K increases to infinity. (Note that for $K = 15$ it is almost there.) The maximum channel throughput rate occurs at $G = 1$ for each K and the channel capacity $S_{max} = e^{-1}$ in the $K \to \infty$ limit.

The average packet delay D is computed using Eq. (3.10) (and assuming $R = 12$). In Fig. 3-3, we plotted the loci of several constant delay values in the S,G plane. Note the way these loci bend over sharply defining a maximum channel throughput rate for a fixed value of D; observe the cost in channel throughput if we want to limit the average packet delay. This effect is clearly seen in Fig. 3-4, which is the fundamental display of the throughput-delay tradeoff for the infinite population model. This figure shows the throughput-delay equilibrium contours for fixed values of K. The minimum envelope of these contours defines a tight lower bound on throughput-delay performance for this system and thus, represents the optimum channel performance for the infinite population model. Considering this optimum curve, we note how sharply the average packet delay increases near the maximum channel throughput rate $S_{max} = 0.368$; it is clear that an extreme price in delay must be paid here for an infinitesimal incremental gain in throughput. Also shown in this figure are the constant G contours. Thus, Figs. 3-3 and 3-4 are two alternate displays of the relationship among the four critical system variables S, G, D and K under equilibrium conditions.

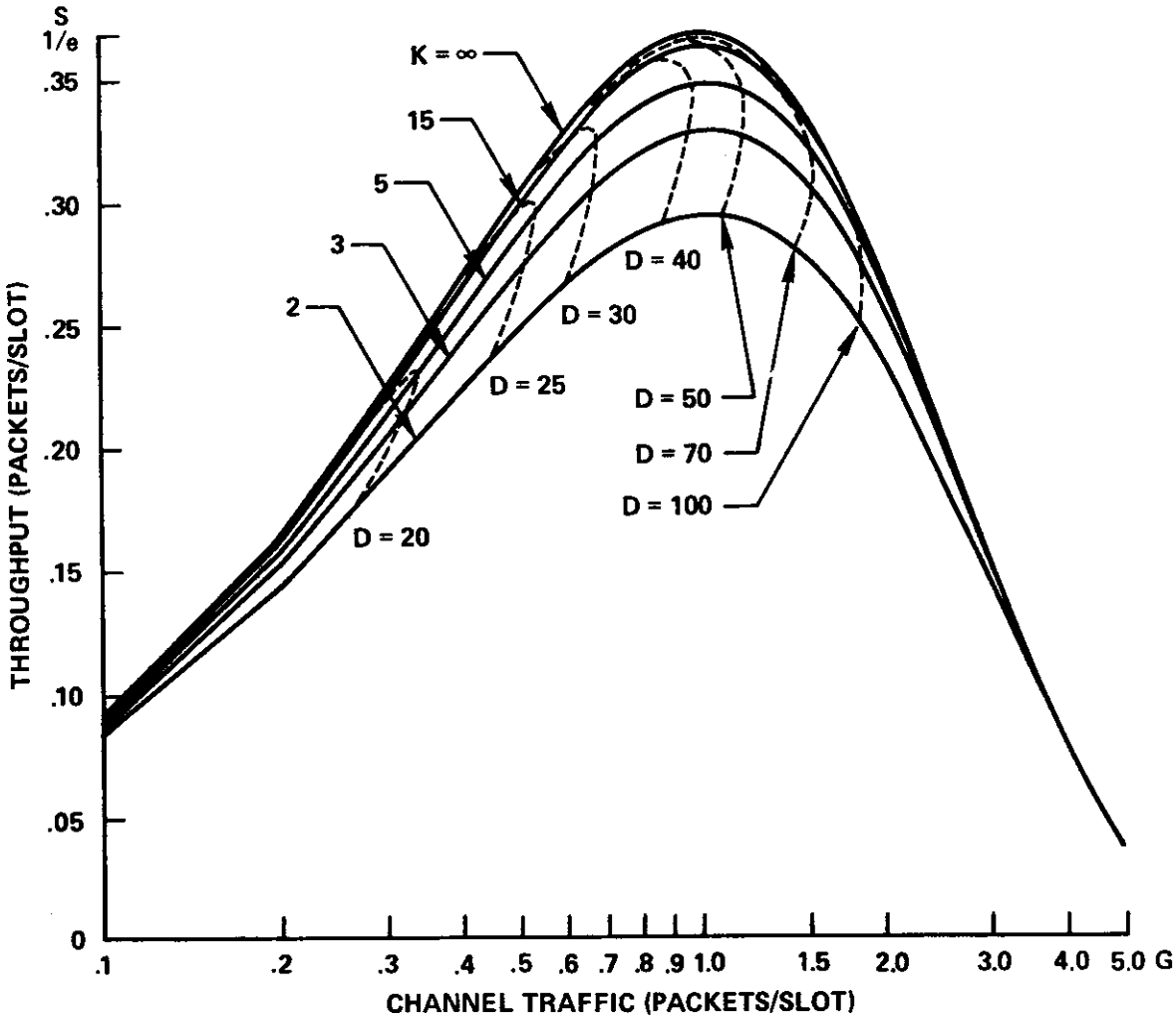

Figure 3-3. S Versus G.

