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.

# CHAPTER 1

# INTRODUCTION

In the design of a computer-communication network, two problems may be identified. One is to provide long haul communications among geographically scattered computers and resources. The other is to provide local distribution of the network computing power, communication power and resources to populations of users.

An abstract model of a computer-communication network is depicted in Fig. 1-1. The first problem mentioned above corresponds to the design of the communication subnet in the figure for <u>computer-</u>

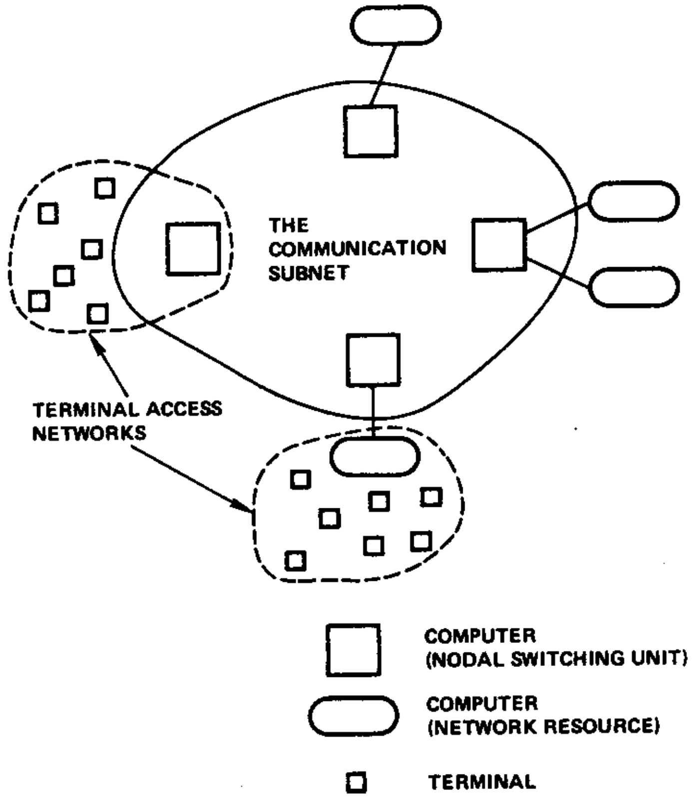

**Figure 1-1.** **An Abstract Model for a Computer-Communication Network.**

computer communications. The second problem corresponds to the design of the terminal access networks for terminal-computer communications. Two kinds of computers are distinguished in the model: (1) autonomous computer systems which constitute resources to be shared over the network, and (2) special purpose processors dedicated to network communication functions and acting as nodal switching units for data flow. (These nodal switching units will be referred to as the nodes of a communication subnet.)

The above abstract model description of a computer-communication network is consistent with the design philosophy of the ARPA (Advanced Research Projects Agency) Network [BUTT 74, CARR 70, CROC 72, FRAN 70, FRAN 72A, HEAR 70, KLEI 70, ORNS 72, ROBE 70, ROBE 72A].

In this dissertation, a packet switching technique based upon the random access concept of the ALOHA System [ABRA 70] will be studied in detail. This technique enables efficient sharing of a communication channel by a large population of users, each with a bursty input source (large ratio between the peak and average data rate). This packet switching technique may be applied to the use of satellite and ground radio channels for computer-computer and terminal-computer communications respectively. The multi-access broadcast capabilities of these channels render them attractive solutions to (1) large communication subnets with nodes over wide geographically distributed areas, and (2) large terminal access networks with potentially mobile terminals.

## 1.1 Present Computer-Communication Schemes

The simplest solution to providing communication between two points is to assign a dedicated channel for their use. This method is expensive in computer communications especially over long distances. Measurement studies [JACK 69, FUCH 70] conducted on time-sharing systems indicate that both computer and terminal data streams are bursty. That is, the peak data rate is much larger than the average data rate. (The ratio between them may be as high as 2000 to 1 [ABRA 73].) Consequently, if a high-speed point-to-point channel is used, the channel utilization is low since the channel is idle most of the time. On the other hand, if a low-speed channel is used, the transmission delay is large.

