IFIP Congress
Tokyo 1980

#147

INTERFERING QUEUEING PROCESSES IN PACKET-SWITCHED BROADCAST COMMUNICATION

L. Kleinrock
Computer Science Department
School of Engineering and Applied Science
University of California
Los Angeles, CA 90024 USA

Y. Yemini
USC/Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90291 USA

We consider packet-queueing processes in two interfering buffered Packet Radio Units (PRUs) that share a Slotted ALOHA broadcast channel. It can be shown that the problem of interfering queueing cannot be solved using classical queueing theory. Here we show that classical approximation schemes, i.e., heavy- and light-traffic or diffusion approximations, fail to provide adequate predictions. Therefore a novel approximation scheme, which we call *topological approximation*, is presented. The idea is to replace approximate solutions of an exact model with an exact solution of an approximate model, obtained by perturbing the topology of interference. Finally, our analysis shows that the curse of interference, i.e., channel waste caused by collisions and/or unnecessary delays, may paradoxically be cured when the interference is increased! In fact, "maximum interference" provides a superb flow-regulating mechanism obtaining the *best* delay-throughput performance ideally possible.

1. INTRODUCTION

Consider a packet-switched store-and-forward broadcast (e.g., radio) communication network [2, 3, 4]. Packets are queued in buffered Packet Radio Units (PRUs), which act as the store-and-forward nodes, and attempt to obtain the channel according to the rules of some access scheme. Simultaneous transmissions on the channel result in collisions and effective loss of service. Analyzing the queueing behavior of the buffered packets is a typical problem of interfering (through the service mechanism) queueing processes.

Problems of interference between queueing processes arise in computer networks through the communication protocol (which conditions the activities of one process on the state of the others)* and/or through shared communication media. Interference is a fundamental mechanism for decentralized sharing of a server (e.g., a channel) and/or of load (e.g., dynamic routing). Therefore, analysis of interfering queueing processes is of prime importance to the understanding of decentralized resource-sharing mechanisms.

In some cases the communication protocol or the communication medium eliminates the dependencies between the queueing processes, and then the interaction problem decomposes into a set of simple classical queueing problems [7] (or can be approximated as a decomposable problem). However, if the queues interact properly, sharing a server and/or arrival processes, it is usually impossible and undesirable to eliminate the interaction in modelling or in practice, since this may be precisely the instrument through which distributed resource sharing is carried out.

2. THE TWO BUFFERED PRUS

Consider two PRUs communicating packets to a common destination--the *station*--over a time-slotted shared channel (see Figure 2 (1)). At each slot, packets arrive at PR_i from a Bernouili source of rate λ_i (i=1,2). The PRUs use a Slotted-ALOHA channel-access scheme [1]; that is, a busy PRU decides independently whether or not it should transmit by tossing a biased coin. It transmits if its coin shows Heads.* Let μ_i (i=1,2) designate the probability of heads on PR_i's coin. If the two PRUs decide to transmit at the same slot, then a "collision" occurs and the two transmissions destroy each other.

Let Q_i^t be the number of packets buffered at PR_i at the beginning of slot t. The queueing process $Q^t=(Q_1^t,Q_2^t)$ is a nearest-neighbor random walk (RW) on the positive quadrant of the two-dimensional integer lattice (TDRW). Figure 1 depicts the transition diagram of a general TDRW (note that in our case the probability α_0 is 0). We are interested in finding the steady-state average number of packets in the queues, $Q_i \triangleq Q_i(\lambda_1,\lambda_2,\mu_1,\mu_2)$. Using Little's result [3, 4], we can then compute the average delays $T_i=Q_i/\lambda_i$. Unfortunately, while the problem of the one-dimensional random walk is thoroughly researched, the passage to two dimensions leads to a terra incognita. The two-dimensional problem is inherently more difficult in two respects: first, the queueing processes are dependent upon each other, and second, the interaction between the behavior on the boundary

* A typical example of such interfering queueing processes is that of routing policies such as "join the shortest queue".

This research is supported by the Defense Advanced Research Projects Agency under Contract Nos. DAHC15 72 C 0308 and MDA 903 77 C 0272. Views and conclusions contained in this paper are the authors' and should not be interpreted as representing the official opinion or policy of DARPA, the U.S. Government or any person or agency connected with them.

*Note that our model of Slotted-ALOHA does not distinguish between "new" and "retransmitted" packets.

(when one queue empties) and the interior behavior of the RW may be very complex.