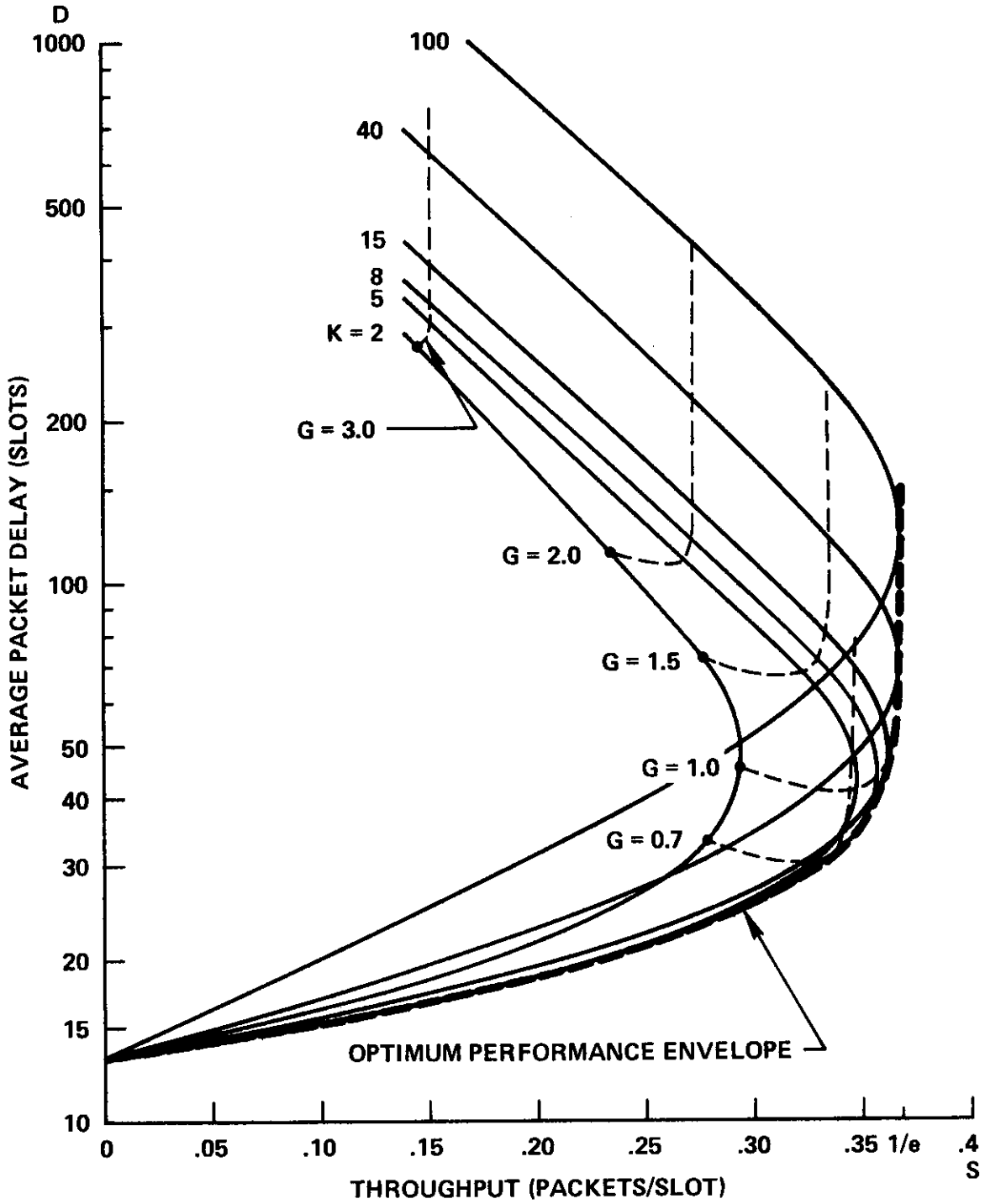

Figure 3-4. Throughput-Delay Tradeoff.

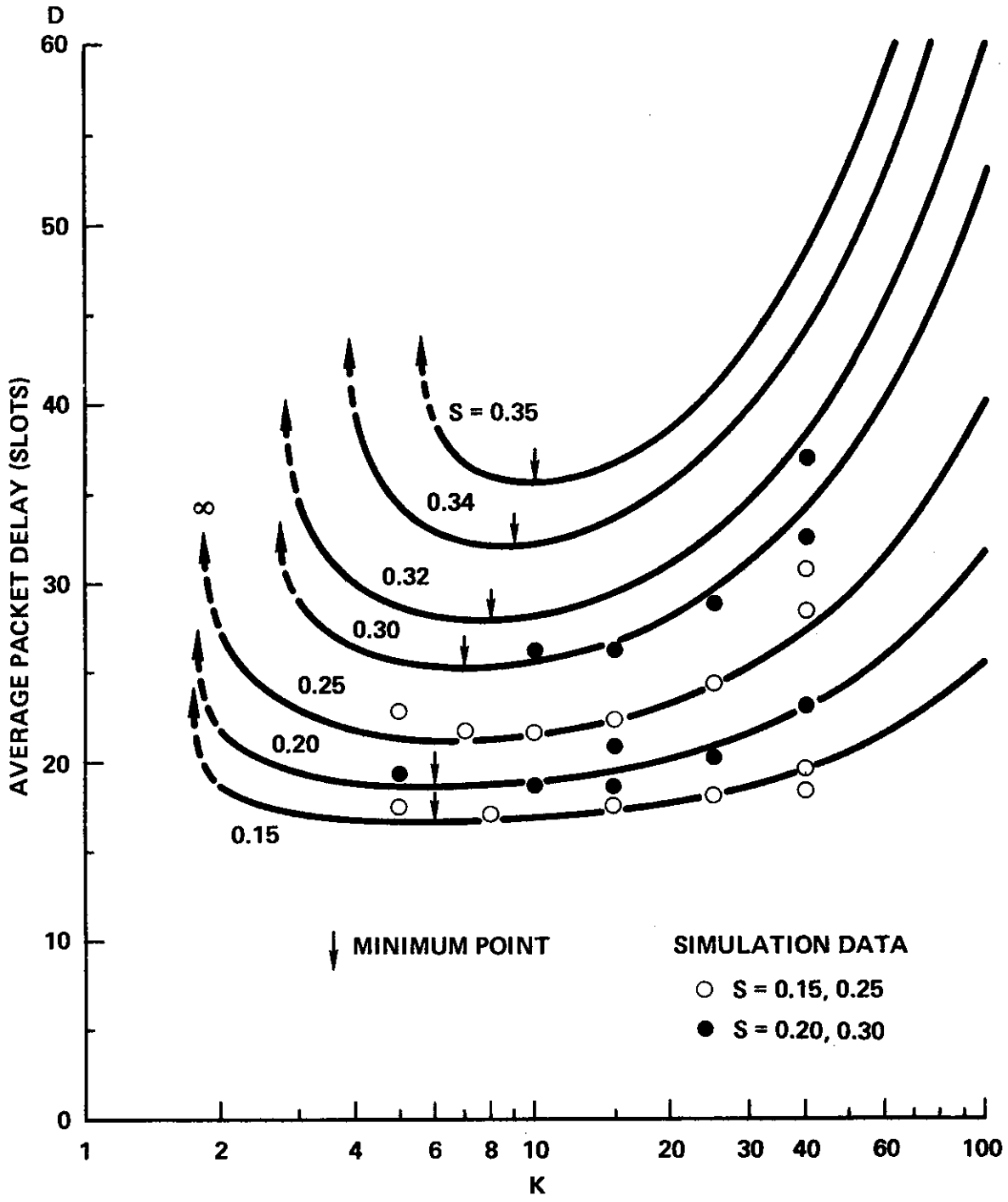

Figure 3-5. Average Packet Delay Versus K.

In Fig. 3-5, the average packet delay is plotted as a function of K for constant values of S. For a fixed S, the curve is quite flat near K_{opt}. Thus, a K value much bigger than K_{opt} can be used without increasing D appreciably. A large K is preferable since it increases the maximum channel throughput rate and improves channel stability (as we shall see in Chapter 5). In Fig. 3-6, we show K_{opt} as a function of S. Note that K_{opt} is a nondecreasing function in S and is bounded below by 6 as $S \downarrow 0$, which is predicted by Eq. (3.15) for $R = 12$.

