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 7

MULTI-PACKET MESSAGE DELAY AND SATELLITE RESERVATION SCHEMES

In a packet switched network, "messages" generated by external sources for transmission over the network are broken into fixed size packets. Up to now we have assumed that all messages generated are fixed length single packets. We have also used the average packet delay as our channel performance measure. This assumption is indeed justified in an interactive computer communications environment. Measurement results [KLEI 74B] indicate that 96 percent of the ARPA network traffic consists of messages shorter than a single packet. However, there are situations in which the average packet delay is not an appropriate channel performance measure: for example, the transfer of long data files and the transmission of digital voice messages [BAYL 73]. In these cases, a more appropriate performance measure may be the average message delay, namely, the delay incurred by a message from the time it is ready for transmission until when all packets in the message have been correctly received at the message destination.

A satellite reservation system has been studied for multi-packet message arrivals by Roberts [ROBE 73]. In this system, the satellite channel is dynamically partitioned into a slotted ALOHA channel for broadcasting reservation requests and a scheduled channel for transmitting multi-packet blocks of data. Since the minimum delay in this case is two satellite round-trip propagation times

($\cong$ 0.5 sec.), this system is preferable if a significant fraction of the channel input source consists of multi-packet messages and if the average message delay is the relevant channel performance measure.

In this chapter, we first derive an approximate formula for the average message delay in a slotted ALOHA channel. Next, Roberts' reservation system will be introduced. Two other satellite reservation schemes will also be described; these two schemes may be used if there is only a small number of channel users and if the channel input source has constant as well as random components. The reservation schemes, by reducing the amount of channel collisions, are capable of providing channel throughput rates well in excess of the slotted ALOHA channel capacity.

7.1 Multi-Packet Message Delay

In this section, we consider a slotted ALOHA model such as the infinite population model in Chapter 3. However, each arrival is now a message of $\tilde{L}$ packets (where $\tilde{L}$ is an integer-valued random variable specified by some probability distribution). A good approximation for the message delay is the delay incurred by the packet in the message with the most number of retransmissions. Define

p_n = Prob[collision/transmission of a new packet]

p_t = Prob[collision/transmission of a previously collided packet]

Thus, $p_n = 1 - q_n$ and $p_t = 1 - q_t$ where q_n and q_t are specified by S and K in Chapter 3. Let C be a random variable representing the number of channel collisions a packet incurs before its successful transmission. We then have

$$\text{Prob}[C = i] = \begin{cases} 1 - p_n & i = 0 \\ p_n p_t^{\,i-1}(1 - p_t) & i \geq 1 \end{cases} \tag{7.1}$$

$$\text{Prob}[C \leq i] = 1 - p_n p_t^{\,i} \tag{7.2}$$

We shall assume that all packets in a message have independent indentically distributed numbers of channel collisions. Let C_ℓ be the maximum of ℓ independent random variables with identical distributions given by Eq. (7.2). Hence,

$$\begin{aligned} \text{Prob}[C_\ell \leq i] &= (\text{Prob}[C \leq i])^\ell \\ &= (1 - p_n p_t^{\,i})^\ell \end{aligned} \tag{7.3}$$

Define the expectation of C_ℓ to be

$$\begin{aligned} E_\ell = E[C_\ell] &= \sum_{i=0}^{\infty} \text{Prob}[C_\ell > i] \\ &= \sum_{i=0}^{\infty} [1 - (1 - p_n p_t^{\,i})^\ell] \\ &= \sum_{i=0}^{\infty} [\ell\, p_n p_t^{\,i} - \binom{\ell}{2} p_n^{\,2}\, p_t^{\,2i} + \binom{\ell}{3} p_n^{\,3}\, p_t^{\,3i} - \\ &\qquad \ldots\ldots (-1)^{\ell+1}\, p_n^{\,\ell}\, p_t^{\,\ell i}] \\ &= \sum_{j=1}^{\ell} (-1)^{j+1} \binom{\ell}{j} \frac{p_n^{\,j}}{1 - p_t^{\,j}} \end{aligned} \tag{7.4}$$