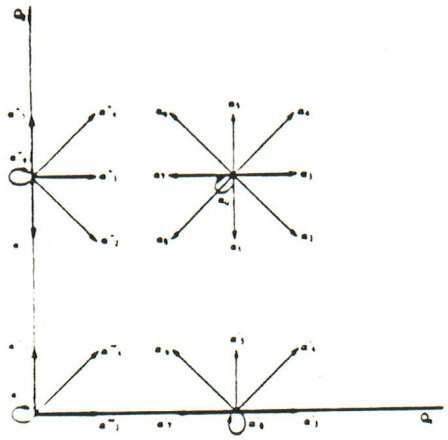

Figure 1. Transition diagram of the general TDRW

The intricacies of deriving a closed-form solution are illuminated in Appendix I. At the present there is no closed-form solution of the general TDRW problem as described above.

In the absence of a closed-form solution, an approximate solution is desirable. A natural path to adopt is to apply heavy- or light-traffic approximations [3, 4]. Unfortunately, as shown below, classical approximation schemes fail to provide an adequate approximation of the delay-throughput performance.

The failure of classical methods to provide either exact or proper approximate solutions demonstrates the need for novel methods. One possible approach is to imbed the problem in a class of problems that in some sense approximate it. This approximation can be obtained in our case by considering a set of interfering queueing problems in which the interference is gradually increased.

We consider four* models of interference for the two-buffered PRUs problem. In all four models, the two PRUs interfere with each other's service (transmission) mechanism, i.e., simultaneous transmissions at the same slot result in total loss. The increase in interference will be assumed to occur between the packet transmission and arrival mechanisms. These four models are illustrated in Figure 2.

Model 1 The two PRUs are assumed to be terminals generating new packets independently of transmissions. There is no interference other than collisions of transmissions.

Model 2 The two PRUs are assumed to be repeaters, receiving packets from a large population of terminals, which they store and forward to the station. Thus, during the transmission by a given repeater, new packets arriving from the terminals cannot be heard.

Model 3 Suppose the two repeaters in the previous model are within hearing range of each other. In addition to the previous interference, packets generated by the terminals are destroyed by the transmissions of either repeater.

Model 4 If, in addition to the interference in the previous model, we assume that all terminals are heard by both PRUs, then an attempt by any two terminals to deliver new packets to either repeater results in collision of arrivals.

Our original problem is concerned with the solution of either model 1 or 2; models 3 and 4 may be considered as approximations.

The increase of interference among the models causes a threshold behavior of the delay-throughput performance; beyond a certain level of interference (model 2 and above) *the performance of the models is identical* (see Fig. 4). As we show, the fourth ("*maximum interference*") model can be solved exactly in terms of closed-form formulae; therefore, its behavior serves as an excellent approximation of the others. We adopt the name *topological approximation* because the four models approximate each other in the sense of the topology of interference.

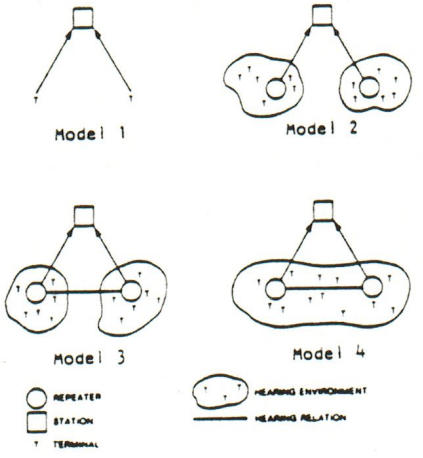

Figure 2. Four models of increasingly interfering two-buffered PRUs

*It is possible to stretch the spectrum of interference to more models; however, the four presented are sufficient to illustrate the power of topological approximation.

3. APPROXIMATE SOLUTIONS

Our point of departure in developing approximate methods is to simplify the relation between the boundary and interior behavior of the queueing RW. A reasonable simplification arises if we adjust the transitions at the boundaries or the interior so that the projections of the TDRW on the two axes perform one-dimensional RW and can be solved using classical techniques. We call such a TDRW *projectable*.

There are two extreme approximations of a TDRW in terms of a projectable TDRW. The *heavy-traffic* approximation assumes that the transitions at the boundaries (i.e., when one or both queues empty) are projections of the interior transitions (i.e., when both queues are busy). Under the heavy-traffic assumption, each PRU sees the other as a Bernoulli source of interfering noise. We ignore the details of the interaction of the two queues when any of them empties. The interaction is reduced to a constant interference. At the other extreme lies the *light-traffic* approximation that imparts the boundary behavior to the interior. Under the light-traffic approximation, interactions between the two queues are completely eliminated. It is assumed that no collisions occur. The two queues become independent.