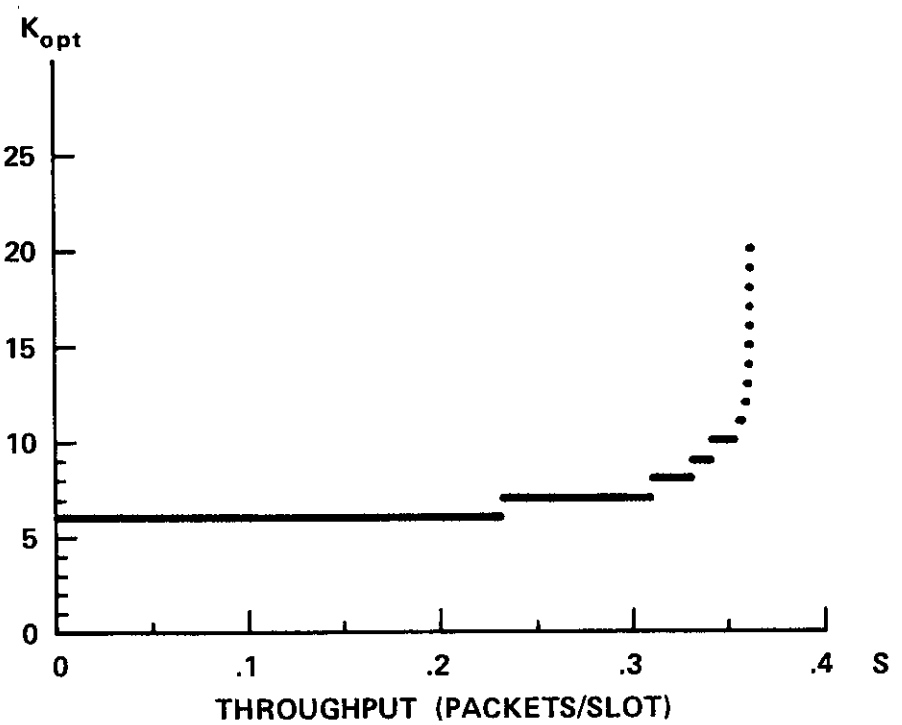

Figure 3-6. K_{opt} Versus S.

Simulations

A simulation program was developed to test the accuracy of the approximations introduced by assumptions in the above analysis. In the simulation program, new packets are generated from a Poisson distribution at a constant rate S which, together with the uniform randomization interval K, constitute the simulation input parameters. Packet delays are obtained by time-stamping each packet at the time of its creation. The exact delay a packet incurs can then be computed when it is successfully transmitted. Both long-term statistics for the duration of the simulation run and short-term statistics for consecutive time intervals (of, say, 400 slots each) are available. Short-term statistics serve to protray approximately the dynamic channel behavior.

Recall that the analytic results we have obtained so far are all based upon the assumption that the channel is in equilibrium. Referring to Fig. 3-4, we see that given S and K, there are two possible equilibrium solutions for D corresponding to a small delay value (say D_A) and a much larger delay value (say D_B). We shall refer to the equilibrium point given by S, K and D_A as the channel operating point, since this is the desired channel performance given S and K.

Each simulation run was observed to behave in the following manner. Starting from an initially empty system, the channel stays in equilibrium at the channel operating point for a finite period of time until stochastic fluctuations give rise to some high traffic rate which reduces the channel throughput rate which in turn further increases the channel traffic rate. As this vicious cycle continues,

the channel becomes "flooded" with collisions and retransmissions. The channel throughput rate vanishes rapidly to zero. This phenomenon will be referred to as channel saturation. (For the equilibrium point corresponding to D_B, no channel equilibrium for any length of time has been observed.) Thus, simulations indicate that we can assume channel equilibrium at the channel operating point, but only for some finite time period. Such time period is a random quantity and will be characterized in Chapter 5 as a measure of channel stability. The expected value of this random time period increases as K increases or S decreases. For a sufficiently small value of S or large value of K, the assumption of channel equilibrium was always valid for the simulation duration we considered. In Fig. 3-7, we show a simulation run for $S = 0.35$ and $K = 15$, which give rise to a relatively short duration of channel equilibrium. As we see in the figure, after 3000 time slots, the channel traffic rate increases very rapidly as the channel throughput rate decreases to zero.

In Fig. 3-5, simulation points are indicated. We show only those simulation runs in which the channel stays in equilibrium for the duration of the run (assumed to be 8000 slots). The (heuristic) criterion we used for channel equilibrium is that the average channel traffic in each of the short-term statistics intervals (400 slots each) must be less than one. Observe that the largest channel input rate used for these simulations is 0.3. For a larger input rate, our criterion of channel equilibrium is often not satisfied unless a very large K (say, $K = 60$) is used, which gives rise to a

SLOTTED SATELLITE COMMUNICATION SIMULATION USING POISSON DISTRIBUTED ARRIVAL GENERATION

INPUT PARAMETERS => => =>
INPUT RATE = 0.350
PROPAGATION DELAY = 12
K = 15

AVERAGE VALUES IN 200 TIME SLOT PERIODS

TIME PERIOD	THROUGHPUT RATE- S	TRAFFIC RATE- G	PACKET DELAY- D	FRACTION EMPTY	AVERAGE BACKLOG
1 - 200	0.330	0.510	17.924	0.595	3.1
201 - 400	0.370	0.605	30.892	0.530	5.3
401 - 600	0.360	0.860	32.764	0.425	9.5
601 - 800	0.340	0.840	43.147	0.435	10.5
801 - 1000	0.315	1.415	58.889	0.250	21.5
1001 - 1200	0.380	1.260	72.066	0.260	18.0
1201 - 1400	0.325	0.455	37.215	0.610	3.0
1401 - 1600	0.355	0.480	20.803	0.590	2.5
1601 - 1800	0.275	0.405	20.600	0.665	2.6
1801 - 2000	0.360	0.560	27.528	0.550	4.5
2001 - 2200	0.330	0.430	18.561	0.620	1.8
2201 - 2400	0.310	0.545	22.065	0.580	4.3
2401 - 2600	0.335	0.840	44.866	0.455	9.9
2601 - 2800	0.320	0.705	34.703	0.500	7.5
2801 - 3000	0.325	1.085	43.815	0.350	14.1
3001 - 3200	0.310	1.715	67.161	0.180	26.9
3201 - 3400	0.105	3.255	147.143	0.050	61.6
3401 - 3600	0.015	5.910	220.333	0.000	113.4
3601 - 3800	0.000	9.155	0.000	0.000	179.3
3801 - 4000	0.000	12.400	0.000	0.000	243.8

LONG TERM STATISTICS =>=>=>=>=>=>=>

THROUGHPUT = 0.270
TRAFFIC = 2.259
IDLE SLOTS = 0.371
AVERAGE DELAY = 41.287
AVERAGE BACKLOG = 38.9

TIME FOR LONG TERM STATS = 3800

Fig. 3-7 Simulation run with a short duration of channel equilibrium.

large average delay. Note that the simulation and analytic results agree very well, thus lending validity to approximations in our analysis (the independence assumption and the Poisson assumption for the channel traffic X^t). Further simulation results on the Poisson approximation are examined in Appendix A.