The above dilemma is caused by channel users imposing bursty random demands on their communication channels. By the law of large numbers in probability theory [FELL 68], the total demand at any instant of a large population of users is, with a high probability, approximately equal to the sum of their average demands. Thus, if a channel is dynamically shared in some fashion among many users, the required channel capacity may be much less than the unshared case of dedicated channels. This concept is known as statistical load averaging and has been applied in many computer-communication schemes to various degrees of success. These schemes include: polling systems [MART 72], loop systems [HAYE 71, PIER 71], Asynchronous Time Division Multiplexing (ATDM) [CHU 69], the random access scheme in the ALOHA System, and the store-and-forward packet switching concepts [BARA 64, KLEI 64, DAVI 68] implemented in the ARPA Network.

For almost a century, circuit switching dominated the design of communication networks. Only with the speed and cost of modern computers did packet communication become competitive. It was not until 1970 that the computer (switching) cost dropped below the communication (bandwidth) cost [ROBE 74]. This also marked the first appearance of packet switched computer-communication networks.

In a circuit switched network, a complete path of communication links must be established between two parties before they can communicate. The path (of links) is allocated for as long as the two parties want. In a store-and-forward packet switched network, the communication is broken into convenient size packets of information with addresses of source and destination attached to each packet. Packets are individually routed through the network to their destinations "hopping" from one node to another. In this case, the communication links are not allocated into paths for specific source-destination pairs of nodes; instead, each link is statistically shared by many nodes. The large savings possible from fuller utilization of the communication links justify the extra computer switching cost.

## 1.2 Satellite and Radio Communications in Large Networks

We are currently facing a booming demand for computer networks. For example, a survey for 17 European nations entitled "Eurodata--A Market Study on Data Communications in Europe, 1972-1985" estimates that data communication volume in those countries will soar twelvefold in the next dozen years. The total number of terminals was 79,600 in 1972; it will rise to 235,600 by 1976 and to 815,000 by 1985 [WRIG 73]. The feasibility of packet switched networks with up to 1000 nodes and tens of thousands of terminals is being investigated [NAC 73, FRAN 73].

These numbers are at least an order of magnitude larger than any other system design attempted. Extension of current computer-communication techniques to networks of such magnitude cannot be easily done. For instance, the adaptive routing techniques currently implemented in the ARPA network cannot be directly utilized in a very large network because of excessive IMP processing time, memory requirements and traffic overhead [NAC 73]. The system overhead in conventional polling schemes is directly proportional to the number of terminals sharing the communication channel; such schemes are thus not appropriate for a large number of terminals.

To design cost-effective computer-communication networks for the future, new techniques are needed which are capable of providing efficient high-speed computer-computer and terminal-computer communications in a large network environment.

The application of packet switching techniques to radio communication (both satellite and ground radio channels) provides a solution.

Radio is a multi-access broadcast medium. A signal generated by a radio transmitter may be received over a wide area by any number of receivers. (This is the broadcast capability.) Furthermore, any number of users may transmit signals over the same channel.* (This is the multi-access capability.) Hence, a single ground radio channel provides a completely connected network topology for a large number

---

*However, if two signals (packet transmissions) at the same carrier frequency overlap in time at a radio receiver, we assume that neither is received correctly.

of nodes within line of sight of each other. On the other hand, a satellite transponder in geosynchronous orbit above the earth acts as a radio repeater. Any number of earth stations may transmit signals up to the satellite at one carrier frequency (the multi-access channel). Any signal received by the satellite transponder is beamed back to earth at another frequency (the broadcast channel). This broadcasted signal may be received by all earth stations covered by the transponder beam. Thus, a satellite channel (consisting of both carrier frequencies)* provides a completely connected network topology for all earth stations covered by the transponder beam (see Fig. 1-2).

