#160

NINTH INTERNATIONAL TELETRAFFIC CONGRESS MADRID, SPAIN OCTOBER 1979

ON A NEW CLASS OF QUEUEING MODELS FOR DISTRIBUTED ENVIRONMENTS*

Leonard Kleinrock
Computer Science Department
University of California, Los Angeles

ABSTRACT

A new class of queueing problems arises when one considers random demands for service which arise in a geographically distributed environment, such as access to communication channels in computer networks. Not only must we suffer the usual consequences of queues and delays due to the randomness in the demand process, but also we must pay a price for organizing these demands into a cooperating queue. It is this second problem which is usually ignored in classical queueing theory.

In this paper, we study these problems associated with geographically distributed access to a common broadcast communication channel in a packet switching environment. We present some solutions to this multi-access broadcast problem, giving the throughput-delay profile both for long-range communication systems (such as satellite packet switching) and for local access in a ground radio packet switching environment. Of interest is the optimum profile one can ever achieve; to this end, we conjecture a lower bound on the mean delay for these systems.

1. INTRODUCTION

Queueing theory is generally concerned with evaluating performance measures as seen by customers who compete for access to a server. As is well known, performance degrades even in the case when the server is able to keep up with the demands on the average due to the random arrival patterns and random service requirements. Queueing theory has devoted itself to evaluating waiting times, queue lengths, busy period durations, server utilization, throughput, etc., in terms of the arrival and service time distributions.

In recent studies involving computer-communication systems, we have now identified a new (and relatively unstudied) class of queueing problems which are extremely rich in their challenge and extremely important in their application [KLEI 77]. This class of queueing problems introduces a new source of performance degradation, namely, it considers the effect due to the *geographical distribution* of customers competing for the attention of a server. Indeed, in classical queueing theory, it has always been assumed that customers will organize themselves into a cooperating queue (e.g., first-come-first-served, last-come-first-served, priority queueing, etc.) *at no cost.* In this paper we focus on the additional cost (measured in terms of throughput and/or delay) which comes about due to the necessity to provide control information in one form or another which orders people in the queue; this includes the case when no control information is transmitted, in which case the lack of an orderly queue discipline may in itself cause a performance degradation. Basically the problem is that the geographically distributed demands for access to a server are unaware of other demands also requiring access, and so it is clear that contention will exist not only due to the random phenomena of arrivals and service times but also due to the inability to observe other queued demands. Indeed this brings us to another generalization beyond that of classical queueing theory which has to do with the method of resolving conflicts. In classical queueing theory, simultaneous demand for a server is usually handled either by delaying other customers while one of the competing requests is being served or by causing new arrivals to be "lost" when there is no storage capacity for queueing them. In both of these cases, or in systems which permit delay as well as loss, we see that one of the competing demands does indeed receive service. However in a distributed environment, it is possible (see below) that two demands actually attempt to seize the server simultaneously and thereby destructively interfere with each other resulting in the case where *neither* receives service; this corresponds to the case of a temporary loss of service to all demands. To use a homely example, let us consider the case of a coffee urn which dispenses its contents by way of a spigot. If we were to blindfold a collection of people, place coffee cups in their hands, point them in the direction of the coffee urn and ask that they attempt to receive coffee, then it is clear that occasionally more than one cup will be jammed under the open spigot at the same time resulting in more coffee on the floor than in anyone's cup; this corresponds to the case of service loss.

In summary, then, we see that these distributed contention systems contain two factors which contribute to a degradation in performance: first there are the usual queueing effects due to the random nature of the arrival and service processes; second there is the cost due to the fact that our sources are geographically distributed. If all sources were co-located (that is, if communications between them were free and instantaneous) then we could form a common queue of arrivals and achieve the optimal delay-throughput profile given by the behavior of the G/G/1 queueing system. Unfortunately we have a number of terminals which are distributed and which independently generate traffic. We are faced with controlling access to a common server (for example, a communications channel) from these distributed sources in which the control information for organizing the queue must pass over the same channel which is being controlled.