3.3 The Large User Model

3.3.1 The Large User Effect

The 1/e limitation on the capacity of a slotted ALOHA channel supporting input from a large number of small users (i.e., the infinite population model above) is due to the loss of all packets whenever simultaneous transmissions are made by two or more users. On the other hand, when the channel is dedicated to a single large user with buffering and scheduling capabilities, simultaneous demands from the large user's input sources can be queued up and served according to some priority rule.* In this case, a channel throughput rate arbitrarily close to unity can be achieved at the expense of a very large average packet delay. In fact, the absolute optimum throughput-delay tradeoff performance of the communication channel can be obtained by modeling it as a single server queue. Intermediate throughput-delay tradeoff performances are possible which lie between the two extremes of the infinite population model and the single server queueing model. A continuum of such intermediate tradeoff performances will be given below for the large user model in which the random access channel is shared by a large user and the small users of an infinite user

*We are only interested in the average packet delay which is independent of the exact priority rule as a result of the conservation law [KLEI 64].

population. Further intermediate tradeoff performances will be given in Section 3.4 below for a finite number of large users. In Fig. 3-8, we show a picture of the large user model in a possible satellite communications system. The large user model also represents a terminal access network in which a single radio channel is used for both terminal-to-computer and computer-to-terminal communications.

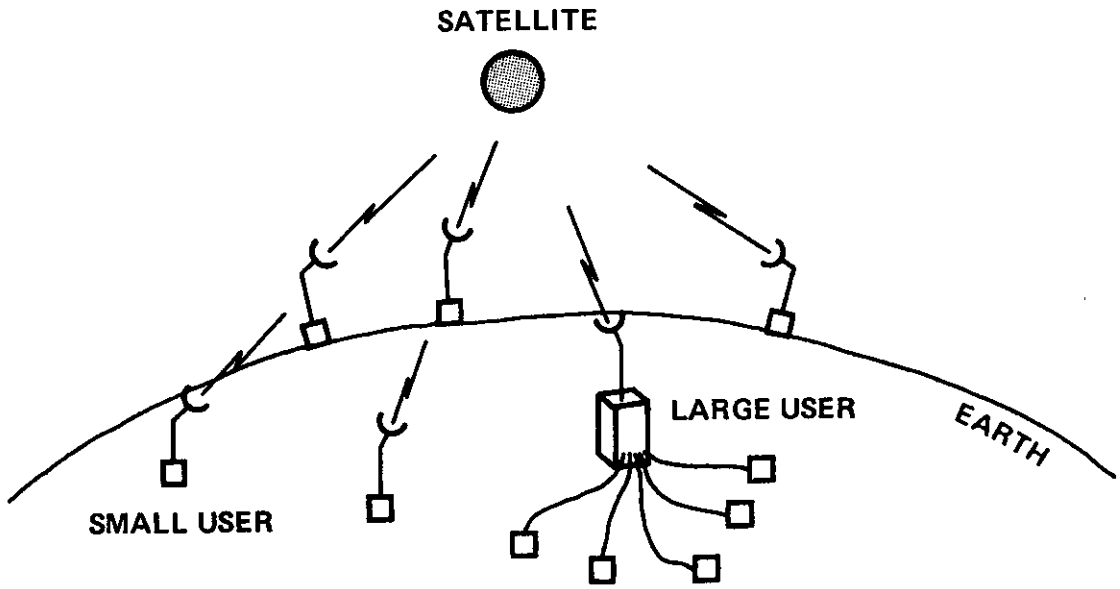

Figure 3-8. The Large User Model.

3.3.2 Throughput-Delay Results

We consider a channel user population consisting of a single large user with buffering and scheduling capabilities as described in Section 2.3.2, and a population of small users as in the infinite population model. Hence, we distinguish corresponding to the large user and the smaller users two channel input sources, both of which are assumed to be independent processes with stationary Poisson distributions. The (combined) input source to the small users

is at a rate of S_1 packets/slot. The input source to the large user is at a rate of S_2 packets/slot. The channel input rate is then given by

$$S = S_1 + S_2$$

The channel traffic in a time slot consists of packet transmissions by both the small users and large user. The large user resolves any conflict among its own packets competing for transmission in a time slot. The highest priority packet is transmitted and the rest of the competing packets are rescheduled for a later time. We define station traffic to be a random variable representing the number of packets in a time slot vying for transmission (i.e., for the transmitter) at the large user. The average station traffic is defined to be G_s packets/slot. Uniform randomization is assumed for both retransmitting packets which had a channel collision and rescheduling packets at the large user. Both station traffic and the portion of channel traffic due to the small users are assumed to be independent processes, Poisson distributed and have stationary distributions (within the time period of interest). As in the infinite population model, these assumptions represent approximations in our analytic model and will be examined by simulations.

We let G be the channel traffic rate such that

$$G = G_1 + G_2$$

where G_1 is the traffic rate due to the small users and G_2 is the traffic rate due to the large user. Since we assume that the large

user attempts no transmission in a time slot if no packet is scheduled for then (although there may be packets scheduled for a later time) and always transmits if one or more packets are scheduled for the time slot, G_2 can be interpreted as the probability that station traffic is greater than or equal to one. We must have $0 \leq G_2 \leq 1$.

We shall solve for equilibrim solutions such that the throughput rates for the small users and the large user are equal to their respective input rates S_1 and S_2. The analysis is similar to the first order approximation analysis for the infinite population model such that the effects of the uniform randomization intervals K and L are included in our model. The analytic results are summarized below. Details of the analysis are presented in Appendix B.

Similar to Eq. (3.9), equilibrium channel input rates and traffic rates are related by the following equations:

$$S_1 = G_1 \frac{q_{1t}}{q_{1t} + 1 - q_{1n}} \tag{3.16}$$

and

$$S_2 = G_2 \frac{q_{2t}}{q_{2t} + 1 - q_{2n}} \tag{3.17}$$

where q_{in} and q_{it} $(i = 1, 2)$ are the probability of success for the transmission of a new packet or a previously collided packet respectively. Note that variables indexed by 1 refer to the small users and variables indexed by 2 refer to the large user. The complete set of nonlinear implicit equations involving S_i, G_i, q_{in} and q_{it} are derived and presented in Appendix B. These equations have been solved numerically and the results are given below.

The average packet delay for the two classes of users are given by

$$D_1 = R + 1 + E_1 \left[R + \frac{K+1}{2} \right] \tag{3.18}$$

$$D_2 = R + 1 + E_2 \left[R + \frac{K+1}{2} \right] + (E_n + E_2 E_t) \frac{L+1}{2} \tag{3.19}$$

where E_1 and E_2 are the average number of retransmissions per packet for the small users and large user respectively; E_n and E_t are the number of reschedules per packet transmission at the large user conditioning on a new packet and a previously collided packet respectively. Recall that the average retransmission delay is $R + \frac{K+1}{2}$ and the average reschedule delay is $\frac{L+1}{2}$.