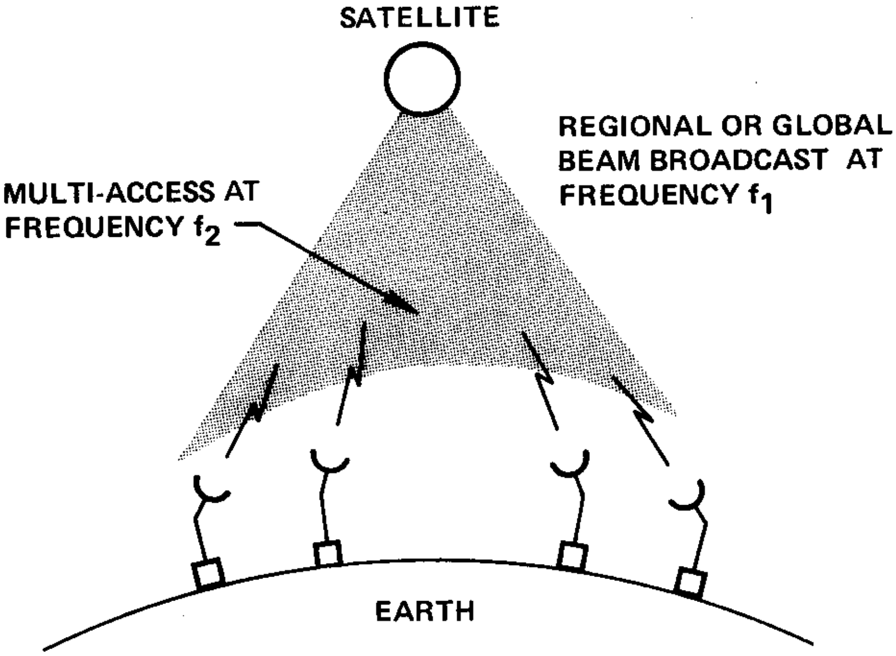

**Figure 1-2. Packet Switch in the Sky.**

---

*This is actually a satellite circuit.

The provision of a completely connected network topology by a satellite or radio channel eliminates complex topological design and routing problems in large networks [FRAN 72B, GERL 73]. Moreover, the use of packet switching techniques enables a large population of users to statistically average their total load at the high-speed multi-access channel. Each user also transmits data at the (wideband) data rate of the channel. Thus, both high channel utilization and small packet delays are possible through the use of appropriate packet switching techniques. We shall elaborate upon the advantages of packet switched radio communication systems in the next chapter.

We give here a description of the ALOHA System which is one of the first packet radio communication systems.

The ALOHA System is an experimental terminal access network at the University of Hawaii [ABRA 70, KUO 73]. Two 24 KBPS radio channels are used. One of the two channels is used by all remote terminals for transmitting data into the central computer. (This is a multi-access channel.) The other channel is used for transmitting data out of the central computer to the remote terminals. (This is a broadcast channel.) The transmission of data from the central computer to the terminals is relatively simple since the central computer can schedule its own use of the broadcast channel. The multi-access channel, however, uses the following radically new random access packet switching technique. (This scheme will be referred to as pure ALOHA.) Each terminal transmits data to the central computer over the same 24 KBPS channel in 30 msec. bursts (packets) in a completely unsynchronized manner. A transmitted packet can be received incorrectly as a result of two types of errors: (1) random noise errors, and (2) errors caused by interference (at the radio receiver of the central computer) with a packet transmitted by another terminal. If and only if a packet is received with no error, it is acknowledged by the central computer. After transmitting a packet, a terminal waits a given amount of time (time-out interval) for an acknowledgment; if none is received, the packet is retransmitted. This process is repeated until successful transmission and acknowledgment occur or until the process is stopped by the terminal. It was estimated that the ALOHA System could theoretically support more than 300 active terminals [ABRA 70].

There is currently an immense worldwide interest in the development of satellite communications systems. In addition to the worldwide INTELSAT system [PUEN 71], there are currently in operation two domestic satellite systems: Anik in Canada [GRAY 74] and Molniya in the U.S.S.R. With the advent of domestic satellite systems in the United States [CACC 74], various satellite computer-communication systems based upon the packet radio communication concept of the ALOHA system have been proposed [ABRA 73, CROW 73, KLEI 73A, ROBE 73]. In particular, Abramson suggested that a single transponder in a domestic satellite system could easily provide 10 MBPS for a public packet switched service with 100 earth stations over the U. S.; each earth station has an average data rate of 15 KBPS and a maximum transmission rate of 10 MBPS [ABRA 73]. Dunn and Eric gave a comparison of illustrative costs for some of the above proposed packet switched satellite systems assuming the use of small earth stations for 100 nodes serving the 40 largest metropolitan areas in the U. S. [DUNN 74]. In a recent application to the FCC for approval of a public packet switched network, a 1.5 MBPS satellite channel was included in the proposed network configuration based on land lines [TELE 73].