2. THE MODEL

We begin by describing a classical queueing model of the form G/G/1 (for purposes of this paper we will mostly consider only single server systems). Below we also comment on the case where we consider a finite population of M customers each of which generates a demand independent of the others. Further, we must characterize the geographical distribution of these sources. We introduce the distance metric d_{ij} which represents the distance between sources i and j. We let v be the basic system velocity (such as the speed of light for electromagnetic energy or the speed of information propagation in a communications system, etc.). As a result we see that $\tau_{ij} = d_{ij}/v$ represents the propagation time to transmit information between sources i and j. As usual we denote the average service time of a customer by $\bar{x}$ and define

$$a_{ij} = \frac{\tau_{ij}}{\bar{x}} \tag{2.1}$$

*This research was supported by the Advanced Research Projects Agency of the Department of Defense under contract MDA 903-77-C-0272.

a_{ij} represents the ratio of the propagation delay to the average service time for two sources and we denote by $\mathbf{A}$ the matrix of such values. In an environment of M distributed sources, which are attempting to share a common server, let us assume that the m^{th} such source generates demands at a rate λ_m per second. The load placed on the server by this source is therefore $\rho_m = \lambda_m \bar{x}$ and the total load placed on the server is simply

$$\rho = \sum_{m=1}^{M} \rho_m. \qquad (2.2)$$

We have now characterized this distributed access problem in terms of the following parameters: M, $\{\rho_m\}$, ρ and $\mathbf{A}$ in addition to the interarrival time distribution $A(t)$ and the service time distribution $B(x)$. We wish to calculate the loss and performance capabilities of a system with this set of parameters. If we consider two extreme cases, namely when $a_{ij} = 0$ (for all i,j) or when $M = 1$, then we no longer have a distributed problem and the cost of creating a common queue disappears. Except for these two limiting cases (and the degenerate cases where all but one of the ρ_m's go to zero) we are faced with some loss in server capacity due to the geographical separation which must be devoted to organizing these sources into a cooperating queueing structure.

3. APPLICATIONS

That which makes communications necessary is simply the fact that two or more individuals or devices wish to exchange information but are separated by some geographic distance. Thus it is natural that the applications which have generated the class of problems discussed in this paper come from a communications environment and, in particular, a computer-communications environment [KLEI 76]. Consider the case of message sources which wish to share a common communications channel of say C bits per second. A typical application is that of a radio channel in which all message sources are within range and line-of-sight of each other so that communication from any one of them may be heard by all of them. Suppose further that message generation times are random and that a given source will select its destination according to some distribution. If all message sources transmit in the same frequency band, then overlapping transmissions will destroy each other (the case of the spilled coffee); it often pays to control access to this communication channel so that at most one source transmits at a given time, thereby guaranteeing successful reception to its intended recipient (the recipient is identified by means of her name or address). One example of such a system is the citizen's band (CB) radio channel. Another example has come to be known as ground radio packet switching [KAHN 78] in which mobile digital packet radio units communicate with each other over a common broadcast channel. A third example is the case of satellite communications whereby a number of earth stations transmit digital information to a satellite which relays the information back to earth in a broad beam covering the entire set of stations [JACO 78]. A fourth example involves wire-based computer-communication networks; here we find that the cost due to geographically distributed switching mechanisms comes about due to the lack of information regarding congestion at distant parts of the network. The issues in this last application usually involve problems of routing procedures and flow control procedures to throttle and direct traffic flow over a network with delayed and inaccurate information regarding the state of congestion of that network [KLEI 78].

Perhaps the best way to describe the class of channels we are concerned with in this paper is to refer to them as *multi-access broadcast distributed* channels. They are multi-access in the sense that many sources wish to use the channel at possibly the same time. They are broadcast channels in that information transmitted from any source may be heard by many or all sources. They are distributed in the sense that the sources are distributed geographically and thereby cannot form a cooperating queue without some form of control information being provided.