Limiting results

In the limit as $K, L \to \infty$, it is shown in Appendix B that our first order approximation results reduce to explicit solutions which could have been derived by direct arguments using the zeroth order approximation approach. (These results correspond to infinite average packet delay.) We have in the limit as $K, L \to \infty$,

$$q_{1n} = q_{1t} = e^{-G_1} (1 - G_2) \tag{3.20}$$

$$S_1 = G_1 e^{-G_1} (1 - G_2) \tag{3.21}$$

$$q_{2n} = q_{2t} = e^{-G_1} \tag{3.22}$$

$$S_2 = G_2 e^{-G_1} \tag{3.23}$$

$$G_2 = 1 - e^{-G_s} \tag{3.24}$$

The limiting channel throughput rate is then

$$S = (G - G_1 G_2) e^{-G_1} \tag{3.25}$$

where we recall $S = S_1 + S_2$ and $G = G_1 + G_2$. From the last equation, it can easily be shown that given either S_1 or S_2, S is maximized if the condition

$$G = G_1 + G_2 = 1$$

is satisfied. This proof was first given by L. Roberts in an unpublished note and was later generalized by Abramson [ABRA 73] to various other channel user populations. Abramson's result will be discussed in the next section.

In Fig. 3-9, we show a qualitative diagram of the 3-dimensional surface for S as a function of G_1 and G_2 (for the limiting case K, L approaching infinity). Consider the following equations:

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

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

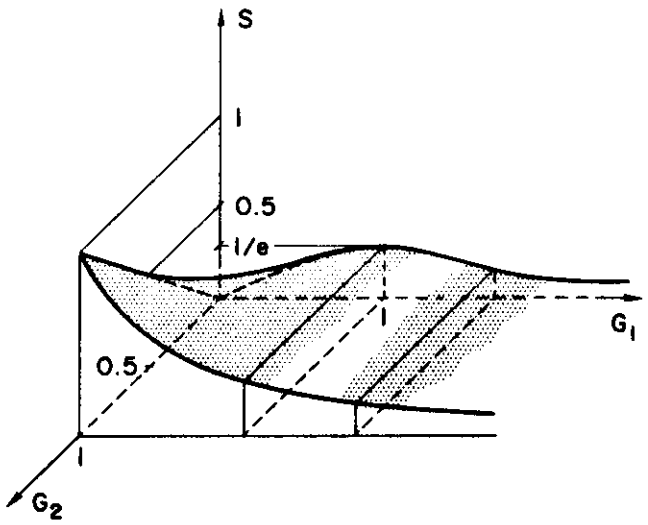

Figure 3-9. Throughput Surface.

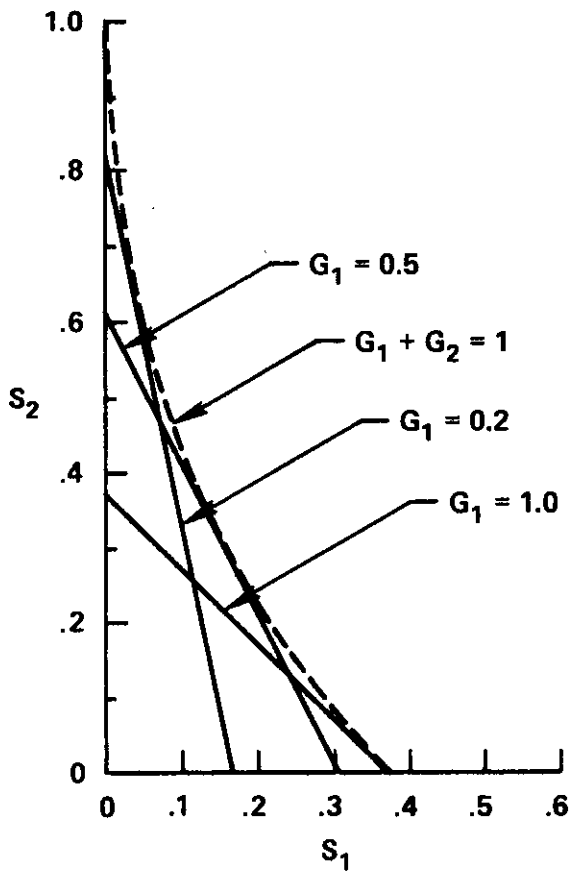

Figure 3-10. Allowable Throughput Rates for the Large User Model.

We see that for constant $G_1 < 1$, S increases linearly with G_2. For constant $G_1 > 1$, S decreases linearly as G_2 increases. In addition, for constant $G_2 < \frac{1}{2}$, S has a maximum value at $G_1 = (1 - 2G_2)/(1 - G_2)$, and for constant $G_2 > \frac{1}{2}$, S decreases as G_1 increases and the maximum throughput occurs at $S = G_2$ in the $G_1 = 0$ plane.

Numerical results

The maximum throughput contour given by letting $G = G_1 + G_2 = 1$ is shown in Fig. 3-10 along with throughput contours at constant G_1. We note in these last two figures that a channel throughput rate equal to 1 is achievable whenever G_1 (and hence, the throughput rate S_1 of the small users) drops to zero, in which case $S = S_2 = G = 1$; this then corresponds to the use of a dedicated channel.

We next present numerical results on throughput-delay tradeoffs for the finite K case; in all of these computations, we let $L = K$, thereby eliminating one parameter. In Fig. 3-11, we show the tradeoff between channel throughput rate and average packet delay for $S_1 = 0.1$, where the average packet delay D is defined to be $(S_1D_1 + S_2D_2)/S$. We show in this figure the equilibrium contours of D at constant values of K. The optimum performance envelope is given. Also shown are optimum performance envelopes for D_1 and D_2. We see that if we are willing to reduce the throughput of the small users from its maximum of $S_1 = 0.368$ to $S_1 = 0.1$, then we can drive the total throughput up to approximately $S = 0.52$ by introducing additional traffic from the large user. Note that the D_1 envelope

