INFORMATION TO USERS

This was produced from a copy of a document sent to us for microfilming. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help you understand markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure you of complete continuity.
2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame. If copyrighted materials were deleted you will find a target note listing the pages in the adjacent frame.
3. When a map, drawing or chart, etc., is part of the material being photographed the photographer has followed a definite method in "sectioning" the material. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.
4. For any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.
5. Some pages in any document may have indistinct print. In all cases we have filmed the best available copy.

University
Microfilms
International
300 N. ZEEB RD., ANN ARBOR, MI 48106

Guha, Dilip Kumar

DYNAMIC RESERVATION MULTIPLE ACCESS TECHNIQUE FOR DATA TRANSMISSION VIA SATELLITES

City University of New York PH.D. 1982

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106

DYNAMIC RESERVATION MULTIPLE ACCESS TECHNIQUE FOR DATA TRANSMISSION VIA SATELLITES

by

Dilip Kumar Guha

A dissertation submitted to the Graduate Faculty
in Electrical Engineering in partial fulfillment of the
requirements for the degree of Doctor of Philosophy,
The City University of New York.

1981

This manuscript has been read and accepted for the Graduate Faculty in Engineering in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

1/6/82
date

Chairman of Examining Committee

1/6/82
date

Paul R Karmel
Executive Officer

T. N. Saadawi

N. Schienberg

T. Apelewicz
Supervisory Committee

The City University of New York

Abstract

DYNAMIC RESERVATION MULTIPLE ACCESS TECHNIQUE FOR DATA TRANSMISSION VIA SATELLITES

by

Dilip Kumar Guha

Advisor: Professor Donald L. Schilling

A new reservation-based multiple access packet switching technique applicable to packet communications using a satellite channel is proposed. The objective of this research is to optimize the system performance in terms of throughput and delay and to develop analytical models to analyze the scheme.

Simulation results show markedly improved delay-throughput characteristics over other existing and proposed multiple access techniques with especially significant performance gains at high traffic and also as the traffic imbalance among the users increases. A good agreement between the simulation results and the analytical results is obtained.

ACKNOWLEDGEMENT

I welcome this opportunity to express gratitude to my advisor, Professor D. L. Schilling. I am also grateful to Professor T. N. Saadawi for his support and suggestions. Many of their ideas are present in this research.

I would also thank Professors J. Salz, N. Schienberg and T. Apelewicz for their criticism and help, and especially my wife, C. Guha, for her continuous support and encouragement.

I gratefully appreciate the financial support of Bell Telephone Laboratories during my Ph.D. program and also for preparing this dissertation.

CONTENTS

LIST OF FIGURES AND TABLES

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 resources to populations of users.

An abstract model of a computer-communication network is shown in Fig. 1-1. The first problem mentioned above corresponds to the design of the communication subnet in the figure for computer–computer communications. The second problem corresponds to the design of the terminal access networks for terminal–computer communications.

The field of large data networks has seen a tremendous growth in the last decade, and in many ways heuristic development has preceded and outdistanced optimal design. In the subarea of multiple access schemes for networks that include broadcast (and particularly satellite) links a great deal of work has been done in inventing and analyzing strategies and protocols for the shared use of a common channel by geographically separated users. The problems of choosing the optimum access protocol (in the sense of minimizing both average and maximum delays per message for a given throughput) have not been formulated yet.

In this dissertation, a packet switching technique based upon dynamic reservation multiple access concept 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.

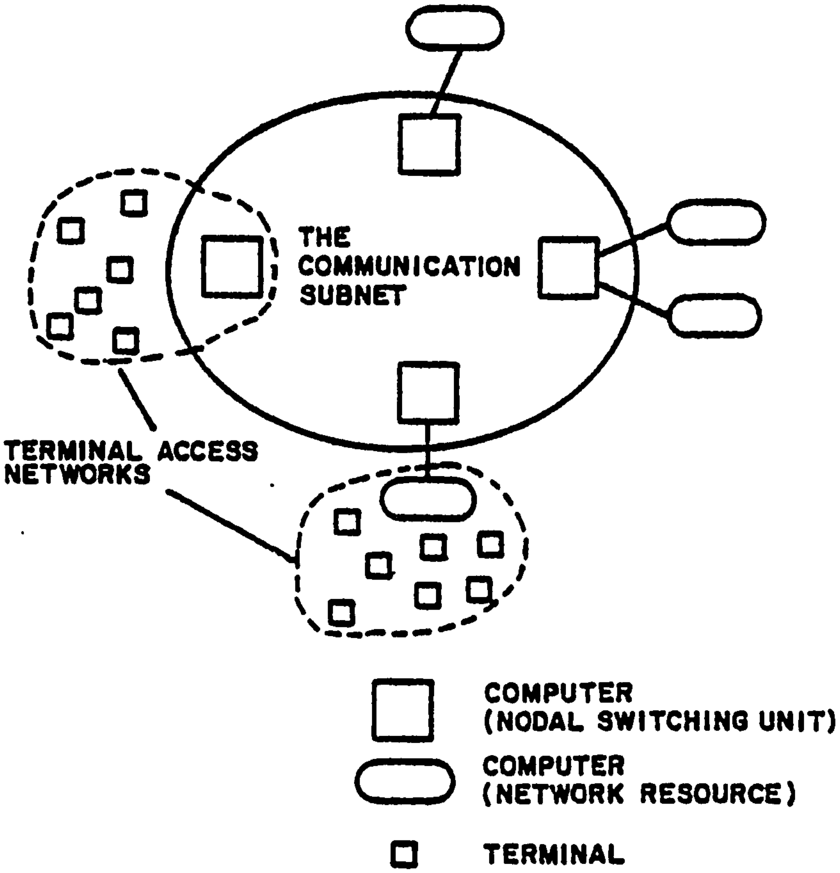

FIGURE 1·1. AN ABSTRACT MODEL FOR A COMPUTER-COMMUNICATION NETWORK

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 [1,2] 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 [3].) Consequently, if a point-to-point channel is used, the channel utilization is low since the channel is idle most of the time.

By the law of large numbers in probability theory [4], 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 [5], loop systems [6,7], Asynchronous Time Division Multiplexing (ATDM) [8], the random access scheme in the ALOHA System, and the store-and-forward packet switching concepts [9,10] 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 [11]. 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 Communications in Large Networks

We are currently facing a booming demand for computer networks. The feasibility of packet switched networks with up to 1000 nodes and tens of thousands of terminals is being investigated [12,13]. 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 [12]. 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.) 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 (this is the multi-access capability.) 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).

The provision of a completely connected network topology by a satellite eliminates complex topological design and routing problems in large networks [14,15]. 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.

1.3 Summary of Results

The basic goal of this dissertation is to develop control schemes that can be employed when using a geostationary satellite channel for intercommunications between a set of geographically distributed nodes and, more importantly, modeling and analysis of such schemes. Although many time-division multiplexed schemes have been proposed, very few of the schemes have been modeled analytically. In this dissertation, we develop and analyze a dynamic reservation-based multiple access packet switching technique, which minimizes both average and maximum delays and maximizes traffic throughput. The analysis is based on a combination of two Markov chain model; one Markov chain describes the status of the buffer contents of a typical user, we refer to it as the User Markov Chain and one that describes the status of all the users of the channel, we refer to it as the System Markov Chain.