## 1.3 Packet Switching Techniques

Consider a radio communication system such as the satellite system depicted in Fig. 1-2 or the ALOHA System. In each case, there is a broadcast channel for point-to-multipoint communication and a multi-access channel shared by a large number of users. Each user is assumed to have a small average data rate relative to the channel transmission rate, but each transmits packets of data at the channel transmission rate. (In other words, the users have bursty input sources.)

Since the broadcast channel is used by a single transmitter, no transmission conflict will arise. All nodes covered by the radio broadcast can receive on the same frequency, picking out packets addressed to themselves and discarding packets addressed to others.

The problem we are faced with is how to effect time-sharing of the multi-access channel among all users in a fashion which produces an acceptable level of performance. As soon as we introduce the notion of sharing in a packet switching mode, we must be prepared to resolve conflicts which arise when simultaneous demands are placed upon the channel. There are two obvious solutions to this problem:

the first is to form a queue of conflicting demands and serve them in some order; the second is to "lose" any demands which are made while the channel is in use. The former approach is taken in ATDM and in a store-and-forward network assuming that storage may be provided economically at the point of conflict. The latter approach is adopted in the ALOHA System random access scheme; in this system, in fact, all simultaneous demands made on the radio channel are lost.

Let us define channel throughput rate $S_{out}$ to be the average number of correctly received packet transmissions per packet transmission time (assuming stationary conditions). We also define channel capacity $S_{max}$ to be the maximum possible channel throughput rate.

The channel capacity of a pure ALOHA multi-access channel was estimated to be $\frac{1}{2e} \cong 18\%$ for a fixed packet size [ABRA 70]. Under similar assumptions, Gaarder showed that a pure ALOHA channel with a fixed packet size is always superior (in terms of channel capacity) to one with different packet sizes [GAAR 72].

Since various propagation delays are involved in a geographically distributed radio communication system, let us define a global reference time called channel time. The channel time will be assumed to be the satellite transponder time in a satellite system and to be the central computer time in a terminal access network. Note that if two or more packet transmissions overlap in time at the radio receiver (of the satellite transponder or the central computer), none is received correctly. This event will be referred to as a channel collision.

Roberts suggested that the channel time may be slotted by requiring all channel users to synchronize the leading edge of each packet transmission to coincide with an imaginary channel time slot boundary. The duration of a channel time slot is equal to a packet transmission time. The resulting scheme will be referred to as "slotted ALOHA random access" or "slotted ALOHA." (In Fig. 1-3, we show packet transmissions and retransmissions in a slotted ALOHA system consisting of four users.) The channel capacity of a slotted ALOHA channel was estimated* to be $\frac{1}{e} \cong 36\%$ [ROBE 72B].

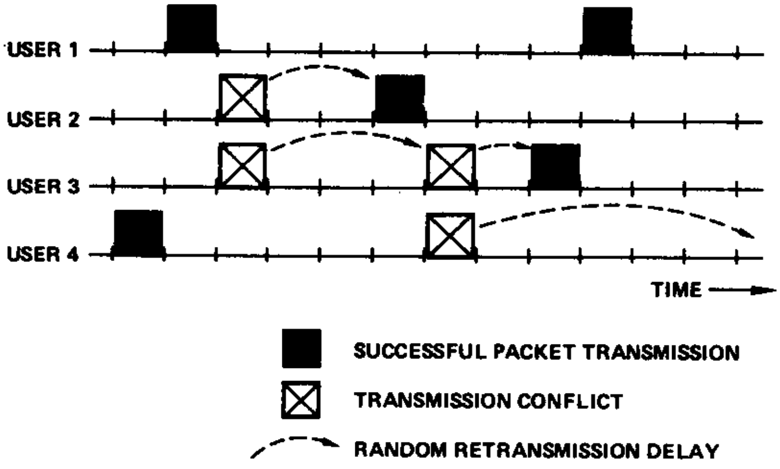

**Figure 1-3. Slotted ALOHA Random Access.**

Several variants of the ALOHA random access technique have been proposed for ground radio systems. One technique is known as FM capture [ROBE 72B]. In the event of a channel collision, the