4. SOME ACCESS SCHEMES

In this section we describe some known procedures for providing access to a multi-access broadcast distributed communications channel from a collection of M sources. Figure 4.1 shows the system configuration under consideration. Here we see a collection of M sources, each consisting of a user at a computer terminal which is equipped with a radio transceiver and some digital logic for implementing a channel access algorithm. The users generate data packets at unpredictable times and they wish to transmit these packets over the common broadcast channel. In much of what we discuss below, we will assume an M/D/1 queueing model in the sense that the interval of time between generation of packets to the system (that is the summed arrival process) is exponentially distributed and that each packet is of a fixed length (say of b bits). The server is the channel itself and indeed the resource being accessed is the capacity of the channel to serve packets. In Figure 4.1 we have shown the channel time axis as a sequence of slots each of which is capable of carrying exactly one packet. It is the allocation of these slots to the individual terminals which provides the service process. In that figure we have shown three of the terminals as having packets to send (one packet each - assuming unbuffered terminals) and have indicated these by drawing their antennas as black triangles. The other terminals are assumed to have no data at the present time (hollow triangles). The object, of course, is to assign each of the channel slots to exactly one of the busy terminals; this could easily be accomplished were the busy terminals to form a cooperating queue in which case they would be served one at a time and no idle slots would occur whenever traffic had to be sent. However since these terminals are distributed in space, the formation of such a queue is a non-trivial task.

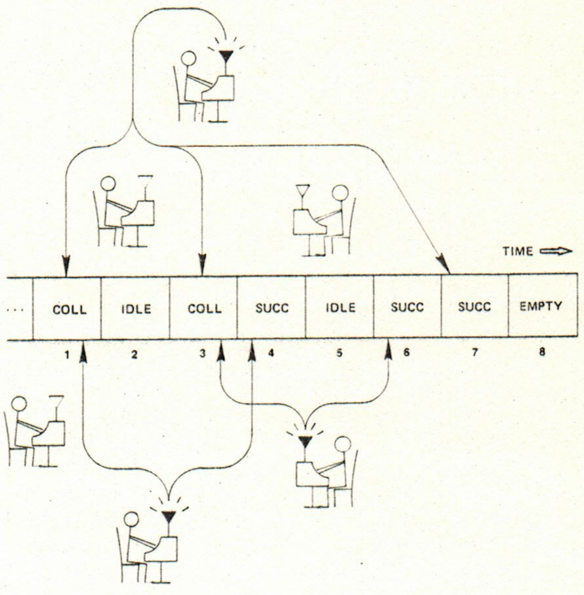

Figure 4.1 The Environment of a Multi-Access Broadcast Distributed Channel.

As a result of the behavior of any particular access scheme, we can identify four types of slots. First, there may be a "collision" slot in which more than one terminal attempts to transmit; slots number 1 and 3 are such slots. Second there may be an "idle" slot which passes unused at a time when some terminals did indeed have data to send; such slots represent wasted capacity as shown by slots number 2 and 5. Third there are "success" slots in which exactly one terminal transmits; slots 4, 6 and 7 are such. Lastly, there are "empty" slots which pass by unused at a time when no terminals require transmission; such a slot is no problem to us and an example is shown as slot number 8. The object of any access scheme is to maximize the number of success slots and to minimize the number of collision and idle slots. These are the ways in which "data" slots may be used; in addition to this there may be portions of the channel which are devoted to control information as we shall see below.

Let us define $T(\rho)$ to be the normalized average time from when a packet is generated until it is successfully received; this is the mean response time of the channel. Our main concern is the way in which the normalized average response time $T(\rho)$ varies with the overall system load ρ. $T(\rho)$ is expressed in packet transmission times for a data channel whose capacity is C bits/sec. That is, $T(\rho)$ is normalized with respect to b/C seconds. Further, since all access methods will be assumed to require the same propagation time, τ, we omit this additional delay time from all of our expressions. Thus $T(\rho) = [T_u(\rho) - \tau]C/b$ where $T_u(\rho)$ is the unnormalized average response time. Note that $\rho = \lambda b/C$.