Analysis of the heavy- and light-traffic approximations is provided in [8]. The results of the analysis were extensively compared for the four models of interaction with those of simulation and were found in general to provide a poor approximation. Figure 3 depicts typical disappointing results.

What about other approximations? It is possible to consider projectable TDRWs that are "between" the two extremes of heavy- and light-traffic. The problem with this approach is that we do not know, on the basis of analysis, which point "in between" to select. Another classical approach is to use Diffusion Approximation. Alas, the diffusion approximation provides worse results than the heavy-traffic analysis. There are two reasons. First, the diffusion approximation subsumes a heavy-traffic to start with. Second, the ability to model the interaction between the boundary and the interior is greatly reduced. The only successful existing solution [5, 6] assumes that as soon as the diffusion process hits the boundary, it is reflected orthogonally. One needs a diffusion approximation that allows sticky boundaries and reflection at other angles. This soon leads beyond the scope of simple theory of partial differential equations and the raison d'etre of the diffusion approximation (i.e., simplified closed-form solution) is lost.

The failure of classical approximation methods is disheartening; novel methods of approximation are necessary.

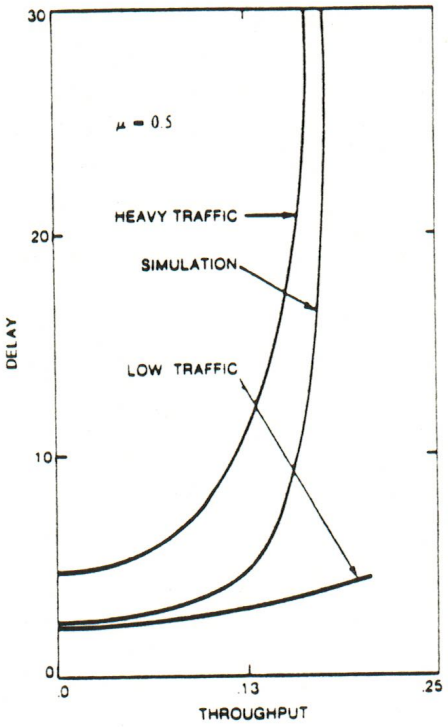

Figure 3. Approximate delay-throughput for Model 1

4. ANALYTIC SOLUTION OF THE "MAXIMUM INTERFERENCE" MODEL

One possible alternative to approximate solutions of exact models is exact solutions to approximate models. In our case we have imbedded the problem within a class of interference problems (models 1-4), each of which may be considered a topological approximation of the others. Figure 4 depicts the delay-throughput performance of the four models obtained from simulation. The performance curves of the last three models overlap each other, demonstrating the quality of topological approximation. This threshold behavior of the delay-throughput curves is of interest in itself. It is left for future research to explain this behavior analytically.

Consider the TDRW of the maximum-interference model; its transition behavior is depicted in Figure 5.* This TDRW is almost projectable; that is, the transition probabilities at each boundary are not exactly the projections of the interior movements but are projections multiplied by constants. This similarity to a projectable TDRW renders this model solvable in a simple product form.

Now consider the general steady-state transition equation of the TDRW given in Appendix I. Let us

* Henceforth we use the notation $\bar{x} \triangleq 1-x$.

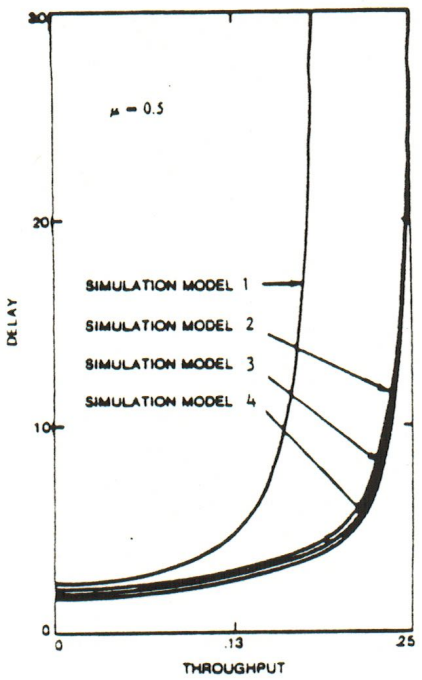

Figure 4. Threshold behavior of the delay-throughput curves as interface increases