Simulation results have been obtained for the system under various traffic distributions among nodes, and are presented later in the paper. These results show markedly improved delay-throughput characteristics over other existing and proposed multiple access techniques with especially significant performance gains as the traffic imbalance among the users increases. A good agreement between the simulation results and the analytic results is obtained.

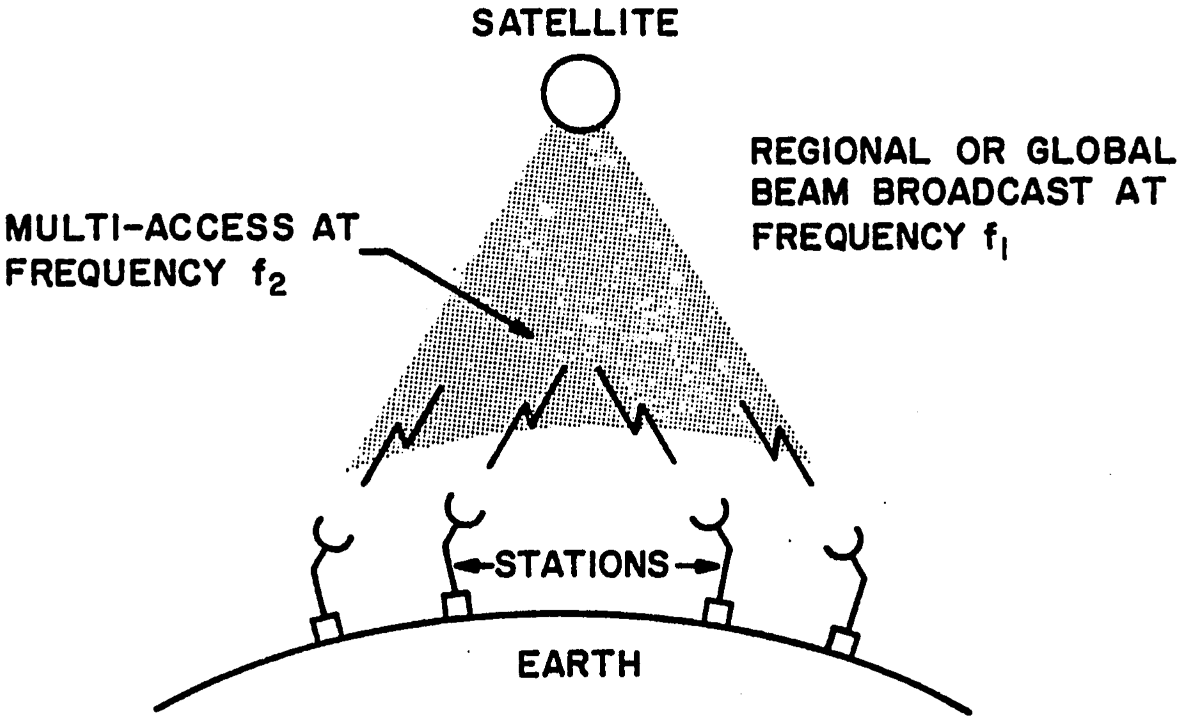

FIGURE 1.2. PACKET COMMUNICATION VIA SATELLITES

For equal traffic distribution among nodes, each node needs approximately two (2) slots at most to transmit packets in a time frame to achieve the best performance. It is found that the average number of frame size is approximately 1 slot per node at low throughput and less than 2 slots per node at high throughput. In other words, the knowledge of the second preceeding reservation vector for a node is not necessary to have two slots at most in a time frame.

For unequal traffic distribution among nodes, on the other hand, the delay-throughput characteristic is improved by assigning variable slots to the nodes in a time frame. The knowledge of preceding reservation vectors is, therefore, necessary to assign variable slots to a node in a time frame. Optimum values of variable slots to be assigned to the nodes are obtained for unequal traffic among nodes in which the arrival rate of packets for one node is eight times higher than that of other nodes. However, these values will depend on the traffic distribution among nodes. Further analysis of these should be made.

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

In Chapter 3, we examine several existing and proposed satellite communication packet switching schemes. A new reservation multiple access scheme is then presented.

In Chapter 4, this proposed scheme is modeled and analyzed in a queuing theoretic framework. A numerical solution to the mathematical model is obtained. Steady-State System performance parameters, such as average queue sizes and waiting times which are useful in system design, are obtained.

In Chapter 5, a FORTRAN simulation model is developed, and a comparison is made between the analytical solution and the simulated performance. A good agreement between these two is obtained. The simulation results are compared with other existing and proposed satellite schemes and show markedly improved delay-throughput characteristics over them.

Simulation results are also obtained for the system using different arrival rates of two classes of traffic under various loads and node distributions, and are presented later in this chapter.

In Chapter 6, conclusions and suggestions for further work are finally presented.

CHAPTER 2

THE CHANNEL MODEL

Consider a radio communication system such as the satellite system depicted in Fig. 1-2. In this 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 among all users in a fashion which produces an acceptable level of performance. However, as we mentioned before, we are interested in the use of satellite packet communication for large populations of users over wide areas. With this in mind, we discuss below some advantages of satellite communications over conventional wire communications. In the next section, we shall examine some satellite channel characteristics and cost trends. Finally, in the last section, abstract models for the channel will be given.

2.1 Advantages of Satellite Packet Communications

Consider the use of packet communication in a computer-communication network environment to support large populations of (bursty) users over wide areas. We can identify the following advantages of satellite channels over conventional wire communications:

(1) Elimination of complex topological design and routing problems

Topological design and routing problems are very complex in large networks [14, 15]. Existing implementations suitable for a (say) 50 node network may become totally inappropriate for a 500 node network required to perform the same functions [13]. On the other hand, satellite channels used in the multi-access broadcast mode provides a completely connected network topology, since every user may access any other user covered by the broadcast.

(2) Wide geographical areas

Wire communications become expensive over long distances (e.g. transcontinental, transoceanic). Even on a local level, the communication cost for an interactive user on an alphanumeric console over distances of over 100 miles may easily exceed the cost of computation [16]. On the other hand, satellite communication is relatively distance-independent.

(3) Mobility of users

Since satellite is a multi-access broadcast medium, it is possible for users to move around freely. This consideration will soon become important in the development of personal terminals in future telecommunication systems [17,18].

(4) Large population of active and inactive users

In wire communications, the system overhead usually increases directly with the number of users (e.g., polling schemes). The maximum number of users is often bounded by some hardware limitation (e.g., the fan-in of a communications processor). In satellite communication, since each user is merely represented by an ID number, the number of active users is bounded only by the channel capacity and there is no limitation to the number of inactive (but potentially active) users.

(5) Flexibility in system design

A satellite packet communication system can become operational with two or three users. The size of the user population can be increased up to the channel capacity. More users can be accommodated by increasing the satellite channel bandwidth. In other words, the communication system can be made bigger or smaller without major changes in the basic system design and operational schemes.

(6) Statistical load averaging

In wire communications, the use of adaptive routing techniques [19] in a store-and-forward packet switched network, for example, enables communication links to be utilized than in a circuit switched network. However, at any instant, there may still be unused channel capacity in some parts while congestion exists in other parts of the network. The application of packet switching techniques to a single high-speed satellite channel permits the total demand of all user input sources to be statistically averaged at the channel. Note also that each user transmits data at the (high-speed) channel rate.