The recent literature describes a number of multi-access schemes. Below we describe some of these schemes and for each we give a reference and an extremely concise definition. Before describing these schemes however we must recognize that we have quite a collection of choices for introducing the control information (or lack of it) which contributes to the formation of a cooperating queue in this environment. This control essentially ranges from no control at all to an extremely tight static or dynamic control. At one end of this spectrum, where no control is enforced, then more than one terminal may transmit in the same slot causing the collisions described above; such *uncontrolled* schemes are extremely simple, involve little or no control function or hardware, but extract a price from the system in the form of wasted channel capacity due to collisions. At the opposite extreme, we might introduce an extremely rigid system of *fixed control* in which each terminal is permanently assigned a portion of the overall channel for its exclusive use. Whereas such a scheme avoids collisions, it is inefficient for two reasons: first because the terminals tend to be bursty terminals and therefore much of their permanently assigned capacity may well be wasted due to their high peak-to-average ratio; and second the response time will be far worse in this channelized case due to the scaling effect [KLEI 74]. Such schemes lead to the creation of idle slots. Midway between these two extremes we find the class of *dynamic control* schemes in which a portion of the channel is set aside for control and this control is used to make reservations for data slots; this permits dynamic allocation of channel capacity according to a terminal's demand but extracts a price in the form of overhead due to the control channel. In one form or another, nature extracts her price. This price appears in the form of collisions due to poor or no control, idle (therefore wasted) slots due to rigid fixed control, or overhead due to dynamic control.

Let us now describe a number of access schemes:

PURE (UNSLOTTED) ALOHA [ABRA 73]: a newly generated packet will be transmitted by its terminal at the instant of its generation; collided packets destroy each other and must be retransmitted.

SLOTTED ALOHA [KLEI 76, KLEI 73]: the same as PURE ALOHA except that new packet transmissions must begin at the next slot point, where time is slotted into lengths equal to a packet transmission time.

CSMA (Carrier Sense Multiple Access) [KLEI 76, KLEI 75A]: same as PURE ALOHA except that a terminal senses (listens to) the channel and can hear the carrier of any other terminal's transmission; if such a carrier is detected, then the terminal refrains from transmitting and follows one of many defined protocols for deferred transmission.

POLLING [MART 70]: a central controller sends a "polling message" to each terminal in turn; when a terminal is polled, it empties all of its data before indicating its empty buffer condition whereupon the next terminal is polled in sequence.

FDMA (Frequency Division Multiple Access) [MART 70]: the bandwidth of the channel is divided into M equal sub-channels, each reserved for one of the M terminals.

TDMA (Time Division Multiple Access) [MART 70]: time is slotted and a periodic sequence of the M integers is defined such that when a terminal's number is assigned to a slot, then that terminal (and only that terminal) may transmit in that slot; typically each terminal is given one out of every M slots.

MSAP (Mini-Slotted Alternating Priority) [SCHO 76]: a carrier-sense version of polling whereby a polling sequence is defined and when the i^{th} terminal's buffer is empty, it simply refrains from transmitting; after a_{ij} (normalized) time units, the next terminal (say j) in sequence senses the channel idle and proceeds with its transmission, etc. (This is also known as hub go-ahead polling.)

URN [KLEI 78a]: The number of busy terminals (say N) is assumed to be known. k terminals are selected from the population of M and are given permission to transmit if they have packets to transmit. The optimum memoryless choice for k is $k = \lfloor M/N \rfloor$ which is the greatest integer less than or equal to M/N.

M/D/1 [KLEI 75]: the classical first-come-first-served single-server queueing system with Poisson arrivals and constant service time equal to a packet transmission time.

A number of the above-mentioned schemes have rather complicated analytic expressions representing their delay-throughput performance. The following have simple analytic expressions (we assume $\rho_m = \rho/M$ and $a_{ij} = a$ for simplicity):

$$T_{POLL}(\rho) = \frac{2-\rho}{2(1-\rho)} + \frac{a}{2}\left(1-\frac{\rho}{M}\right)\left[1 + \frac{M[2+(t_p/\tau)]}{1-\rho}\right] \quad (4.1)$$

$$T_{FDMA}(\rho) = M\left[\frac{2-\rho}{2(1-\rho)}\right] \quad (4.2)$$

$$T_{TDMA}(\rho) = 1 + M\left[\frac{1}{2} + \frac{\rho}{2(1-\rho)}\right] \quad (4.3)$$

$$T_{MSAP}(\rho) = \frac{2-\rho}{2(1-\rho)} + \frac{a}{2}\left(1 - \frac{\rho}{M}\right)\left(1 + \frac{M}{1-\rho}\right) \quad (4.4)$$

$$T_{M/D/1}(\rho) = \frac{2-\rho}{2(1-\rho)} \quad (4.5)$$

where t_p is the time to transmit a polling message. Note that $T_{MSAP}(\rho) \leqslant T_{POLL}(\rho)$.