The average delay for a message of ℓ packets is thus approximated by

$$D_\ell = R + \ell + E_\ell[R + (K + 1)/2] \quad (7.5)$$

where $R + (K + 1)/2$ is the average retransmission delay. Note that no buffer scheduling delays are included in this estimate. Thus, the actual average message delay will probably be slightly larger than D_ℓ. Using the above estimate, the average message delay for the channel is given by

$$D_{mess} = \sum_\ell D_\ell \cdot \mathrm{Prob}[\tilde{L} = \ell] \quad (7.6)$$

D_ℓ has been evaluated for $\ell = 1, 2, 4, 8, 20$ and plotted in Fig. 7-1 using numerical values of q_n and q_t for $K = 15$ in the infinite population model in Chapter 3. Thus, the $\ell = 1$ contour in Fig. 7-1 is the same as the $K = 15$ contour in Fig. 3-4. Several simulation points are also shown for $\ell = 4$ and 8 . These simulations were performed for the finite population model in Chapter 3 with 20 users and $K = L = 15$. Note that all simulation delay values are larger than their corresponding analytic values since buffer scheduling delays are included in the simulation message delay. Assuming that the channel input is equally divided between single-packet and eight-packet messages, the average message delay for the channel is shown in Fig. 7-2.

Simulations also indicate that when the channel input consists of many multi-packet messages, the slotted ALOHA channel is "more unstable" than before. It may be possible to extend the stability