(7) Multi-access broadcast capability

This capability in satellite communication may be useful for certain multi-point to multi-point communication applications.

(8) Reliability

The normal bit error rate of a satellite channel using forward error correction techniques is estimated to be $P_{be} = 1 \times 10^{-9}$ and better, compared to $P_{be} = 1 \times 10^{-5}$ for typical terrestrial links [20].

2.2 Satellite Channel Characteristics and Cost Trends

In addition to their multi-access broadcast capability, satellite channels have other characteristics which distinguish them from conventional communication channels and must be taken into consideration in any satellite communication system design.

The Satellite

We quote the following information on the Anik satellites [21]:

> "The satellites are about 6 feet in diameter and 11 feet high. At launch they weighed about 1250 lbs. and their orbiting weight is about 600 lbs. Each satellite's electronics system is powered normally by about 23,000 solar cells with sufficient on-board battery capability to provide power during eclipse periods when the satellite is in shadow... . The life expectancy of the batteries is a minimum of seven years. Each spacecraft consists of an electronic communications system, literally a microwave receiving and transmitting station in space, and on-board propulsion systems to inject it into its synchronous orbit and correct for wobble or spin."

Round-Trip Delay (RTD)

A satellite in geosynchronous orbit is stationed approximately 36,000 kilometers above the equator. A signal transmitted by an earth station to the satellite transponder (at one frequency) is beamed back earth (at another frequency) and can be received by all stations covered by the transponder beam. The round-trip propagation delay (RTD) is approximately a quarter of a second. Depending on a station's geographical location on each, a difference of 15 milliseconds exists. Furthermore, the satellite drifts approximately 200 miles in range during the day, which produces an additional two milliseconds difference in RTD. Without loss of generality, we shall assume the maximum RTD value for all stations in our work.

Burst Synchronization and Channel Slotting

Despite differences in the RTD values of earth stations, tests performed with an Experimental TDMA system over INTELSAT I (Early Bird) during August 1966, indicate that transmission bursts from different stations can be synchronized at the satellite transponder requiring guard times less than 200 nanoseconds [22]. In our case of a packet switched system, the satellite transponder time was assumed to be the global reference time (channel time) for all earth stations. The very small guard time required for burst (packet transmission) synchronization demonstrates the feasibility of channel slotting. Several slotting techniques have been examined by Rettberg [23].

Automatic Acknowledgement

To ensure data integrity in a communication channel, a very reliable method is the use of an error detecting block code in conjunction with positive acknowledgment of each message by its recipient. In a satellite channel, any signal relayed by the transponder is received by all earth stations including the sender(s).

Data Rates and Small Earth Station

An excellent introduction to the currently operational SPADE system (using an INTELSAT IV global-beam transponder) is available in [24]. We summarize here some relevant information on channel data rates and consideration for small earth station operation.

The SPADE system utilizes single-channel-per-voice-carrier transmission. 7-bit PCM encoding is used for voice with the encoded output at 56 KBPS (8000 samples/sec.). The channel transmitted bit rate is 64 KBPS. Since 4-phase coherent PSK modulation is used, the transmitted symbol rate is 32,000 symbols per second using a bandwidth of 38 KHz. The SPADE channel unit can be operated in continuous or voice-activated mode depending on whether data or voice is transmitted.

Costs and Other Considerations

We emphasize again that we are primarily interested in systems involving fairly large populations of users. In such a packet switched satellite broadcast system, the cost of earth stations dominates the satellite bandwidth cost. A standard INTELSAT earth station with a 97-foot antenna costs between $3-3.5 million dollars. We note that if a node has enough traffic to justify the cost of a large satellite station, its traffic is probably high enough and consequently, sufficiently "smooth" to warrant its own satellite channel. On the other hand, an earth station for a domestic satellite system (such as Anik and future U. S. systems) can use a 30-foot antenna which costs from $150,000 upwards. This figure is comparable to the costs of peripheral devices in present large computer installations. In a recent study [25], even smaller earth stations (with antenna between 10 to 15 feet) were suggested. The annual cost per station was estimated to be approximately $5,000 to $15,000.

We also note that there is an existing regulatory restriction on the use of an INTELSAT IV channel in the multi-access broadcast mode for several stations. Discussions are under way with various agencies to remove these regulatory barriers in either the INTELSAT system or one of the domestic systems [3].

With domestic satellite systems, data rates are not limited to that of a single voice channel. For example, data rates ranging up to 60 MBPS will be available over the American Satellite Corporation system. Furthermore, specialized network configurations will be available to suit a user's customized requirements [20].

We quote the following remarks on projected satellite technology cost trends by Roberts [11]:

> "Although terrestrial communications cost appears to limit the future price of computer-communication service, including packet-switching networks, the situation is rapidly changing with the introduction of domestic satellites... . Applying the least-mean-square exponential fit to this data, the rate of technological improvement in the cost performance of satellites is found to be 40.7 percent per year, or a factor of ten every 6.7 years. This can only be treated as a crude estimate of the cost trend for satellite communication, but since it is quite in keeping with the general cost trend for electronics, it is a quite credible growth rate... . Satellites will play an important role in reducing the future cost of packet switching service... ."

2.3 An Abstract Model

In order to evaluate the performance of these communication systems via model building and theoretical analysis or simulation, it is desirable to define abstract models which include only the salient properties and operational features. We define the following models for the multi-access broadcast channel and its users.

We are considering a satellite communication channel in which the satellite is used as a transponder; messages are received (uplink) from the terminals on one frequency band, and then retransmitted (downlink) back to the terminals on a different frequency band. The satellite must be maintained in a geosynchronous (stationary) orbit at roughly 36000 km above the earth. This long transmission path results in a round-trip propagation delay that is approximately 0.27 s for terminals near the horizon.

Since users of this channel are in general geographically distributed, we assume a global reference time called channel time. We assume fixed size packets. Channel time is *slotted* such that all users synchronize their packet transmissions into channel slots. A channel slot length is exactly equal to the duration of a packet transmission. Any guard time required to separate pocket transmissions in the channel is neglected. From now on, time will be expressed in channel slots. All rates will be normalized with respect to a channel time slot.

The channel traffic (or throughput or utilization factor) in a channel time slot is a random variable representing the total number of a packet transmissions by all users in that time slot. The channel traffic rate is the average number of packet transmissions per time slot (assuming stationary conditions).

CHAPTER 3

PACKET SWITCHING TECHNIQUES

3.1 Existing Multiaccess Protocols

Multiaccess protocols differ by the static of dynamic nature of the bandwidth allocation algorithm, the centralized or distributed nature of the decision-making process, and the degree of adaptivity of the algorithm to changing needs. Accordingly, these protocols can be grouped into three classes. The first class, labeled fixed assignment techniques, consists of those techniques which allocate the channel bandwidth to the users in a static fashion, independently of their activity. The second class is that of random access techniques. In this class the entire bandwidth is provided to the users as a single channel to be accessed randomly; since collisions may result which degrade the performance of the channel, improved performance can be achieved by either synchronizing users so that their transmissions coincide with the boundaries of time slots, by sensing carrier prior to transmission, or both. The third class corresponds to demand assignment techniques. Demand assignment techniques require that explicit control information regarding the users' need for the communication resource be exchanged.