The ALOHA schemes and CSMA correspond to uncontrolled schemes. FDMA and TDMA correspond to static control schemes. Polling, MSAP and the URN scheme correspond to dynamic control schemes. M/D/1 of course corresponds to the ideal scheme in which no price is extracted for distributed control.

If one studies the behavior of $T(\rho)$ for these various access schemes (as in [KLEI 77]) we find that no one access scheme is best for all values of load; rather, as the load changes, we find that the preferred access scheme is the pure ALOHA system at extremely light loads or is a heavily controlled scheme (such as TDMA) at loads approaching saturation. One may further note that the URN scheme has the characteristic of behaving like ALOHA in light loads and like TDMA in heavy loads and is an example of a class of schemes currently being studied in which this load-sensitive adaptive behavior is present.

5. A CONJECTURED LOWER BOUND TO THE OPTIMAL BEHAVIOR

It is clear that a system with Poisson input and deterministic service in a multi-access broadcast distributed environment (in which all terminals are within range and line-of-sight of each other) can behave no better than that of the simple M/D/1 system since we are neglecting the price one must pay to nature for the fact that the terminals are distributed geographically. One wonders how close to this over-idealized behavior one can approach. In this section we provide a conjectured lower bound to the optimal behavior which is an improvement over the M/D/1 bound. (The lower bound is a bound on the mean delay).

First we must point out that schemes such as MSAP and CSMA in which one is taking advantage of the ability to sense the state of the channel requires a further refinement in their performance evaluation. In particular it is clear that the detection of silence, upon which both of these schemes depend, is really the detection of another symbol in the alphabet of symbols transmitted. This being the case, one must not allow the parameter a to shrink below that of the time required to transmit a symbol, and so one must not accept the performance evaluation equations when $a \rightarrow 0$; rather one must then introduce an additional time for detecting the silence symbol. This being the case, then we find that none of our performance curves approach the performance of M/D/1 at all values of load.

In order to obtain a lower bound on performance, we observe that it is sufficient for a terminal to be aware of the exact number of busy terminals (say N) at the time when that terminal itself becomes busy. In such a case, the terminal will know its exact position on queue, namely it is in position $N+1$ (assuming first-come-first-served). Let us imagine that an all-knowing gremlin is available to provide this information to a terminal as soon as it becomes busy and further assume that terminals with no data to send do not eavesdrop on the channel gremlin. We define $\mathbf{P}$ to be the row vector describing the equilibrium probability for the number of busy terminals in equilibrium, that is

$$\mathbf{P} = [P_1, P_2, \cdots] \tag{5.1}$$

where P_m is the following equilibrium probability

$$P_m = P[m \textit{ terminals are busy}] \tag{5.2}$$

It is clear that on the average the amount of information which must be transmitted to a terminal is simply the entropy (say, in bits) of the distribution given above. This entropy we define as $H(\mathbf{P})$ where

$$H(\mathbf{P}) = -\sum_{m=0}^{\infty} P_m \log_2 P_m \tag{5.3}$$

We are here assuming a straightforward M/D/1 model with an infinite population of terminals. (Were we instead considering a finite population of M terminals, then the appropriate distribution to use in this entropy calculation would be that for a system containing $M-1$ terminals.) For the case M/D/1, it is well known [GROS 74] that the generating function for the equilibrium probabilities is given through

$$P(z) = \sum_{m=0}^{\infty} P_m z^m = \frac{(1-\rho)(1-z)}{1-ze^{\rho(1-z)}} \tag{5.4}$$

and the expression for P_m is given as

$$P_0 = (1-\rho)$$

$$P_1 = (1-\rho)(e^{\rho}-1)$$

$$P_m = (1-\rho)\left[\sum_{k=1}^{m} e^{k\rho}(-1)^{m-k}\frac{(k\rho)^{m-k}}{(m-k)!} + \sum_{k=1}^{m-1} e^{k\rho}(-1)^{m-k}\frac{(k\rho)^{m-k-1}}{(m-k-1)!}\right] \quad m \geqslant 2 \tag{5.5}$$