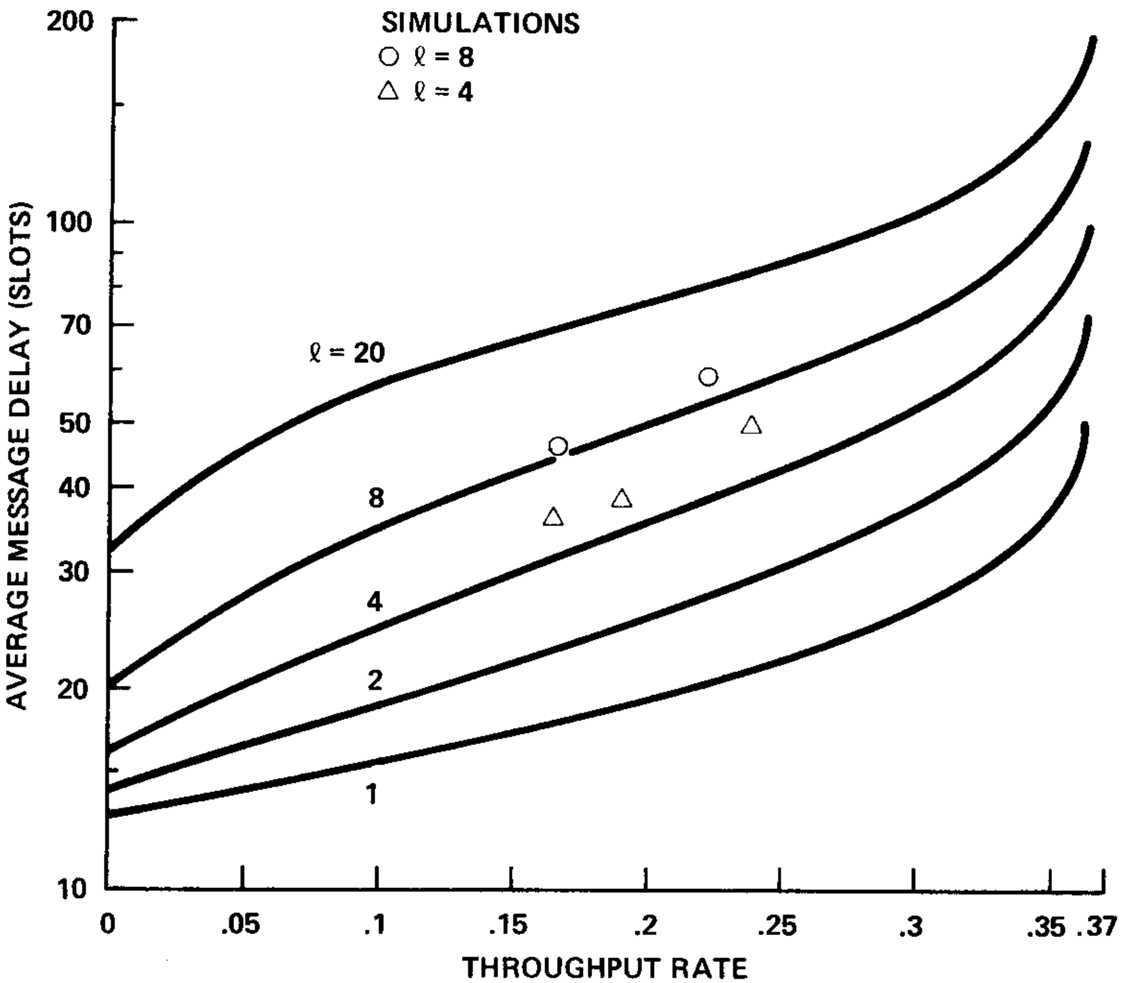

Figure 7-1. **Multi-Packet Message Delay Versus Throughput.**

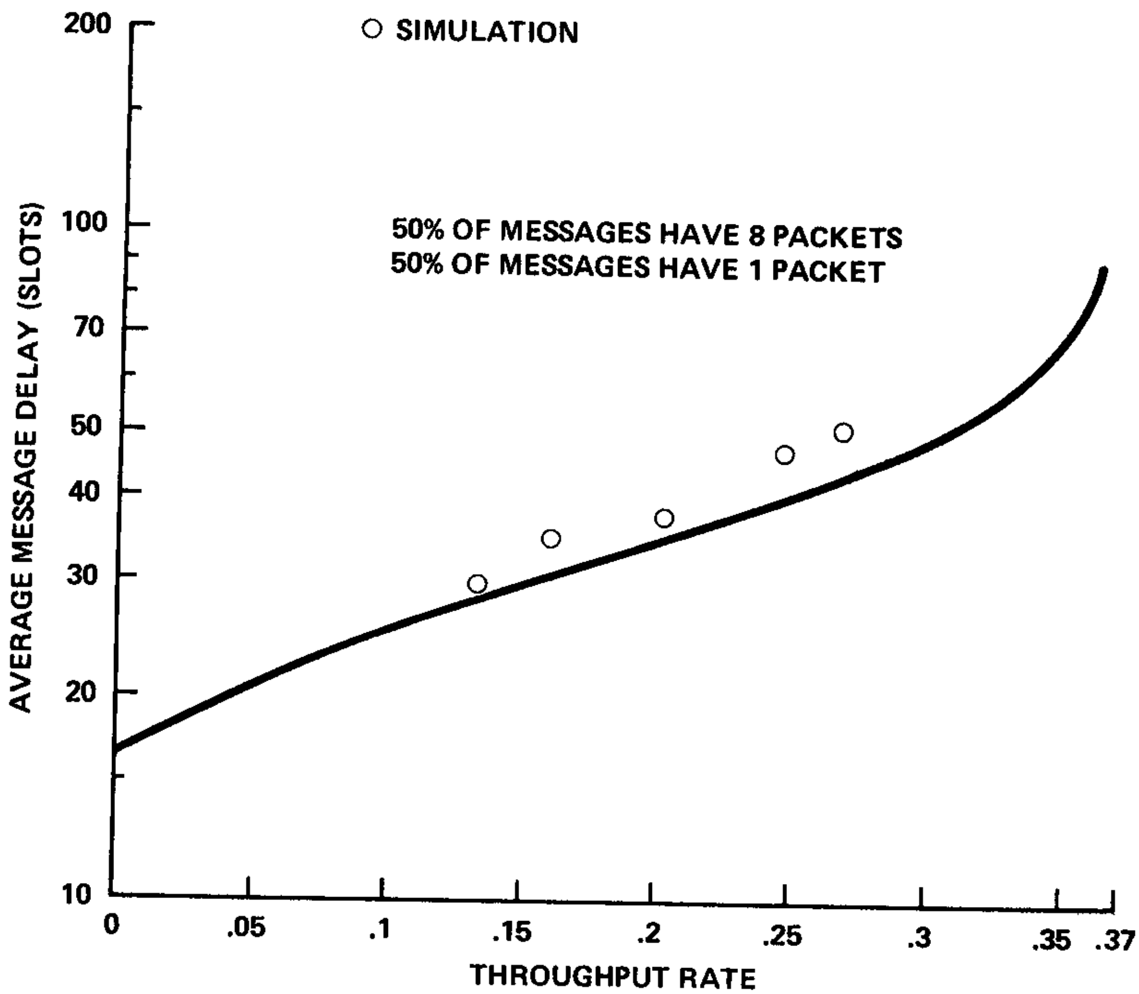

Figure 7-2. Throughput-Delay Tradeoff for Single-Packet and Eight-Packet Messages.

and control analyses in Chapters 5 and 6 to account for multi-packet messages. Such a study, however, is beyond the scope of this dissertation.