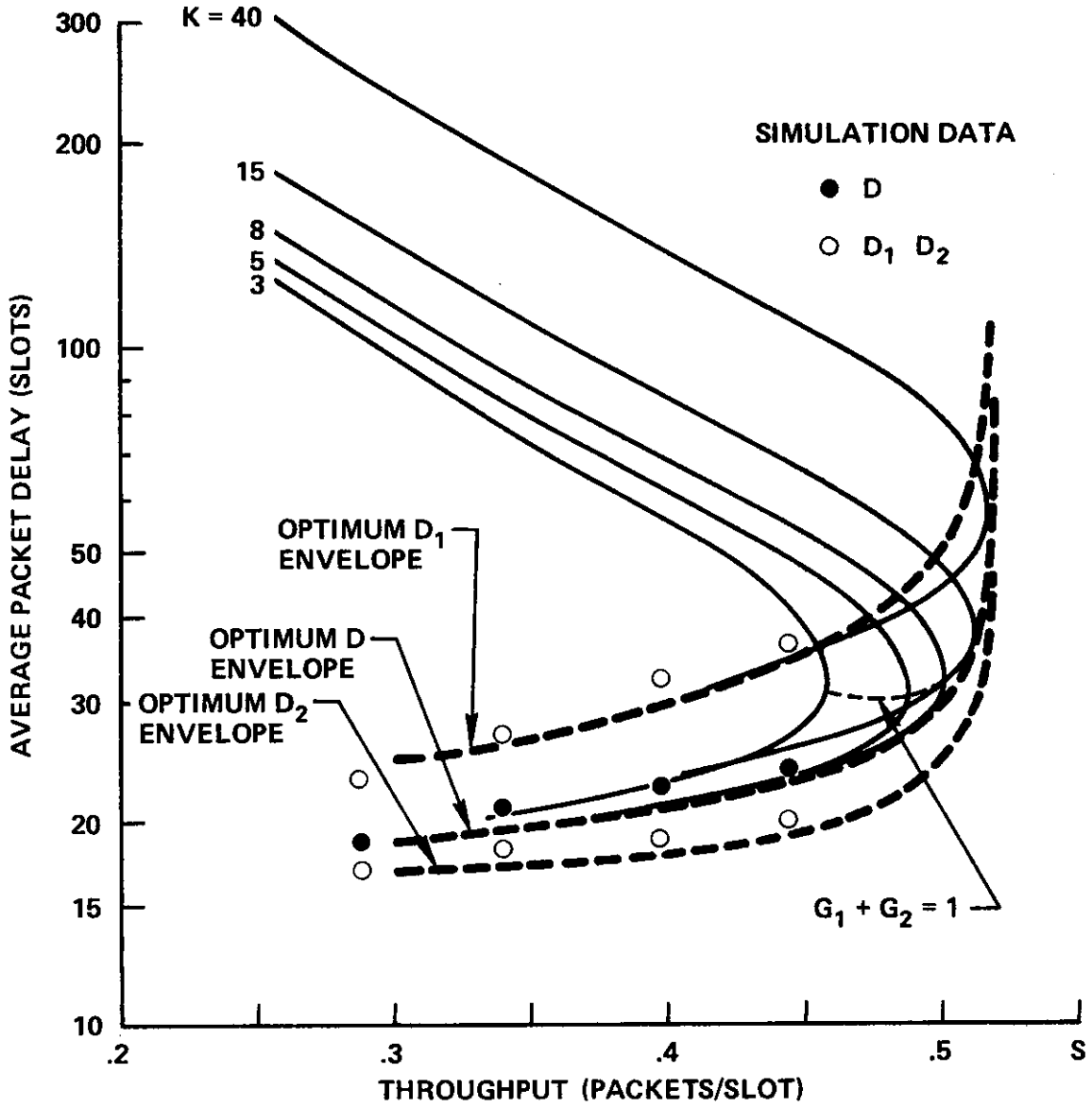

Figure 3-11. Throughput-Delay Tradeoff at $S_1 = 0.1$

is much higher than the D_2 envelope. Thus, our net gain in channel throughput is also at the expense of long delays for the small users. Once again, we note the sharp rise in average packet delay when S approaches the channel capacity.

In Fig. 3-12, we display a family of optimum throughput-delay performance envelopes for the large user model at fixed values of S_1 bounded by the optimum performance envelope of an infinite population model and that of a dedicated channel (modeled as a M/D/1 queue [KLEI 74D]). Note that 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. Also observe that each curve "peels off" from the infinite population model envelope at a value of $S = S_1$. The M/D/1 queue performance curve represents the absolute optimum performance contour for any method of using the channel when the channel input is Poisson; for input sources characterized by other probability distributions, we may use the G/D/1 queueing results to compute this absolute optimum performance contour.

Simulations

A simulation program was developed for the large user model. As in the infinite population model simulations, we found that the assumption of channel equilibrium is valid for the duration of a simulation run if a sufficiently small value of S or large value of K (and L) is used. Simulation points are indicated in Figs. 3-11 and 3-12 for those simulation runs which satisfied our channel equilibrium criterion (described earlier). The duration of each run was 5000 slots. Note that the analytic and simulation results agree very

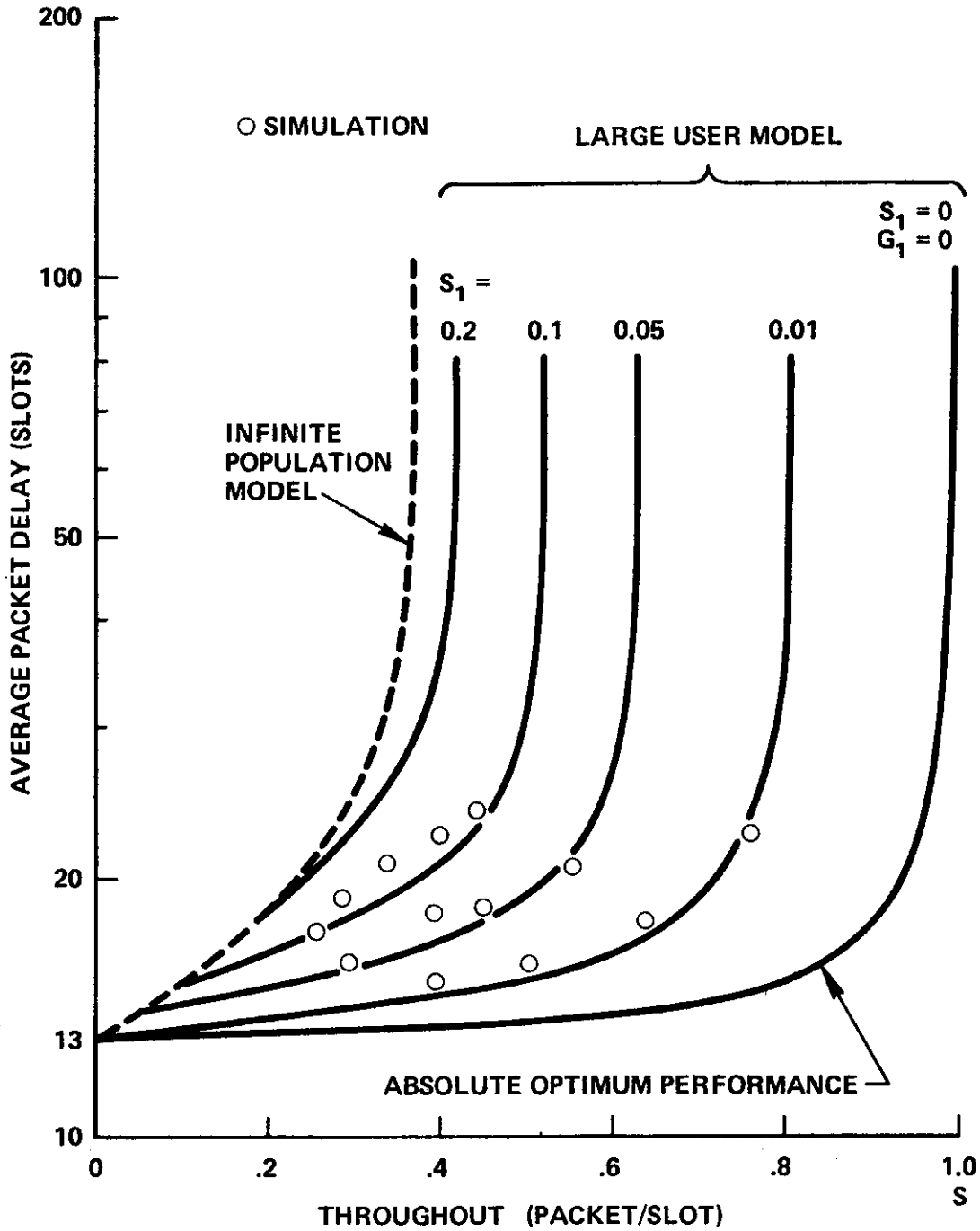

Figure 3-12. Optimum Throughput-Delay Tradeoffs.

well, thus justifying our analytic approximations. The channel input rates used in the simulations are much below the channel capacity; larger input rates can be used only with a very large K resulting in average delays much above the optimum performance envelopes.