Since the arrival rate of terminals to the busy population is simply λ, then the rate at which the gremlin must provide information to the population of terminals is simply $\lambda H(\mathbf{P})$ assuming that the gremlin takes care to code the information he must transmit in the most efficient form according to Shannon's noiseless coding theorem [SHAN 49]. This operation on the part of the gremlin involves three activities. First the gremlin must observe the state of the system; we assume this is done at no cost to the system. Second, in encoding the information to be transmitted, some delay will be incurred due to the coding procedure; this too we assume costs the system nothing in terms of delay. Third the gremlin must use some of the system channel capacity in transmitting this information and it is this price which we include in order to calculate the lower bound on performance.

Now we know that the M/D/1 system incurs a delay $T(\rho)$ when the load is at the value ρ. However of this load, we now assume that some of the capacity is used for the control information transmitted by the gremlin; as a result only a portion of the load is useful data and this portion we define as ρ' where

$$\rho' = \frac{b}{b+H(\mathbf{P})}\rho \tag{5.6}$$

This last is true since each newly activated terminal will transmit b useful bits and the gremlin will be required to transmit $H(\mathbf{P})$ control bits per busy terminal. Note that $H(\mathbf{P})$ is a function of ρ. If we now charge our system for this reduction in useful throughput, we find that the lower bound for the delay-throughput profile of any access scheme is simply given as follows:

$$T_{LB}(\rho') = T_{M/D/1}(\rho) \tag{5.7}$$

Thus the behavior of the delay-throughput profile for any access scheme is lower bounded by

$$T(\rho') \geqslant T_{M/D/1}(\rho) \tag{5.8}$$

We note for $\rho \rightarrow 0$ that the lower bound approaches that of the M/D/1 curve; this is true since in this limit, the entropy of the distribution approaches zero and no capacity is lost in the transmissions due to the gremlin. Furthermore, for a finite population of terminals, it is clear that as $\rho \rightarrow 1$ then the entropy will once again approach zero and the lower bound will approach that of the pure queueing curve.

The behavior of three infinite population cases is given in Figure 5.1 (for b = 10, 100, 1000) and they are compared to the classical M/D/1 curve. Note the minimal loss when b = 1000 and the significant loss when b = 10.

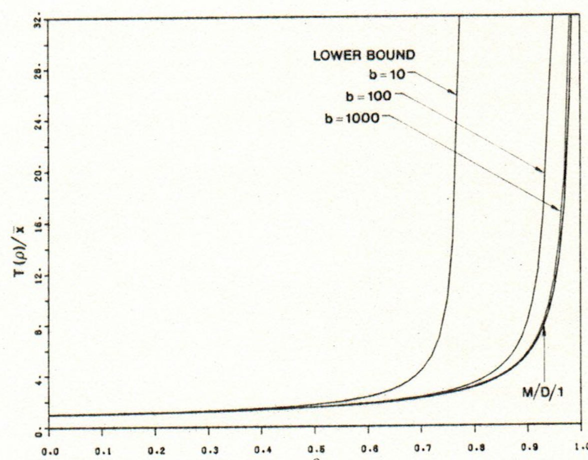

Figure 5.1 The Conjectured Lower Bound (b = 10, b = 100, b = 1000) Compared to M/D/1.

6. CONCLUSIONS AND EXTENSIONS

In this paper we have introduced a new class of queueing systems problems in which performance is degraded not only due to the random arrival and service processes but also due to the fact that the sources are geographically distributed in space and therefore require some control information to organize themselves into a cooperating queue whereby the server is used in an efficient fashion; this efficient fashion corresponds to a maximization of the number of successful slots and the minimization of the number of collided and wasted slots.

In addition we have introduced a conjectured lower bound on the delay-throughput profile for any access scheme and have expressed it in terms of the entropy of the underlying distribution of busy terminals.