The various protocols known today, either implemented or proposed, and their performance and applicability to the satellite communication environment are discussed in Reference 26 and summarized below. For this we consider the (conceptually) simplest situation consisting of M users wishing to communicate over a channel. This situation arises typically in a satellite communication environment.

3.1.1 Fixed Assignment Techniques

Fixed assignment techniques consist of allocating the channel to the user, independently of their activity, by partitioning the time-bandwidth space into slots which are assigned in a static predetermined fashion. These techniques take two common forms: frequency division multiple access (FDMA) and synchronous time division multiple access (TDMA).

FDMA and TDMA: FDMA consists of assigning to each user a fraction of the bandwidth and confining its access to the allocated subband. Orthogonality is achieved in the frequency domain. FDMA is relatively simple to implement and requires no real time coordination among the users. TDMA consists of assigning fixed predetermined channel time slots to each user; the user has access to the entire channel bandwidth, but only during its allocated slots. Here, signaling waveforms are orthogonal in time.

A number of disadvantages exist for FDMA when compared to TDMA. FDMA wastes a fraction of the bandwidth to achieve adequate frequency separation. FDMA is also characterized by a lack of flexibility in performing changes in the allocation of the bandwidth and certainly the lack of broadcast operation. The major disadvantages in TDMA are the need to provide A/D converters for overlap traffic such as voice, and rapid burst synchronization and sufficient burst separation to avoid time overlap. TDMA is more complex to implement than FDMA, but an important advantage is the connectivity which results from the fact that all receivers listen to the same channel while senders transmit on the same common channel at different times. Accordingly, many network realizations, both in ground and satellite environments, are easier to accomplish [27].

From the performance standpoint it has also been established that TDMA is superior to FDMA in many cases of practical interest. I. Rubin has shown that the random variable representing packet delay is always larger in FDMA than in TDMA [28] for comparable systems. Lam derived the average message delay for a TDMA system with multipacket messages and a nonpreemptive priority queue discipline [29]. There, too, it was shown that TDMA is superior to FDMA.

For both FDMA and TDMA, the fixed preallocation of the frequency or time resource does not have to be equal for all users, but can be tailored to fit their needs (assumed constant). Kosovych studied two TDMA implementations [30]. In the first, called contiguous assignment, the users are cyclically ordered in the time sequence in which they have access to the channel.

Each user is periodically assigned its own fixed time duration. In the second implementation, called distributed allocation, all access periods are of equal time duration, but the frequency of accesses can be different from one user to the other. It was shown that for situations in which the transmission overhead (defined as guard time and synchronization preamble time) is large, the contiguous fixed assignment implementation is better suited and provides substantially better performance than distributed fixed assignments, while when the transmission overhead is small, distributed fixed assignments provide slightly better performance.

Finally we note that, even though the allocation can be tailored to the relative need of each user, fixed allocation can be wasteful if the users' demand is highly bursty, as we shall explicitly see in the sequel. Given these limitations, one may increase the channel utilization beyond FDMA and TDMA by using asynchronous time division multiple access (ATDMA), also known as statistical multiplexing [8]. Basically the technique consists of switching the allocation of the channel from one user to another only when the former is idle and the latter is ready to transmit data. Thus the channel is dynamically allocated to the various users according to their need. The performance of ATDMA in packet communication systems corresponds to that of a work-conserving single server queueing system, and is the best we can achieve under unpredictable demand. Unfortunately, it is not always possible to accomplish the necessary coordination among the users. This mode of multiplexing is possible only when several calocated users (such as at the same earth station) are sharing a single point-to-point channel.

3.1.2 Random Access Techniques

As stated in the introduction, packet communication is a natural means to achieve sharing of the common channel. When dealing with shared channels in a packet-switched mode, one must be prepared to resolve conflicts which arise when more than one demand is placed upon the channel. For example, in packet-switched radio channels, whenever a portion of one user's transmission overlaps with another user's transmission, the two collide and "destroy" each other (unless a code division multiple-access scheme is used). The existence of some positive acknowledgment scheme permits the transmitter to determine if his transmission is successful or not. The problem is how to control the access to the common channel in a fashion which produces, under the physical constraints of simplicity and hardware implementation, an acceptable level of performance. The difficulty in controlling a channel which must carry its own control information has given rise to the so-called random-access protocols, among others. We describe these here by considering again single-hop environments.

ALOHA[32]–[33]: Historically, the pure ALOHA protocol was first used in the ALOHA system, a single-hop terminal access network developed in 1970 at the University of Hawaii, employing packet-switching on a radio channel [31], [16]. The simplest of its kind, pure ALOHA permits a user to transmit any time it desires. If they do so, and within some appropriate time-out period it receives an acknowledgment from the destination (the central computer), then it knows that no conflict occurred. Otherwise it assumes that a collision occurred and it must retransmit. To avoid continuously repeated conflicts, the retransmission delay is randomized across the transmitting devices, thus spreading the retry packets over time. A slotted version, referred to as slotted ALOHA, is obtained by dividing time into slots of duration equal to the transmission time of a single packet (assuming constant-length packets) [32], [33]. Each user is required to synchronize the start of transmission of its packets to coincide with the slot boundary. When two packets conflict, they will overlap completely rather than partially, providing an increase in channel efficiency over pure ALOHA. Due to conflicts and idle channel time, the maximum channel efficiency available using ALOHA is less than 100 percent, 18 percent for pure ALOHA and 36 percent for slotted ALOHA. Both schemes are theoretically applicable to satellite, ground radio and local bus environments. The slotted version has the advantage of efficiency, but in multihop ground radio, it has the disadvantage that synchronization may be hard to achieve.

3.1.3 Demand Assignment Techniques

There are two reasons why demand assignment with distributed control is desirable in satellite communications. The first is reliability; with distributed control the system is not dependent on the proper operation of a central scheduler. The second is improved performance, especially when dealing with systems with long propagation delays. Indeed, if an earth station were to play the role of a scheduler, the minimum packet delay in a packet reservation scheme would be three times the round-trip propagation delay. (Of course, this can be decreased if on-board processing is available.) With distributed control, this minimum delay can be brought down to twice the round-trip delay or less without affecting the bandwidth utilization. Clearly, in slotted ALOHA, the best random access scheme available for satellite channels, the minimum packet delay is exactly one round-trip delay; but this is guaranteed only for a channel utilization approaching zero. In fact, the inherent long propagation delay in satellite channels is really the nasty characteristic that makes this environment "more distributed" than the single-hop ground radio or local area environments.

The basic element underlying all distributed algorithms is the need to exchange control information among the users, either explicitly or implicitly. Using this information, all users then execute independently the same algorithm resulting in some coordination in their actions. Clearly, it is essential that all users receive the same information regarding the demand placed on the channel and its usage in order to achieve a global optimum, and thus distributed algorithms are most attractive in fully connected systems. This attribute is not always present in ground radio environments, but certainly exists in satellite environments due to their inherent broadcast nature. The long-delay/broadcast combination of attributes has been one of the reasons why many distributed control algorithms have been proposed in the context of satellite environments. We examine in this subsection distributed control algorithms suitable for satellite channels.