scale the coefficients representing the boundary transitions and the respective transforms. That is, define

$$\hat{G}^{10}(w) \triangleq G^{10}(w)/\bar{\mu}_2 \quad \text{and} \quad \hat{A}^{10}(z,w) \triangleq A^{10}(z,w)\bar{\mu}_2$$

$$\hat{G}^{01}(z) \triangleq G^{01}(z)/\bar{\mu}_1 \quad \text{and} \quad \hat{A}^{01}(z,w) \triangleq A^{01}(z,w)\bar{\mu}_1$$

$$\hat{G}^{00} \triangleq G^{00}/(\bar{\mu}_1\bar{\mu}_2) \quad \text{and} \quad \hat{A}^{00}(z,w) \triangleq A^{00}(z,w)\bar{\mu}_1\bar{\mu}_2$$

The steady-state equation may then be rewritten in terms of the scaled transforms. The new form is identical to the equation for a projectable TDRW and thus may be solved. The results may be used to recover the original transform. The cumbersome computation results in the following product form expression for transforming the steady-state distribution of the fourth model*

$$G(z,w) = \frac{\left[\, 1/(1-\rho_1 z) - \mu_1 \,\right]\left[\, 1/(1-\rho_2 w) - \mu_2 \,\right]}{\left[\, 1/(1-\rho_1) - \mu_1 \,\right]\left[\, 1/(1-\rho_2) - \mu_2 \,\right]}$$

Here

$$\rho_1 \triangleq \lambda_1\bar{\lambda}_2\bar{\mu}_1/\mu_1 \quad \text{and} \quad \rho_2 \triangleq \bar{\lambda}_1\lambda_2\bar{\mu}_2/\mu_2$$

*The expression is correct when $\mu_1 \neq 1$ and $\mu_2 \neq 1$. We shall consider the case $\mu_1 = \mu_1 = 1$ separately.

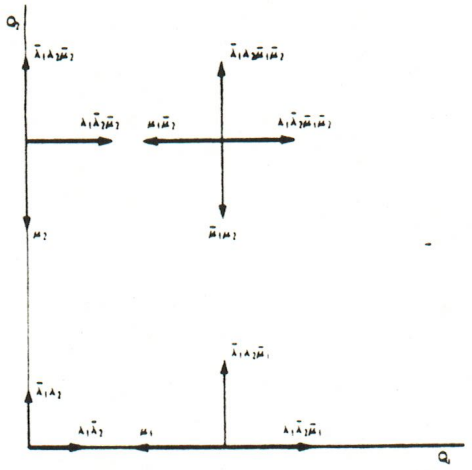

Figure 5. Transition probabilities for the maximum interference model

Let us consider the case when the two PRUs are symmetric (i.e., $\lambda_1 = \lambda_2 = \lambda$ and $\mu_1 = \mu_2 = \mu$). The expected number of packets queued in the buffer of either PRU is given by

$$\bar{Q} = \frac{\partial G}{\partial z}(1,1) = \rho/(1-\rho)(\bar{\mu}+\mu\rho)$$

When the transmission probabilities μ_i (i=1,2) are 1, the behavior of the system is simplified. The number of packets in each queue is at most one. The bivariate queueing process Q^t has 3 states whose steady-state probabilities are easily computed to be $\pi(0,0) = 1/(1+2\lambda\bar{\lambda})$, $\pi(0,1) = \pi(1,0) = \lambda\bar{\lambda}/(1+2\lambda\bar{\lambda})$. From these steady-state probabilities one can readily obtain the different performance measures.

$$Q = \lambda\bar{\lambda}/(1+2\lambda\bar{\lambda}) \qquad (\text{when } \mu = 1)$$

The overall expected throughputs (of each traffic stream) may be computed to be

$$S = \begin{cases} \lambda\bar{\lambda}/(1+\lambda\bar{\lambda})^2 & \mu < 1 \\ \lambda\bar{\lambda}/(1+2\lambda\bar{\lambda}) & \mu = 1 \end{cases}$$

Using Little's result we may compute the expected delay of a packet

$$T = \begin{cases} (1+\lambda\bar{\lambda})/(\mu-\lambda\bar{\lambda}\bar{\mu}) & \mu < 1 \\ 1 & \mu = 1 \end{cases}$$

Note that the expected delay decreases as the probability of transmission μ increases toward 1. Moreover, when the probability of transmission μ assumes the value 1, a *discontinuous improvement of performance occurs; the expected throughput exhibits a jump increase and the expected delay shows a jump decrease.* Therefore, by choosing the *rude*

transmission policy $\mu=1$ the PRUs obtain a singular improvement reaching the best possible performance.