7.2 A Reservation System for Multi-Packet Messages

In Roberts' study of a satellite reservation system, the message arrivals to each station are assumed to be Poisson and equally divided between single-packet and eight-packet messages. The channel is dynamically partitioned into a slotted ALOHA reservation channel and a scheduled channel in the following manner [ROBE 73]:

> "....the satellite channel is divided into time slots of 1350 bits each. However, after every M slots one slot is subdivided into V small slots. The small slots are for reservations and acknowledgments, to be used on a contention basis with the ALOHA technique. The remaining M large slots are for RESERVED data packets. When a data packet or multi-packet block arrives at a station it transmits a reservation in a randomly selected one of the V small slots in the next ALOHA group. The reservation is a request for from one to eight RESERVED slots. Upon seeing such a reservation each station adds the number of slots requested to a count, J, the number of slots currently reserved. The originating station has now blocked out a sequence of RESERVED slots to transmit his packets in. Thus, there is one common queue for all stations and by broadcasting reservations they can claim space on the queue. It is not necessary for any station but the originating station to remember which space belongs to whom, since the only requirement is that no one else uses the slots."

Each small ALOHA slot is 224 bits long $(V = 6)$ and can accommodate an acknowledgment packet, a reservation packet or a small data packet. In a reservation packet, the reservation request is triplicated to improve the probability of error-free reception of the request by all stations.

The queueing delay in the above reservation system may be obtained by modelling the common queue as a M/G/1 queueing system

[KLEI 74D]. The delay incurred by a message consists of the slotted ALOHA (reservation) delay and the queueing delay. Thus, the message delay is bounded below by two satellite round-trip propagation times (≅ 0.5 second). Given that the channel input is equally divided between single-packet and eight-packet messages and assuming a 50 KBPS channel and ten stations, Roberts showed that the slotted ALOHA scheme gives a lower average message delay than the reservation system for $S_{out} < 0.15$; however, for $S_{out} > 0.15$, the reservation system gives a lower delay. Furthermore, a channel throughput rate close to one is achievable in the reservation system. For a large population of stations with low data rates, both the slotted ALOHA scheme and the reservation system are far superior to traditional techniques such as Time Division Multiple Access (TDMA) and Frequency Division Multiplexing (FDM). The latter techniques are competitive only when each individual station has a data rate of 50 KBPS. On the other hand, the packet switching techniques depend upon the total multi-station traffic rather than the individual station traffic for their efficiency.

Simulation results for the reservation system indicate that analytic results given by the M/G/1 queueing model are very accurate. However, the slotted ALOHA reservation channel exhibits unstable behavior [LAM 73]. Since the overall performance of the reservation system depends upon the slotted ALOHA reservation channel performance, some dynamic channel control scheme (such as those in Chapter 6) may be necessary.

When a significant fraction of the channel input consists of multi-packet messages, the reservation system has the following advantages compared to the slotted ALOHA scheme:

(1) The average message delay is smaller (except at a low channel throughput rate).

(2) The channel capacity is larger.

(3) The slotted ALOHA scheme tends to be "more unstable" with multi-packet messages. On the other hand, in the reservation system, the slotted ALOHA reservation channel input consists of only single packets (reservation requests, acknowledgments). For a relatively large V (number of small slots in a large slot), a low reservation channel input rate can be maintained for good channel stability using just a small fraction of the total channel bandwidth.

7.3 Reservation-ALOHA Schemes

We describe in this section two satellite "reservation" schemes based upon the slotted ALOHA scheme. By providing some degree of synchronization among the channel users, they are capable of achieving a channel throughput rate well in excess of the slotted ALOHA channel capacity. However, the channel performance of both reservation schemes depends upon (1) a small number of stations (in the order of R, the number of slots in a round-trip satellite propagation time), and (2) each station's input source (of packets) consists of both constant and random components.