---

*A derivation of this result is given in Chapter 3.

strongest signal (packet transmission) may still be received correctly by a good FM receiver. As a result, the ALOHA random access channel capacity may be larger than the 36% limitation. Another scheme is currently being investigated for ground radio packet switching systems in which the maximum propagation delay is small compared to a packet transmission time (say, less than 0.1). In such systems, the users may "listen before transmitting" in order to determine if the channel is in use by some other user; such systems are referred to as "carrier sense" systems. In these systems, a channel capacity much larger than 36% is possible [KLEI 74C].

Several reservation schemes based upon the slotted ALOHA random access technique have been proposed for satellite packet switching systems. In a satellite channel, the round-trip propagation delay is approximately a quarter of a second and is in the order of many channel time slots. In one reservation system [ROBE 73], 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. The minimum delay in this system is at least twice the round-trip propagation delay (half a second). Thus, this sytem 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 measure of channel performance. Two "reservation-ALOHA" schemes have also been proposed [BIND 72, CROW 73]. These schemes may be used if there is only a small number of channel users (say, in the order of the number of slots in a round-trip propagation time), and if the channel input source has constant as well as random components.

## 1.4 Summary of Results

We examined several radio communication packet switching schemes in the last section. Some of these schemes (FM capture, carrier sense, reservation-ALOHA) are variants of the slotted ALOHA random access concept; some others (e.g., Robert's reservation system) are dependent upon the slotted ALOHA random access technique.

The basic goal of this dissertation is to develop analytic models with which we can predict and optimize the stability-throughput-delay performance of a multi-access channel using the slotted ALOHA random access technique. The analytic models, despite their limitations (due to various mathematical assumptions), suggest a system design methodology and operational strategies for packet switching random access systems. Our emphasis is on a large population of users with bursty input sources; each user has an average data rate which is small relative to the channel transmission rate.

It has been realized that in a slotted ALOHA random access channel, channel "saturation" may occur as a result of time fluctuations in the channel input or inherent channel instability [KLEI 73A, KLEI 73B, KLEI 74A, LAM 73, METC 73A, RETT 72]. However, existing results on the channel capacity [ABRA 73] and throughput-delay tradeoff [KLEI 73A] have all assumed steady-state conditions. A channel control strategy derived from a steady-state analysis has been proposed which may prevent channel saturation [METC 73A].

Considering the state of the research, only fragmented results are available on the performance (channel capacity, delay, dynamic behavior and stability) of the slotted ALOHA random access channel. Little attention has been paid to the problem of dynamic channel control. In this dissertation, we attempt to give a coherent theory of channel behavior and to develop techniques to optimize the system design and dynamically control the channel performance.

In Chapter 2, we summarize various advantages of satellite and radio communications over conventional wire communications. Satellite channel characteristics and cost trends are examined. Abstract models are then given for the random access channel and channel users to be considered in the dissertation.

In Chapter 3, an analytic model is developed to predict the equilibrium throughput-delay tradeoff. The minimum throughput-delay performance envelope and the corresponding optimal retransmission delays are characterized. These results are generalized to a model which includes a "large" user in the channel user population.*
In this case, significant improvements in the channel throughput-delay performance are possible. A channel throughput rate equal to one may be achieved. A continuum of throughput-delay performance envelopes are presented. Abramson's result [ABRA 73] on channel capacity will also be given. Simulation results have been obtained which agree very well with analytic results. However, the assumption of channel

---

*This situation arises when, for example, in a terminal access packet radio system, a single radio channel is used for both terminal-to-computer (multi-access) and computer-to-terminal (broadcast) communications [GITM 74].

equilibrium may be valid only for finite time periods beyond which the channel goes into saturation.

In Chapter 4, the complexity of an exact mathematical analysis of channel dynamics is illustrated. This serves to motivate our use of approximations. The channel traffic (packet transmissions and retransmissions in a channel slot) is shown to be Poisson distributed in the limit of an infinite average retransmission delay and under the "weak independence assumption." A difference equation is derived which gives a deterministic approximation of the dynamic behavior of the channel subject to time varying inputs.