Reservation--ALOHA [34]: Reservation-ALOHA for a satellite channel is based on a slotted time axis, where the slots are organized into frames of equal size. The duration of a frame must be greater than the satellite propagation delay. A user who has successfully accessed a slot in a frame is guaranteed access to the same slot in the succeeding frame and this continues until the user stops using it. "Unused" slots, however, are free to be accessed by all users in a slotted ALOHA contention mode. An unused slot in the current frame is a slot which, in the preceding frame, either was idle or contained a collision. (Note again the effect of long delays on the control procedure.) Users need to simply maintain a history of the usage of each slot for just one frame duration. Since no request is explicitly issued by the user, this scheme has been referred to as an implicit reservation scheme. Clearly Reservation-ALOHA is effective only if the users generate stream type traffic or long multipacket messages. Its performance will degrade significantly with single packet messages, as every time a packet is successful the corresponding slot in the following frame is likely to remain empty.

Roberts' Scheme [35]: In this scheme, reservations are made explicitly. Time division is used to provide a reservation subchannel. The channel time is slotted as before, but every so often a slot is divided into V small slots which are used for the transmission of reservation packets (as well as possibly acknowledgments and small data packets); these packets contend on the V small slots in a slotted ALOHA mode. All other slots are data slots and are used on a reservation basis, free of conflict. The frequency of occurrence of reservation slots can be made adaptive to the load on the channel and the need to make new reservations. This adaptivity can be achieved as a result of the time-division of bandwidth allocation between reservations and data packets.

To execute the reservation mechanism properly, each station must maintain information on the number of outstanding reservations (the "queue in the sky") and the slots at which its own reservations begin. These are determined by the FIFO discipline based on the successful reservations received. Each successful reservation can accommodate up to a design maximum of, say, eight packets, thus preventing stations from acquiring exclusivity of the channel for long periods of time. To maintain synchronization of control information at the proper time, and to acquire the correct count of packets in the queue if out-of-sync conditions do occur, each station sends, in its data packet, information regarding the status of its queue. This information is also used by new stations which need to join the queue. The robustness of this system is achieved by a proper encoding of the reservation packets to increase the probability of their correct reception at all stations. Furthermore, to limit the effect of errors, a station reacquires synchronization if it detects a collision in one of its reserved slots or an error in a reservation packet. FIF0-Reservation offers delay improvements over TMDA. When compared to ALOHA, higher system capacity is achieved but at the expense of a higher delay at low channel throughouts (due to a higher overhead).

Binder's Scheme [36]: The basis of this scheme is fixed TDMA assignment, but with the major difference that "unused" slots are assigned to the active stations on a round-robin basis. This is accomplished by organizing packet slots into equal size frames of duration greater than the propagation delay and such that the number of slots in a frame is larger than the number of stations. One slot in each frame is permanently assigned to each station. To allow other stations to know the current state (used or unused) of its own slot, each station is required to transmit information regarding its own queue of packets piggy backed in the data packet header (transmitted in the previous frame.) A zero count indicates that the slot in question is free. All stations maintain a table of all stations' queue lengths, allowing them to allocate among themselves free unassigned slots in the current frame. A station recovers its slot by deliberately causing a conflict in that slot which other users detect. For a station which was previously idle, initial acquisition of queue information is required and is achieved by having one of the stations transmit its table at various times. However, it is interesting to note that in this scheme, while acquiring queue synchronization, a station can always reclaim and use its own assigned slot.

Balgangadhar and Pickholtz(B&P)'s Scheme [37]: In this proposed scheme, time is divided into frames, each frame consisting of reservation time slots, preassigned (PA) time slots and reservation access (RA) time slots. At the beginning of each frame, there are N small slots, each large enough to transmit a reservation from one of the N stations in the network. Following the reservation slots, there are N PA slots, with each station being permanently allocated one PA slot per frame. The slot size and the number N of nodes, as in Binder's scheme, is such that the duration of these N PA slots is at least as long as one round-trip delay to the satellite. At the end of these PA slots come the reservation access slots. The number of RA slots in each frame depends on the reservation vector sent by the nodes at the beginning of each frame. The idea behind this scheme is to make efficient channel utilization (hence, the asynchronous RA slots) but circumvent the delay problem by introducing the PA slots, forcing the frame size to be no smaller than one round-trip delay.

3.2 A Proposed New Reservation Technique:

The proposed scheme consists of a dynamic time-assignment system which retains the allocation fairness while allowing stations to make use of channel capacity otherwise wasted due to light traffic loads and/or a nonuniform distribution of traffic among the nodes. In this proposed scheme, time is divided into frames, each frame consisting of reservation time slots, preassigned fixed time slots and reservation-based variable time slots. At the beginning of each frame, there are M small slots (M is the number of users in the system), each large enough to transmit a reservation from one of the M nodes in the network. Following these small reservation slots, there are M preassigned fixed slots, with each node being permanently allocated one fixed slot per frame. The slot size and the number of nodes, as in Binder's Scheme, is such that the duration of these M fixed slots is at least as long as one round-trip delay to the satellite. At the end of these fixed slots come the reservation based variable slots (V). The number of these variable slots in each frame depends on the reservation vectors sent by the nodes at the beginning of the frames and the algorithm described in the following discussion. The idea behind this scheme is to make efficient channel utilization (hence, the asynchronous variable slots) but circumvent the delay problem by introducing the fixed slots, forcing the frame size to be no smaller than one round-trip delay. One major advantage of this scheme is that the reservation time is small enough to accommodate two bits of reservation vector for each node and thus a very small percentage of the total frame time, which is not true for other existing schemes. Figure 3.1 illustrates the frame structure and time-slot allocation.

The reservation vector is sent by each node as follows: The nodes examine the buffer contents (number of packets opened) just prior to sending the reservation information, subtract one from it (since they know that they will get one fixed slot in the frame) and sends two bits (0,0) (or only 0 in this case) or (0,1) or (1,0) or (1,1) into the two reservation slots assigned to it. The (0,0) vector represents the situation where the station does not have any packet for transmission, the (0,1) vector represents the station having only one packet to transmit, the (1,0) vector represents the station having only two packets to transmit, and the (1,1) vector represents the station having more than two packets to transmit. The number of variable slots assigned to each station in a time frame will be determined by the following algorithm:

A. There will be no time slot allotted for the station which has sent a (0,0) vector (or only '0' in this case).

B. There will be one time slot allotted for transmission for the station which has sent a (0,1) vector.

C. There will be two time slots allotted for transmission for the station which has sent a (1,0) vector.

D. The station which has sent a (1,1) vector will be allotted time slots (equal to or more than two) which will depend on its prior reservation vectors as follows:

   1. P (which is equal to or more than two) time slots will be reserved for transmission if the first preceding reservation vector is (0,0), or (0,1) or (1,0).

2. Q (which is equal to or more than P) time slots will be reserved for transmission if the first preceding vector is also (1,1) but the second preceding vector is (0,0) or (0,1) or (1,0).
3. R (which is equal to or more than Q) time slots will be reserved for transmission if the preceding two vectors are both (1,1).

The above algorithm is explained in Table 3.1.

Immediately, after all the nodes have transmitted their reservation, each node transmits (if it has any packets in its buffer), sequentially, one packet on the one fixed slot assigned to it. Since the duration of the M fixed slots is longer than a round-trip delay, by the end of the fixed slots all the nodes will have received the information for the variable slots assigned to them. Accordingly, each node will transmit its assigned packets, sequentially, in the variable portion of the time frame.