The first reservation scheme is known as reservation-ALOHA in which the notion of a "time frame" is introduced [CROW 73]. The channel time is divided into consecutive time frames. Each time frame contains at least R slots. Channel slots in which a station had successful packet transmissions in the previous time frame are reserved for it to use again in the current time frame; no other stations are permitted to use these slots. Channel slots which were either empty or contained collisions in the previous time frame are available for random access by all stations in the current time frame. Thus, once a station has acquired a channel slot, it can keep the same slot in every time frame as long as it has something to transmit. Consequently, the channel performance is very good if the stations have deterministic uniform arrivals; the channel performance suffers when packet arrivals to stations are infrequent and random. Simulation results indicate that a channel throughput rate close to one is achievable at the expense of long packet delays. Also, for a given channel throughput, packet delays tend to increase as the number of stations increases [RETT 73B].

The second reservation scheme to be referred to as priority reservation-ALOHA adds a priority mechanism to the frame principle of reservation-ALOHA [BIND 72]. In this scheme, each station owns at least one slot per frame. The owner of a slot has the highest priority in the event of a collision in the slot. Thus, a station that has been idle is guaranteed channel usage within a maximum of two time frames. Beyond ownership, slots are also assigned. When two assignees of a slot are involved in a collision, the conflict is resolved by some globally known priority assignment mechanism. Thus, in all cases, each packet requires at most one retransmission. This scheme is more complex to implement than reservation-ALOHA, but promises to give smaller delays to stations with infrequent packet arrivals. The throughput-delay performance of this scheme has not been demonstrated.

CHAPTER 8

CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

Trends in the growth of computer-communication networks seem to indicate that the next generation of networks will be at least an order of magnitude larger than existing designs. Present implementations, however, are not directly applicable to very large networks. New techniques are needed which can provide cost-effective, high-speed communications for large populations of (potentially mobile) users scattered over wide geographical areas. Under these circumstances, we feel that packet switched satellite and ground radio systems provide attractive solutions to the design of communication subnets and terminal access networks respectively. A packet switching technique which has attracted considerable attention is the slotted ALOHA random access scheme.

The objective of this research was to develop analytic models with which we can evaluate and optimize the performance of a slotted ALOHA channel; our emphasis is on a large population of small users.* Results obtained in this dissertation are summarized in Section 1.4. Major contributions of this research may be classified into three categories:

- a coherent theory of channel behavior in which the key result is the characterization of stable and unstable channels
- evaluation of channel performance such as equilibrium throughput-delay tradeoffs for stable channels and stability-throughput-delay tradeoffs for unstable channels

*Recall that our abstract model of a small user represents a bursty user with buffering space for only one packet.

- dynamic channel control and estimation procedures for optimal control of unstable channels.

To design a slotted ALOHA random access system for small (bursty) users, the following steps may be followed:

(1) Evaluate the equilibrium throughput-delay tradeoff curves. Then, choose an operating value for K (or p) which gives an equilibrium channel throughput-delay tradeoff close to the optimum performance envelope.

(2) Given the average user think time $\frac{1}{\sigma}$, insure that the channel is not overloaded (as shown in Fig. 5-6(d)) by limiting the number of active users (M) who can "sign-on" and use the channel.

(3) For a small enough M , the channel may already be stable according to our stability definition. In this case, the system design is complete.

(4) For bursty users (i.e., $\frac{1}{\sigma}$ is large), a stable channel is associated with a very low channel throughput rate. Increasing M to increase channel utilization will render the channel unstable. In this case, go to either (5) or (6).

(5) If the unstable channel has an acceptable channel failure rate (i.e., FET) or one can be achieved by increasing the operating value of K without significantly increasing delay, the system design is complete. Otherwise, go to (6).

(6) Incorporate into the system (at each channel user) capability for storing channel information within a history window and implementing channel state estimation and dynamic control algorithms. Results in Chapter 6 indicate that with dynamic channel control, a channel throughput-delay performance close to the optimum performance envelope is achievable over an infinite time horizon for (originally) unstable channels.

(7) In a practical system, the load (M) on the channel will probably vary as a function of time with periods of heavy and light loads. The system should be designed for heavily loaded conditions since the performance of a lightly loaded channel is relatively insensitive to the system design. (See, for example, Figures 3-5, 6-12 and 6-13.)