In Chapter 5, stable and unstable channels are characterized and a stability definition is given. For stable channels, previous equilibrium throughput-delay results given in Chapter 3 are actually valid and achievable over an infinite time horizon. For unstable channels, the degree of instability is quantified by the definition of the stability measure FET. An efficient algorithm is developed for the calculation of FET. Unstable channels, in general, are characterized by a large population of users. The "stability" (i.e., FET) of an unstable channel may be improved by reducing the channel input rate or increasing the average packet delay. The appropriate channel performance measure for unstable channels is the stability-throughput-delay tradeoff. Some stability-throughput-delay tradeoff curves are presented.

Under the assumption that channel users have bursty input sources with low data rates (relative to the channel speed), stable channels are characterized by a relatively small population of users and thus, a small throughput rate. To obtain a high channel throughput rate, dynamic channel control is necessary to convert unstable channels into stable channels. In Chapter 6, Markov decision theory is used to formulate three dynamic channel control procedures (ICP, RCP, IRCP). It is shown that optimal stationary policies exist. Furthermore, a theorem is proved that the same stationary control policy will maximize the stationary channel throughput rate and minimize the average packet delay simultaneously. An efficient computational algorithm (POLITE) is developed which utilizes Howard's policy-iteration method [HOWA 71] and is capable of solving for an optimal stationary policy in a small number of computational steps. Numerical results indicate that optimal control policies are of the control limit type, but a rigorous mathematical proof remains an open problem. Throughput-delay tradeoffs given by optimal control policies are presented. These throughput-delay results are very close to the optimum performance envelope in Chapter 3 and are achievable over an infinite time horizon for (originally) unstable channels. Since in a practical system the exact channel state is not known but must be estimated, some channel control-estimation (CONTEST) algorithms based upon the dynamic control procedures are proposed. A heuristic control algorithm is also suggested. Simulations indicate that for a channel throughput rate up to 0.32, throughput-delay results close to the optimum channel performance are achievable through application of the CONTEST algorithms.

In Chapter 7, multi-packet messages are considered. An approximate formula for the average message delay is derived. Roberts' reservation system and two reservation-ALOHA schemes are surveyed.