Throughout this paper we have assumed that all terminals are within range and in line-of-sight of each other. This implied that all terminals would hear (and therefore also be interfered by) all other terminals' transmissions. If this assumption is false, as for example in the case where there are objects which are opaque to radio signals separating the terminals or where the range of the terminals is restricted (by choice or by power limitations) to be less than that required to reach all terminals, then a number of other considerations arise. For example one must then concern oneself with the way in which intermediate terminals should relay traffic destined for distant terminals; this corresponds to what is known as the multi-hop problem and involves routing procedures, more complex distributed control procedures, and search and location problems. On the other hand there is an advantage to a restricted range situation since then one may "spatially reuse" the frequency of the communications channel in the sense that more than one transmission may take place simultaneously and still be successful if the transmissions do not overlap at the receivers. Questions regarding the spatial capacity of such systems have been addressed in [KLEI 78b].

The class of problems introduced in this paper are rich in their difficulty and application to important packet switching systems. It is hoped that this introduction will inspire queueing theorists to apply their expertise in obtaining solutions to some of these problems.

ACKNOWLEDGEMENTS

I would like to thank Frank Heath, George Ann Hornor and Lou Nelson for their help in the preparation of this manuscript.

REFERENCES

[ABRA 73] Abramson, N., "Packet Switching with Satellites," *AFIPS Conference Proceedings,* National Computer Conference, **42,** pp. 711-716, 1973.

[GROS 74] Gross, D. and C.M. Harris, *Fundamentals of Queueing Theory,* John Wiley & Sons, New York, 1974.

[JACO 78] Jacobs, I.M., R. Binder and E.V. Hoversten, "General Purpose Packet Satellite Networks", *Proceedings of the IEEE,* **66,** pp. 1448-1467, November 1978.

[KAHN 78] Kahn, R.E., S.A. Gronemeyer, J. Burchfiel and R.C. Kunzelman, "Advances in Packet Radio Technology," *Proceedings of the IEEE,* **66,** pp. 1468-1496, November 1978.

[KLEI 73] Kleinrock, L. and S. Lam, "Packet-Switching in a Slotted Satellite Channel" *AFIPS Conference Proceedings,* National Computer Conference, **42,** pp. 703-710, 1973.

[KLEI 74] Kleinrock, L., "Resource Allocation in Computer Systems and Computer-Communication Networks," *Information Processing 74,* pp. 11-18, August 1974.

[KLEI 75] Kleinrock, L., *Queueing Systems, Volume I: Theory,* Wiley Interscience, New York, 1975.

[KLEI 75a] Kleinrock, L. and F. Tobagi, "Random Access Techniques for Data Transmission Over Packet-Switched Radio Channels," *AFIPS Conference Proceedings,* National Computer Conference, **44,** pp. 187-201, 1975.

[KLEI 76] Kleinrock, L., *Queueing Systems, Volume II: Computer Applications,* Wiley Interscience, New York, 1976.

[KLEI 77] Kleinrock, L. "Performance of Distributed Multi-Access Computer-Communication Systems, *Information Processing 77,* pp. 547-552, August 1977.

[KLEI 78] Kleinrock, L., "Principles and Lessons in Packet Communications," *Proceedings of the IEEE,* **66,** pp. 1320-1329, November 1978.

[KLEI 78a] Kleinrock, L. and Y. Yemini, "An Optimal Adaptive Scheme for Multiple Access Broadcast Communication," *Conference Record, International Conference on Communications,* pp. 7.2.1-7.2.5, June 1978.

[KLEI 78b] Kleinrock, L. and J. Silvester, "Optimum Transmission Radii for Packet Radio Networks," *Conference Record, National Telecommunications Conference,* pp. 4.3.1-4.3.5, December 1978.

[MART 70] Martin, J., *Teleprocessing Network Organization,* Prentice-Hall, Englewood Cliffs, New Jersey, 1970.

[SCHO 76] Scholl, M., "Multiplexing Techniques for Data Transmission over Packet-Switched Radio Systems," Computer Science Department, University of California, Los Angeles, UCLA-ENG-76123, December 1976. (Also published as Ph.D. dissertation.)

[SHAN 49] Shannon, C.E. and W. Weaver, *The Mathematical Theory of Communication,* The University of Illinois Press, Urbana, Illinois, 1949.