3.4 The Finite Population Model

So far, we have considered the slotted ALOHA channel with a user population consisting of many small users modeled by a Poisson channel input. We have seen that by adding a large user with buffering and scheduling capabilities, the channel performance can be markedly improved if a significant portion of the channel input is due to the large user. In a real system, the identity of this large user may change as time progresses. Moreover, the channel user population may include more than one large user. In this case, the first order approximation approach can still be applied to solve for the throughput-delay results. However, the large number of nonlinear implicit equations that must be solved numerically renders this approach computationally unattractive. In this section, the much simpler zeroth order approximation approach is adopted and some general results are presented on the channel capacity of the finite population model. Throughput-delay tradeoffs will then be examined by simulations.

3.4.1 Channel Capacity

The results in this section were first obtained by Abramson [ABRA 73].

Given M large users with channel input rates $S_1, S_2, \ldots, S_M$ and traffic rates $G_1, G_2, \ldots, G_M$. Note that G_i corresponds to the probability of the i^{th} user transmitting in a time slot (i.e.,

the probability of having one or more packets scheduled for transmission in that time slot as discussed previously in the large user model). The equilibrium values of S_i and G_i are related by

$$S_i = G_i \prod_{j=1, j\neq i}^{M} (1 - G_j) \qquad i = 1, 2, \ldots, M \tag{3.26}$$

For any set of M acceptable traffic rates $G_1, G_2, \ldots, G_M$, these M equations define a set of input (throughput) rates $S_1, S_2, \ldots, S_M$ corresponding to a region in the M-dimensional space whose coordinates are the S_i. In order to find the boundary of this region, we calculate the Jacobian,*

$$J\left(\frac{S_1, S_2, \ldots, S_M}{G_1, G_2, \ldots, G_M}\right)$$

Since

$$\frac{\partial S_j}{\partial G_k} = \begin{cases} \displaystyle\prod_{\substack{i=1 \\ i\neq j}}^{M} (1 - G_i) & j = k \\ -G_j \displaystyle\prod_{\substack{i=1 \\ i\neq j,k}}^{M} (1 - G_i) & j \neq k \end{cases}$$

*This is the determinant of a M x M matrix whose jk^{th} element is $\frac{\partial S_j}{\partial G_k}$.

the Jacobian can be written as

$$J\left(\frac{S_1, S_2, \ldots, S_M}{G_1, G_2, \ldots, G_M}\right) = \alpha^{M-2} \begin{vmatrix} (1 - G_1) & -G_1 & -G_1 & \\ -G_2 & (1 - G_2) & -G_2 & \cdots \\ -G_3 & -G_3 & (1 - G_3) & \\ & \vdots & & \end{vmatrix}$$

$$= \alpha^{M-2} (1 - G_1 - G_2 - \ldots - G_M) \qquad (3.27)$$

where $\alpha = \prod_{j=1}^{M} (1 - G_j)$.

Equating the Jacobian to zero,* the boundary of the M-dimensional region of allowable input rates is defined by the condition

$$\sum_{i=1}^{M} G_i = 1 \qquad (3.28)$$

Examples

Consider two groups of users with M_1 users in group 1 and M_2 users in group 2 and let $M = M_1 + M_2$. Suppose $\frac{S_1}{M_1}$ and $\frac{G_1}{M_1}$ are the input and traffic rates of each user in group 1, and $\frac{S_2}{M_1}$ and $\frac{G_2}{M_2}$ are

* See Section 3.2 of [BEVE 70].

the input and traffic rates of each user in group 2. In this case, the M equations in Eqs. (3.26) become the two equations

$$S_1 = G_1\left(1 - \frac{G_1}{M_1}\right)^{M_1-1}\left(1 - \frac{G_2}{M_2}\right)^{M_2}$$
$$S_2 = G_2\left(1 - \frac{G_2}{M_2}\right)^{M_2-1}\left(1 - \frac{G_1}{M_1}\right)^{M_1} \tag{3.29}$$

which map the region of acceptable traffic rates in the (G_1, G_2) plane into the region of allowable input rates in the (S_1, S_2) plane, the boundary of which is defined by the condition

$$G_1 + G_2 = 1 \tag{3.30}$$

Substituting Eq. (3.30) into Eqs. (3.29), the resulting equation can be solved numerically for the maximum throughput contour (i.e., boundary of the allowable region of input rates) in the (S_1, S_2) plane. Several examples of such maximum throughput contours are shown in Fig. 3-13. Note that the special cases $(M_1, M_2) = (\infty, 1)$ and (∞, ∞) correspond to the large user model and the infinite population model respectively.

3.4.2 Simulation Results

A simulation program was developed for the finite population model. As in previous simulations for the infinite population model and the large user model, the assumption of channel equilibrium is valid for the duration of a simulation run if sufficiently small