The optimality of the rude policy is intuitively clear. Indeed, a packet entering the system is guaranteed immediate, uninterrupted service. A new packet is permitted into the system iff the system is empty and no other packet tries to enter. After entry, the expected delay is exactly one slot, and no channel waste in collisions or empty slots occurs.

The rude policy results in a *perfect synchronization* of the arrivals and transmissions. The system exhibits phased service cycles. An arriving packet is delivered from the second hop (the terminal level) to the first hop (the repeater level); then it is delivered to the station. At the end of each cycle, the system is ready for the next service cycle. Through perfect synchronization, the system obtains the best performance possible for any two-hop system, namely, delay of one slot per accepted packet (minimum possible) and maximal throughput possible (as much as the limits of interference permit).

The surprising effect is the *singularity* of the rude behavior. What is the reason for the jump in performance when the transmission probability is increased from $\mu=0.999999$ to $\mu=1$? The rude policy precludes the possibility that the two queues will ever be busy at the same time. The policy $\mu=0.999999$ renders the event "both PRUs are busy" highly improbable yet possible. However, on those rare occasions when both PRUs become busy, they will keep colliding with each other for a very long period of time, contributing significantly to the expected delay. Therefore, once those lengthy collision periods are excluded (as in the rude policy), the expected delay exhibits a discontinuous decrease.

Another surprise is the sensitivity of the two-buffered PRUs problem to small changes in the interference structure. Indeed neither the first, second, or third models even admits a rude policy. Nor is it possible to solve those models with a simple computational procedure such as the one above. A small change in the combinatorics of interference may lead to a problem that is inherently different.

Finally, let us note that the results of analysis match those of simulation so that our analytic solution of the maximum interference model is indeed a good approximation of the other models. Figure 6 compares the delay-throughput performance of the maximum interference model obtained by simulation and by analysis.

To summarize our findings:

1. Maximum interference models can be solved in terms of a simple product form solution. *This can be trivially generalized to any number of PRUs.*

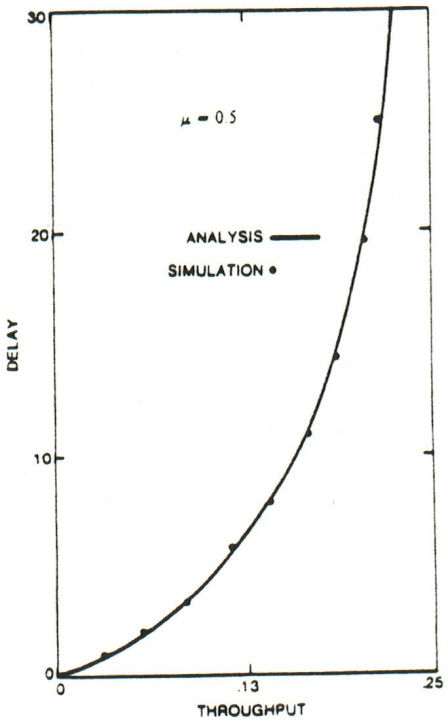

Figure 6. Delay-throughput performance of the maximum interference model simulation versus analysis

2. The delay-throughput performance of maximum-interference models can be used as an excellent approximation for lower interference models.

3. Maximum interference models admit rude policies as the optimal behavior. Such policies exhibit singular improvements of performance and obtain the best possible performance through perfect synchronization.

4. The analytic behavior of multidimensional interference systems can be very sensitive to "small" changes in the topology of interference. Nevertheless, the resulting performance may be almost invariant to these changes.

In [8], other self-synchronizing interference topologies that admit rude policies and obtain the ideally optimal performance are further explored.

Appendix I: The Steady-State Equation of a TDRW

Consider a general TDRW whose transition behavior is illustrated in Figure 1. Let $\pi(Q_1,Q_2)$, $Q_i \geq 0$, be the bivariate steady-state distribution of the numbers of queued packets (i.e., the position of the RW). We consider the following transforms

$$G^{11}(z,w) \triangleq \sum_{Q_1 \geq 1, Q_2 \geq 1} \pi(Q_1,Q_2) z^{Q_1} w^{Q_2}$$

$$G^{10}(z) \triangleq \sum_{Q_1 \geq 1} \pi(Q_1,0) z^{Q_1}$$

$$G^{01}(w) \triangleq \sum_{Q_2 \geq 1} \pi(0,Q_2) w^{Q_2}$$

$$G^{00} \triangleq \pi(0,0),$$

each of which summarizes the steady-state behavior of the TDRW at the respective region of the nonnegative quadrant.