It is to be noted that the slot allocation is highly asymmetrical with regard to the terminals. It is intuitively obvious that the packets in station n have to wait longer than the packets of station (n-1) to get transmitted. To overcome this problem, the assignment of the slots can be done in a cyclic order: in frame (j), the sequence of slot allocations is (1,2,...,M); in frame (j+1), the sequence is changed to (2,3,...,M,1), and so forth; in frame (j+M+1), the sequence is again (1,2,...,M). In such a TDMA based reservation slot the reservation minipackets are quite short, because they would not have to carry identity and synchronization information.

The length of the frame is a random variable and depends on the number of users. requesting transmission at the beginning of the frame. The mathematical model in section 4 is centered primarily around the imbedded Markov chain, imbedded at the special epochs which are the beginnings of the frames. The service order within the frame has no effect on the state of the system at these epochs. But the service order does affect the waiting time distribution for the packets queued at the terminals.

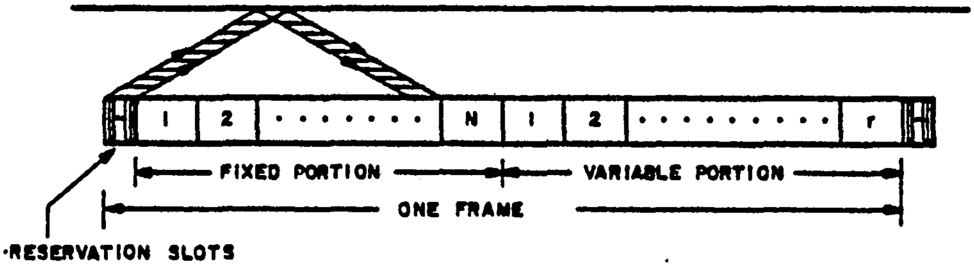

FIGURE 3·1·a· FRAME STRUCTURE OF PROPOSED SCHEME

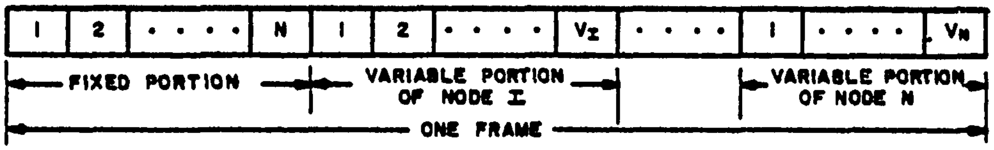

FIGURE 3·1·b DETAILED TIME–SLOT ALLOCATION IN THE PROPOSED SCHEME

(NOTE: IN THE MATHEMATICAL MODEL, THE SMALL SEGMENTS AT THE BEGINNING OF A FRAME, WHERE THE RESERVATION VECTORS ARE SENT BY THE NODES, ARE ASSUMED TO BE OF ZERO DURATION)

TABLE 3·1

ALGORITHM FOR SLOT ASSIGNMENT

RESERVATION VECTORS SENT BY EACH NODE AT THE BEGINNING OF A FRAME			NO. OF RESERVATION BASED VARIABLE SLOTS (V) ASSIGNED TO EACH NODE IN A TIME FRAME
PRESENT VECTOR	FIRST PRECEDING VECTOR	SECOND PRECEDING VECTOR	
(0,0)	—	—	0
(0,1)	—	—	1
(1,0)	—	—	2
(1,1)	(0,0) OR (0,1) OR (1,0)	—	P
(1,1)	(1,1)	(0,0) OR (0,1) OR (1,0)	$Q \geq P$
(1,1)	(1,1)	(1,1)	$R \geq Q$

CHAPTER 4

A MATHEMATICAL MODEL OF THE PROPOSED SCHEME

4.1 Introduction

The proposed system is a non-standard multi-queuing single-server queuing system. Multiqueues attended by a single-server have received a good deal of attention in the queuing theory literature [38-41]. What distinguishes the proposed system is primarily the service discipline.

Chu and Konheim[42] have developed a unified mathematical model for the analysis of synchronous TDM, the asynchronous TDM, and the "Hub polling" techniques. Specifically, they assume that the arrival pattern of packets at the stations is Poisson and, using a generating function approach, have developed equations leading to the queue size distributions of the buffers at the model and the waiting times experienced by the packet. For the STDM and ATDM systems, the generating functions of queue size distributions are explicitly obtained and the moments derived by differentiation. For the Hub Polling System, a functional equation is obtained for the generating function, solution of which is developed in an infinitive product form, and the first and second moments of the queue sizes (and the waiting times) are obtained explicitly.

It is to be noted that the service time of a packet is dependent on the status of the remaining queues. If we let n_k represent the contents of the input buffer of the kth user, then a complete description of all M users requires the specification of the joint probability distribution $P(n_1, n_2, \ldots, n_M)$. The determination of these probabilities is very complex, if not impossible and demands the solution of a large number of sets of equations. Fayolle and Iasnogorodski [43] showed the limitation of this direct approach. They considered the simple example of two parallel M/M/1 queuing systems with infinite capacities and with service rate for each system depending on the status of the other system's queue. They showed that the generating function F(x, y), corresponding to the joint probability of the two queue sizes can be continued as a meromorphic function to the whole complex plane. Using the theory of analytic continuation they reduced the problem to a Rieman-Hilbert problem and were able to obtain a closed form for F(x, y) which includes several elliptical integrals of the third kind. Extension of their analysis to the case of more than two users and to our problem seems unfeasible.

Leibowitz [44] presents an approximate method of treating multiqueue systems. He studied the case of N queues with a single server, in which the queues are served in cyclic order. He used the following argument to derived the probability distribution of the queues; if the server meets the same probability distribution of the number of customers, say P_n, at each of the queues on one cycle of the system, then the same distribution must meet him when he returns to the same queue.

Hashida [45] used Leibowitz's approximate approach in the analysis of multiqueue systems where the server serves, at most, K customers that were waiting when he arrived at a queue. Also in Reference 37, the authors used Leibowitz's approximation along with an independence assumption to obtain the queue size distribution for each user.

We consider that the main contribution of this dissertation, besides introducing a new reservation scheme, is the mathematical model presented here to analyze interacting buffered terminals. The model, as mentioned earlier, is basically a combination of two Markov chains imbedded at the beginning of each frame; one Markov chain that describes the state of the user, referred to as the User Markov Chain and another Markov chain that describes the state of all the users in the systems, referred to as the System Markov Chain. T. T. Saadawi and A. Ephremides [46] used a similar approach to analyze the reservation scheme where they considered that each station transmits maximum one packet in a time frame.

In Section 4.2.1 we describe the system Markov chain while in Section 4.2.2 the user Markov chain is obtained. For the purpose of comparison with simulation results, the mathematical model is studied here for the reservation scheme where each station is allowed to transmit maximum two packets in a ti.e., frame i.e. P = Q = R = 1. The analysis for the situation where we allow maximum one packet for each station in a time frame (i.e. $P = Q = R = 0$), is shown in the appendix. For other values of P, Q and R we will have similar analyses for the system.

4.2 The User Model

The system consists of M terminals, each of which has an infinite buffer. The arrival process at each of the M terminals is assumed to be a Bernoulli process with rate σ, i.e., the probability of arrival of a (single packet) message at any terminal in each slot is σ. The total arrival rate is, therefore, $M\sigma$ packets per slot, which is equal to the throughput rate. Most other studies of multiple access protocals have assumed a Poisson arrival process. The Bernoulli process is a discrete-time analog of the Poisson process, which is well-suited to the discrete time slot structure. The user may be in one of two states: an idle user, where his buffer is empty; or an active user, where his buffer is not empty.