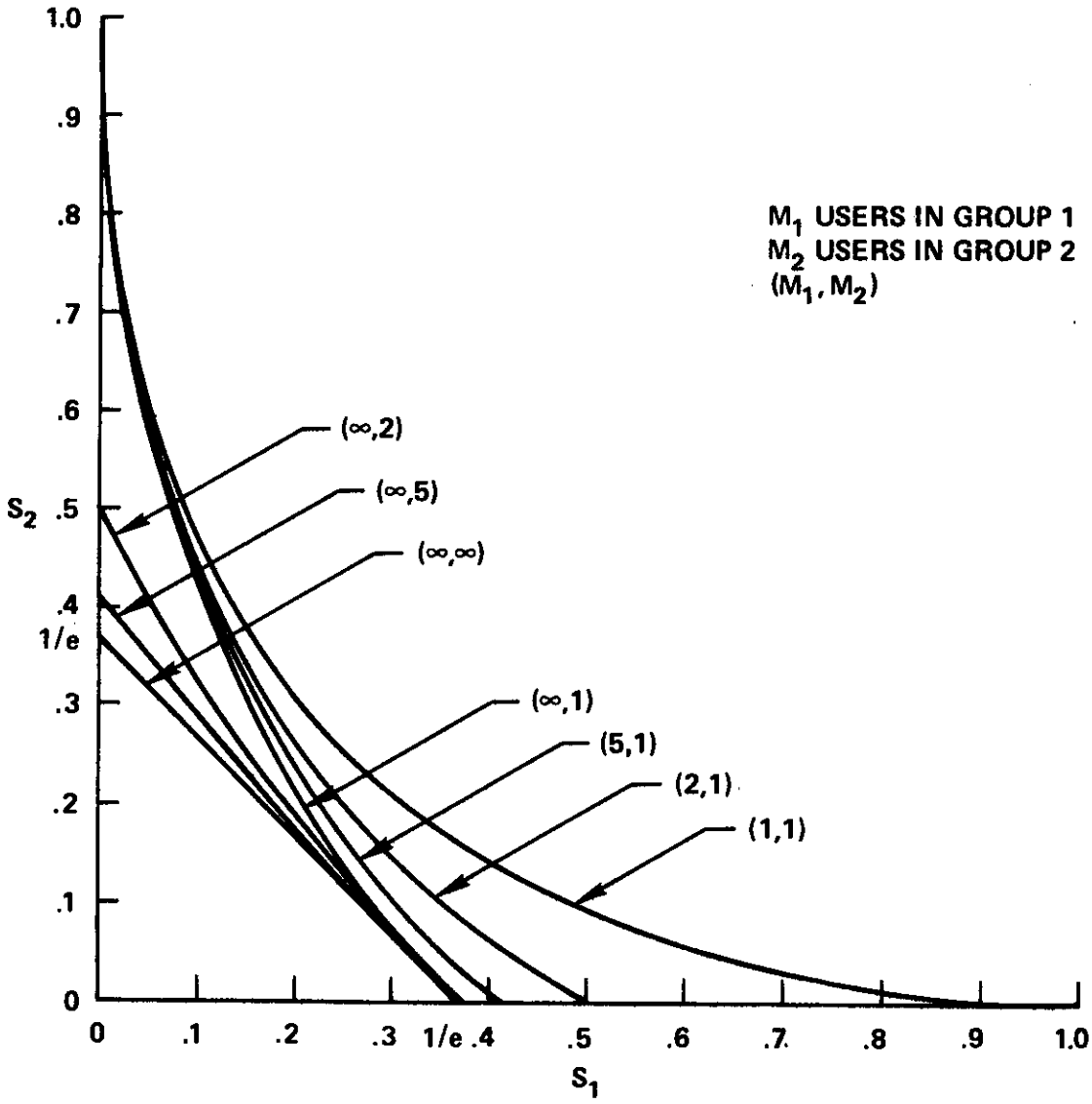

Figure 3-13. Allowable Throughput Rates for the Finite Population Model.

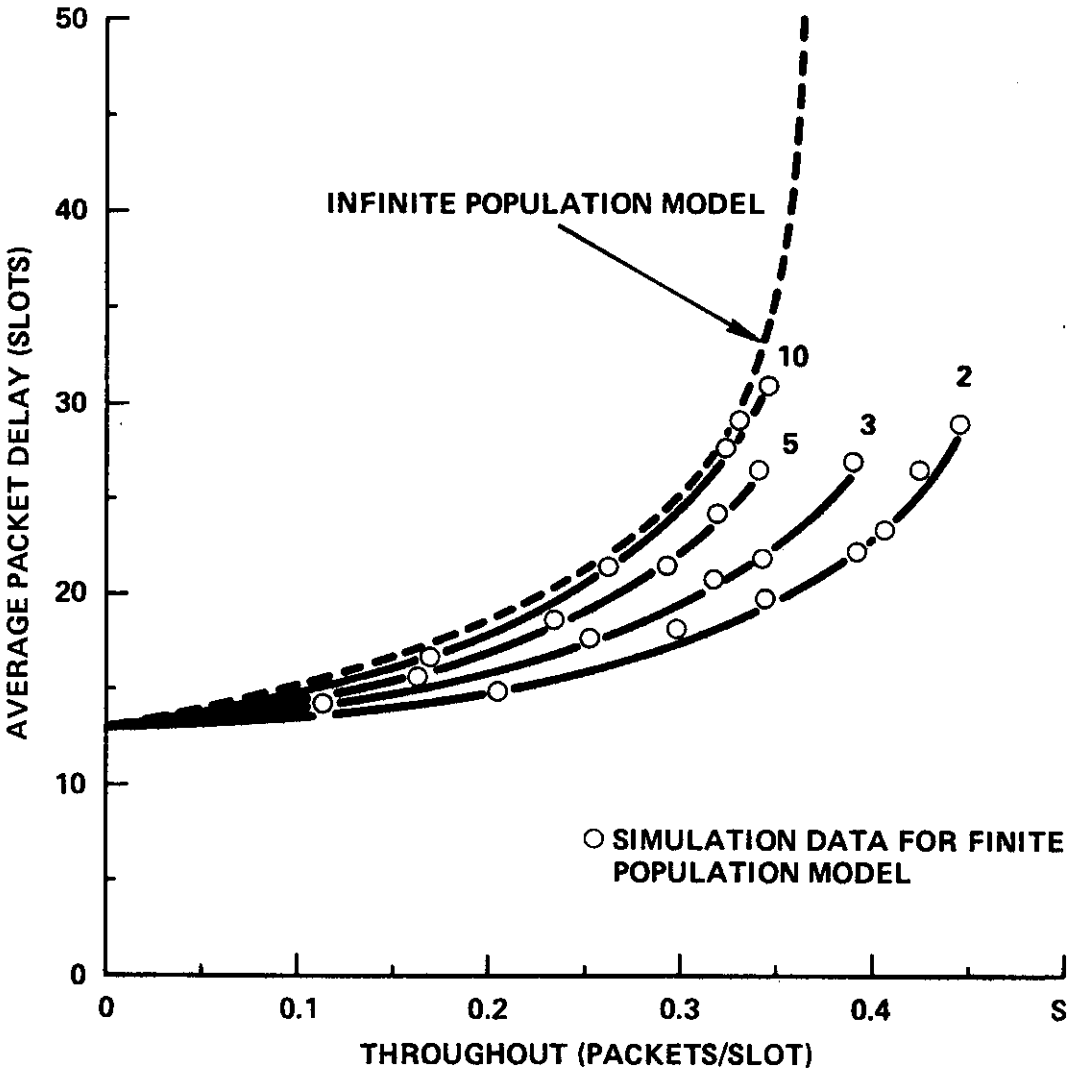

Figure 3-14. Throughput-Delay Tradeoffs for the Finite Population Model.

values of S_i or large values of K (and L) are used. We show in Fig. 3-14 throughput-delay tradeoff performances for the finite population model consisting of 2, 3, 5 and 10 large users; in each case, the channel input rate is assumed to be equally divided among the users. The infinite population model optimum envelope is also shown for comparison. Note that when the channel user population has 10 large users, the large user effect disappears and the throughput-delay results already approximate closely those of an infinite user population.