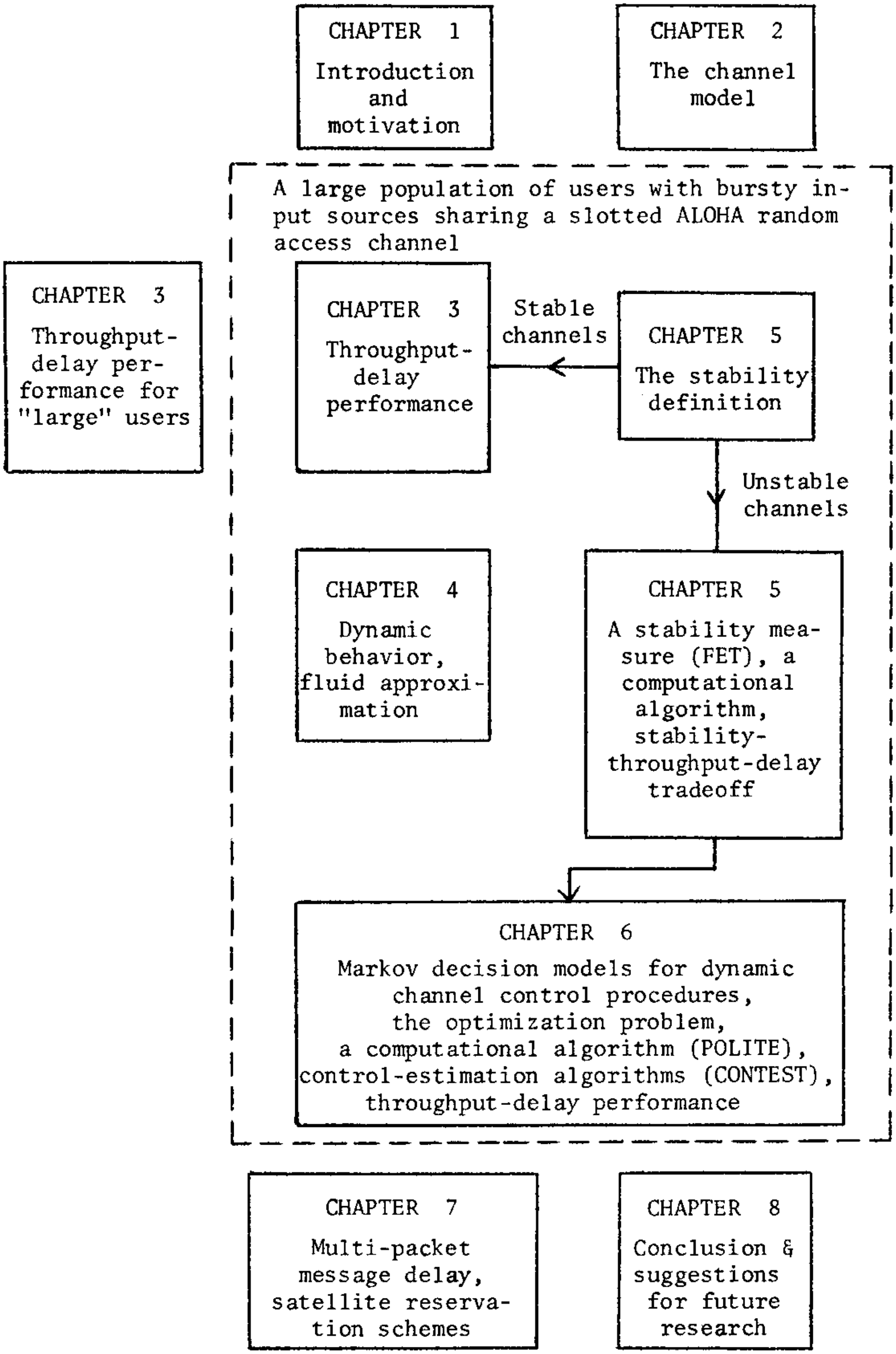

Fig. 1-4 Summary of results in this dissertation.

In Chapter 8, we give some concluding remarks and suggest topics of future research interests.

The above summary of results in this dissertation is depicted in Fig. 1-4.

This research was motivated by the research and development activities of the ARPANET Satellite System intended to incorporate satellite packet communication into the existing ARPA Network [ABRA 72, BUTT 74]. Consequently, the use of a satellite channel is considered in numerical examples throughout this dissertation. A satellite channel is characterized by a large channel propagation delay which will be reflected in all our numerical results. However, the models and methodology developed in this dissertation are applicable to ground radio systems. In fact, before small satellite earth stations become a reality (economically), the assumption of a large population of channel users is more appropriate in a ground radio environment. We also note that application of the random access techniques considered here is not limited to satellite and radio multi-access broadcast channels. They can, for example, also be applied to terminal access networks with multi-drop lines [HAYE 72].

In summary, the major contributions of this research are:

(1) The characterization and performance evaluation of stable and unstable channels--for stable channels, techniques are developed to solve for the optimum throughput-delay performance envelope. For unstable channels, the degree of channel instability is

quantified by the definition of the stability measure FET. An efficient algorithm has been developed to calculate FET. The channel stability-throughput-delay performance is shown.

(2) Dynamic channel control procedures which prevent channel saturation in an unstable channel to give better channel utilization--Markov decision models are developed for various dynamic control procedures. Optimal stationary control policies are shown to exist which will maximize the stationary channel throughput rate and minimize the average packet delay simultaneously. An efficient algorithm (POLITE) based upon the policy-iteration method finds an optimal stationary policy in a small number of computational steps. Control-estimation (CONTEST) algorithms are proposed for practical implementation of the above control procedures. Truly stable channel throughput-delay performance close to the optimum performance envelope is achievable using the dynamic control procedures.

In conclusion, despite model limitations as a result of various assumptions for mathematical convenience, we feel that the results and methodology presented in this dissertation are valuable and will lead to sound design procedures and operational strategies for packet communication systems using radio and satellite channels in a large network environment.