As shown in Fig. 3.1 whenever the user has the packet(s) in his buffer, it sends a reservation request at the beginning of the frame*, the preassigned packet is transmitted to the fixed time slot and the reservation based packet(s) are transmitted to the assigned slot(s) of the variable portion of the frame after it receives messages from the satellite at the end of a round-trip delay.

We need the following definitions;

$\pi_i =$ steady state probability of having i packets in the buffer at the beginning of a frame.

Hence

---

* In the mathematical model, the small segments at the beginning of a frame, where the reservation vectors are sent by the nodes, are assumed to be of zero duration.

π_0 = probability of an empty buffer

= probability of an idle user

$1 - \pi_0$ = probability of an active user

4.2.1 The System Marko v Chain

The state of the system is described by the number of users requesting transmission at the beginning of the frame. Let us first define the following:

j_1 = number of users having only one packet requesting transmission at the beginning of a frame

j_2 = number of users having more than one packet requesting transmission at the beginning of a frame

L_{j_1,j_2} = Length of a frame resulted from requests for transmission of j_1 and j_2 users at its beginning

= $(M+j_2)$ slots, where M is equal to or greater than one round-trip delay in time slots

P_{j_1,j_2} = Steady state probability of having j_1 and j_2 users requesting transmission at the beginning of a frame

σ_{j_0} = Probability of an idle user generating no packet during the frame of length L_{j_1,j_2}

$$= \begin{pmatrix} M+j_2 \\ 0 \end{pmatrix} (1-\sigma)^{M+j_2} \tag{4.1a}$$

σ_{j_1} = Probability of an idle user generating one packet during the frame of length $L_{j_1 j_2}$

$$= \begin{pmatrix} M+j_2 \\ 1 \end{pmatrix} \sigma\ (1-\sigma)^{M+j_2-1} \tag{4.1b}$$

σ_{j_2} = Probability of an idle user generating more than one packet during the frame of length $L_{j_i j_2}$

$$= 1 - \sigma_{j_0} - \sigma_{j_1} \tag{4.2}$$

F_0 = Probability of having two packets only in the buffer at the beginning of a frame given user is active and having more than one packet

$$= \frac{\pi_2}{1-\pi_0-\pi_1} \tag{4.3}$$

F_1 = Probability of having three packets in the buffer at the beginning of a frame given user is active and having more than two packets

$$= \frac{\pi_3}{1-\pi_0-\pi_1-\pi_2} \tag{4.4}$$

K_1 = number of users having one packet remain active out of j_2 users requesting transmission at the beginning of a frame.

K_2 = number of users having more than one packet remain active out of j_2 users requesting transmission at the beginning of a frame

$M - K$ = number of idle users at the beginning of a frame where $K = K_1 + K_2$

$C(K_1, K_2 j_2)$ = Pr [K_1 users having one packet and K_2 users having more than one packet remain active out of j_2 users requesting transmission at the beginning of a frame]

$$= \binom{j_2}{K_1+K_2}(1-F_0)^{K_1+K_2}(F_0)^{j_2-K_1-K_2} \cdot \binom{K_1+K_2}{K_2}(1-F_1)^{K_2}F_1^{K_1} \tag{4.5}$$

In the above binomial distribution, it is assumed that the probability a user remains active is independent from the state of the other users. A similar assumption has been stated and justified in Reference 45, where the case of a single server serving N lines in cyclic order is studied. In that analysis, the number of lines is assumed constant during a cycle and is determined by a binomial distribution.

$q(i_1,i_2,K_1,K_2,j_2) =$ Pr [i_1 idle users receive one packet and i_2 idle users receive more than one packet out of (M-K) idle users during a frame given $K(=K_1+K_2)$ users remain active out of j_2 users requesting transmission at the beginning of a frame]

$$= \binom{M-K}{i_1+i_2-K} \cdot \left(1-\sigma_{j_0}\right)^{i_1+i_2-k} \cdot \left(\sigma_{j_0}\right)^{M-i_1-i_2}$$

$$\times \binom{i_1+i_2-K}{i_2-K_2} \cdot \left[\frac{1-\sigma_{j_0}-\sigma_{j_1}}{1-\sigma_{j_0}}\right]^{i_2-K_2} \cdot \left[\frac{\sigma_{j_1}}{1-\sigma_{j_0}}\right]^{i_1-K_1} \tag{4.6}$$

The system Markov chain is illustrated in Fig. 4.1a. Therefore, P_{n_1,n_2}, the steady state probability of n_1 users having one packet and n_2 users having more than one packet, requesting transmission at the beginning of the frame is given by

$$P_{n_1,n_2} = \sum_{j_1=0}^{M} \sum_{j_2=0}^{M-j_1} A_{j_1j_2}(n_1,n_2)P_{j_1j_2} \tag{4.7}$$

where $A_{j_1j_2}(n_1,n_2)$ is the transition probability from state (j_1,j_2) to state (n_1,n_2) and is given by

$$A_{j_1j_2}(n_1,n_2) = \sum_{k_2=0}^{\min(j_2,n_2)} \sum_{K_1=0}^{\min(j_2-K_2,n_2)} \Pr[K = (K_1+K_2) \text{ users remain active,}$$

(n_1-K_1) idle users receive one packet only, and (n_2-K_2) idle users receive more than one packet during a frame, given $j(=j_1+j_2)$ users request transmission at the beginning of a frame.]

$$= \begin{cases} \sum_{k_2=0}^{\min(j_2,n_2)} \sum_{K_1=0}^{\min(j_2-K_2,n_2)} q(n_1-K_1,n_2-K_2,K_1,K_2,j_2)C(K_1,K_2,j_2) , & \text{for } 1 \leq j = j_1+j_2 \leq M, \\ & \text{and } 0 \leq n = n_1+n_2 \leq M \\ \binom{M}{n_1} \sigma^{n_1}(1-\sigma)^{M-n_1}, & \text{for } j=0, n_2=0, \\ & \text{and } 0 \leq n = n_1+n_2 \leq M \end{cases} \quad (4.8)$$

4.2.2 The User Markov Chain

The state of the user is described by the number of packets in the buffer at the beginning of the frame.