Also, let us define the following transformed transition coefficients in terms of the transition probabilities of the respective regions:

$$A^{11}(z,w) \triangleq [w \; 1 \; 1/w] \begin{bmatrix} \alpha_6 & \alpha_5 & \alpha_4 \\ \alpha_7 & \alpha_0 - 1 & \alpha_3 \\ \alpha_8 & \alpha_1 & \alpha_2 \end{bmatrix} \begin{bmatrix} 1/z \\ 1 \\ z \end{bmatrix},$$

$$A^{10}(z,w) \triangleq [w \; 1 \; 1/w] \begin{bmatrix} \alpha_6' & \alpha_5' & \alpha_4' \\ \alpha_7' & \alpha_0' - 1 & \alpha_3' \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/z \\ 1 \\ z \end{bmatrix},$$

$$A^{01}(z,w) \triangleq [w \; 1 \; 1/w] \begin{bmatrix} 0 & \alpha_5'' & \alpha_4'' \\ 0 & \alpha_0'' - 1 & \alpha_3'' \\ 0 & \alpha_1'' & \alpha_2'' \end{bmatrix} \begin{bmatrix} 1/z \\ 1 \\ z \end{bmatrix},$$

$$A^{00}(z,w) \triangleq [w \; 1 \; 1/w] \begin{bmatrix} 0 & \alpha_5''' & \alpha_4''' \\ 0 & \alpha_0''' - 1 & \alpha_3''' \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/z \\ 1 \\ z \end{bmatrix}.$$

In terms of the above transforms and transitions coefficients, it is possible to derive [8] the following steady-state equation

$$0 = A^{11}(z,w)G^{11}(z,w) + A^{10}(z,w)G^{10}(z) +$$

$$A^{01}(z,w)G^{01}(w) + A^{00}(z,w)G^{00}$$

This is an equation for three unknown functions G^{11}, G^{10} and G^{01} and one unknown normalization factor G^{00}. These unknowns satisfy two additional conditions:

Analyticity The above functions are analytic in the respective unit (poly) disks (i.e., $\{(z,w): |z|,|w| \leq 1\}$, $\{z: |z| \leq 1\}$, and $\{w: |w| \leq 1\}$ respectively).

Normalization $G^{11}(1,1) + G^{10}(1) + G^{01}(1) + G^{00} = 1$

It is well known that the steady-state equation together with the analyticity and normalization conditions *theoretically* determines the steady-state transform uniquely. Unfortunately, when it comes to a practical solution, even relatively degenerate forms of the steady-state equation that can be solved require extremely complex mathematical instruments; a general closed-form solution is yet to be obtained [8].

REFERENCES

1. Abramson, N., "Packet switching with satellites," in *AFIPS Conference Proceedings, National Computer Conference*, pp. 695-702, 1973.

2. Kahn, R. E., "The organization of computer resources into a packet radio network," *IEEE Transactions on Communications* COM-25, (1), January 1977, 169-178.

3. Kleinrock, L., *Queueing Systems.* Volume I: *Theory*, Wiley Interscience, New York, 1975.

4. Kleinrock, L., *Queueing Systems.* Volume II: *Computer Applications*, Wiley Interscience, New York, 1976.

5. Kobayashi, H., "Application of the diffusion approximation to queueing networks I: equilibrium queue distribution," *Journal of the Association for Computing Machinery* 21, (2), 1974, 316-328.

6. Kobayashi, H., "Application of the diffusion approximation to queueing networks II: nonequilibrium distributions and applications computer modeling," *Journal of the Association for Computing Machinery* 21, (3), 1974, 459-469.

7. Baskett, F., K. M. Chandy, R. R. Muntz, and F. G. Palacios, "Open, closed, and mixed networks of queues with different classes of customers," *Journal of the Association for Computing Machinery* 22, (2), 1975, 248-260.

8. Yemini, Y., *On Channel Sharing in Discrete-Time, Multiaccess Broadcast Communication*, Ph.D. thesis, University of California, Los Angeles, 1979.