Equilibrium throughput-delay tradeoffs have also been obtained for the large user model and multi-packet messages. In the former case, substantial improvements in the channel performance are possible if the large user accounts for a significant fraction of the channel input rate. In the latter case, if a large fraction of the channel input consists of multi-packet messages and the average message delay is the relevant performance measure, we concluded that Roberts' reservation system is superior to slotted ALOHA. Note, however, that the reservation system utilizes the slotted ALOHA scheme for broadcasting reservation requests; our slotted ALOHA results apply to the reservation subchannel in this system.

Numerical results in this dissertation were obtained assuming a 50 KBPS satellite channel with 1125 bits/packet and a channel round-trip propagation delay of 0.27 second. The models and methodology developed, however, are independent of these assumptions and may be applied to satellite channels with different data rates, ground radio systems as well as wire communications such as multi-drop lines and loop systems.

Extensions to this research

Before small satellite earth systems become an economic reality, satellite channel users will tend to be "big" and few in number. For example, the Satellite IMP, being designed for the ARPANET Satellite System, will have buffer space for 32 packets [BUTT 74]. This situation corresponds to the finite population model studied in Chapter 3. Our stability and dynamic channel control results in Chapters 5 and 6 may be extended to this case. However, the state description is now a vector consisting of the queue sizes at all satellite stations instead of a single variable such as in the linear feedback model.

For dynamic channel control procedures considered in this research, optimal control policies were found to be of the control limit type in all our numerical examples. A rigorous mathematical proof of this result remains an open problem.

The slotted ALOHA channel is characterized by the throughput-load curve depicted in Fig. 8-1, which is typical of "contention" systems [AGNE 73]. Unlike queueing systems in which the throughput increases to one as the system load increases, the throughput of a contention system increases to a maximum value and then decreases.

In this dissertation, we have characterized the unstable behavior and studied dynamic control schemes for a specific contention system, namely, slotted ALOHA random access. The probabilistic model and techniques employed here can probably be extended to solve stability and dynamic control problems of other contention systems.

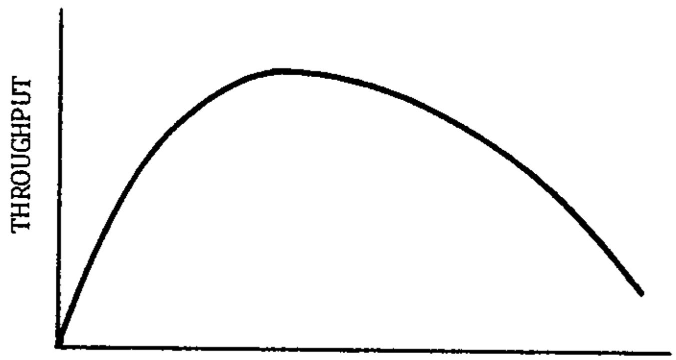

Fig. 8-1 A Typical Throughput-Load Curve for a Contention System

One class of contention systems consists of random access packet switching techniques (relatives of slotted ALOHA!) such as pure ALOHA, FM capture and carrier sense. These systems seem to exhibit unstable behavior similar to that of slotted ALOHA and may be dynamically controlled by similar schemes. As of now, most efforts in the study of these systems have been concentrated on the evaluation of the system capacity and equilibrium throughput-delay tradeoff. Little attention has been given to the problems of stability and control. For example, the ALOHA System at the University of Hawaii has been estimated to be able to support over 300 interactive users (assuming the multi-access channel capacity to be $\frac{1}{2e} \cong 18\%$) [ABRA 70]. We feel that these figures are unrealistic for an uncontrolled system, but may be achieved given appropriate dynamic channel control.

Many other existing systems can also be characterized as contention systems and exhibit unstable behavior similar to that of slotted ALOHA. A highway is a contention system and Fig. 8-1 represents the so-called "fundamental diagram of traffic" [ASHT 66]. Simulation results for store-and-forward packet switching networks show that they have throughput-load curves similar to that depicted in Fig. 8-1 [KAHN 71, DAVI 71]. It is interesting to note that heuristic flow control routing algorithms suggested by Kahn and Crowther [KAHN 71] and the so-called "isarithmic" networks proposed by Davies [DAVI 71] are similar in spirit to our dynamic channel control procedures.

Agnew considered a general deterministic model of a contention system and studied its dynamic control through pricing [AGNE 73]. A topic of future research interest is the formulation of a general probabilistic model of a contention system. It may be possible to extend the stability and dynamic control results in this dissertation to the general model.