Let

$n =$ number of packets in the buffer at the beginning of a frame

$\pi_n =$ steady state probability of having n packets in the buffer at the beginning of a frame

In Fig. 4.1.b, we show the user Markov chain. Note that $B_{n-i}(n)$ is the transition probability from state (n-i) to state (n). Let us now determine these transition probabilities. Let

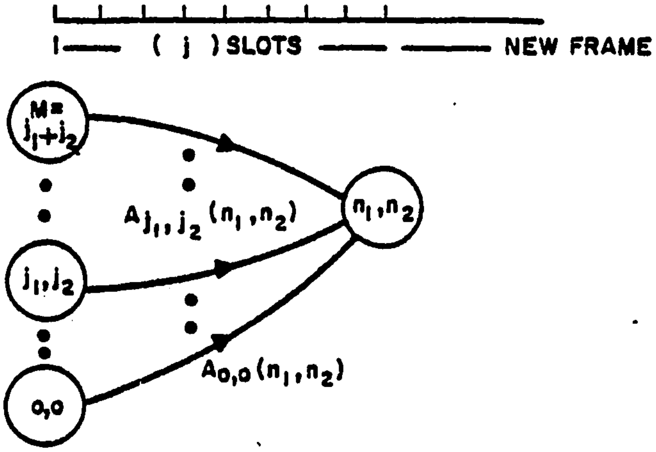

FIGURE 4-1a. SYSTEM MARKOV CHAIN

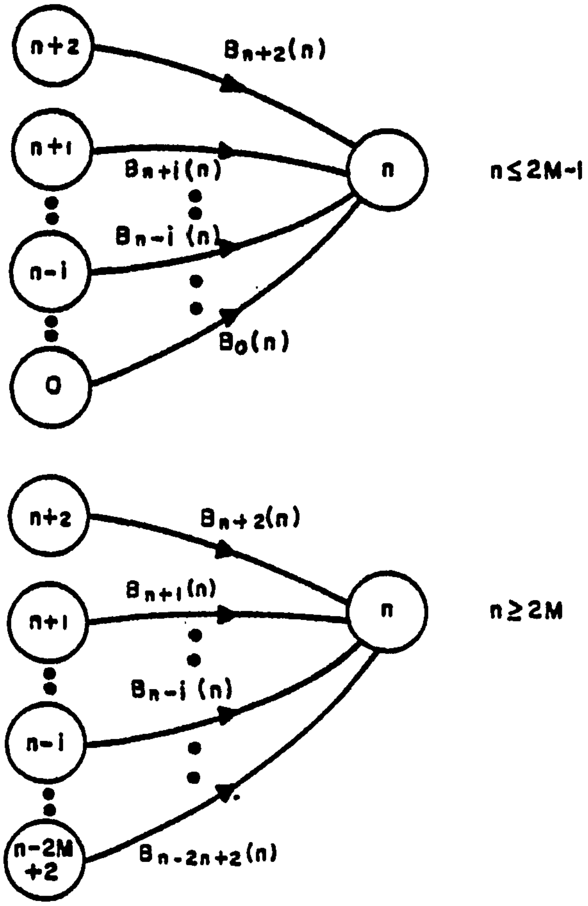

FIGURE 4-1b. USER MARKOV CHAIN

$g(i,L_{j_1 j_2})$ = Pr[i packets arrive during the frame of length $L_{j_1 j_2}$]

$$= \binom{M+j_2}{i} \sigma^i (1-\sigma)^{M+j_2-i} \tag{4.9}$$

$B_{n+2}(n)$ = Pr[Two packets leave and no packet arrives during the frame of length $L_{j_1 j_2}$]

$$= \sum_{j_2=1}^{M} \sum_{j_1=0}^{M-j_2} g(0,L_{j_1 j_2}) P_{j_1 j_2} \Big/ \left[1-\sum_{j_1=0}^{M} P_{j_1,0}\right], \qquad n = 0,1,2\,. \tag{4.10}$$

Similarly,

$B_{n+1}(n)$ = Pr[Two packets leave and one packet arrives during the frame of length $L_{j_1 j_2}$]

$$= \sum_{j_2=1}^{M} \sum_{j_1=0}^{M-j_2} g(1,L_{j_1 j_2}) P_{j_1 j_2} \Big/ \left[1-\sum_{j_1=0}^{M} P_{j_1,0},\right], \qquad n = 1,2,\cdots \tag{4.11}$$

and

$B_1(0)$ = Pr[One packet leaves and no packet arrives during the frame of length $L_{j_1 j_2}$]

$$= \sum_{j_1=1}^{M} \sum_{j_2=0}^{M-j_1} g(0,L_{j_1,j_2}) P_{j_1,j_2} \Big/ \left[1-\sum_{j_2=0}^{M} P_{0 j_2}\right] \tag{4.12}$$

Similarly,

$B_n(n)$ = Pr[Two packets leave and two packets arrive during the frame of length $L_{j_1 j_2}$]

$$= \sum_{j_2=1}^{M} \sum_{j_1=0}^{M-j_2} g(2, L_{j_1 j_2}) P_{j_1 j_2} \Big/ \left(1 - \sum_{j_1=0}^{M} P_{j_1,0}\right), \qquad n = 2,3,\cdots \tag{4.13}$$

$B_1(1)$ = Pr[One packet leaves and one packet arrives during the frame of length $L_{j_1 j_2}$]

$$= \sum_{j_1=1}^{M} \sum_{j_2=0}^{M-j_1} g\left(1, L_{j_1 j_2}\right) P_{j_1 j_2} \Big/ \left(1 - \sum_{j_2=0}^{M} P_{0 j_2}\right) \tag{4.14}$$

and

$B_0(0)$ = Pr[No packet arrives during the frame of length of $L_{j_1 j_2}$]

$$= \sum_{j_2=0}^{M-1} \sum_{j_1=0}^{M-1-j_2} g(0, L_{j_1 j_2}) P_{j_1 j_2} \Big/ \left(1 - \sum_{j_2=0}^{M} P_{M-j_2 j_2}\right) \tag{4.15}$$

Similarly,

$B_{n-i}(n) = \Pr[\text{Two packets leave and } (i+2) \text{ packets arrive during the frame of length } L_{j_1 j_2}]$

$$= \sum_{j_2=1}^{M} \sum_{j_1=0}^{M-j_2} g\left(i+2, L_{j_1 j_2}\right) P_{j_1 j_2} \Big/ \left(1-\sum_{j_1=0}^{M} P_{j_1,0}\right), \qquad (4.16)$$

$$n = 3,4,\ldots$$

$$i = 1,\min(n-2,2M-2)$$

$B_1(n) = \Pr[n \text{ packets arrive during the frame of length } L_{j_1 j_2}]$

$$= \sum_{j_1=1}^{M} \sum_{j_2=0}^{M-j_1} g(n, L_{j_1 j_2}) P_{j_1 j_2} \Big/ \left(1-\sum_{j_2=0}^{M} P_{0 j_2}\right), \qquad (4.17)$$

$$n = 2,\ldots,\ 2M-1$$

and

$B_0(n) = \Pr\ [n \text{ packets arrive during the frame of length } L_{j_1 j_2}]$

$$= \sum_{j_2=0}^{M-1} \sum_{j_1=0}^{M-1-j_2} g(n, L_{j_1 j_2}) P_{j_1 j_2} \Big/ \left(1-\sum_{j_2=0}^{M} P_{M-j_2 j_1}\right),$$

$$n = 1,\ldots,2M-1 \qquad (4.18)$$

We now can have a numerical solution for $\pi_n, n = 0,1,2,3,\ldots$ as described below.

4.2.3 Numerical Solutions

Equations (4.1) through (4.6) provide the values of the quantities that were needed in order to solve the equations (4.7). However, some of these quantities are not expressed in terms of only the system parameter M, σ, but, in terms of F_0 and F_1 which are function of π_0, π_1, π_2 and π_3, which, in turn, on $[P_{n_1 n_2}]$. Thus, in total, we have a set of simultaneous, coupled non-linear equations in π_n and $[P_{n_1,n_2}]$. In Fig. 4.2, we show the flow chart to solve for these values. We start with an initial value for F_0 and F_1, then solve equation (4.7) in $[P_{n_1,n_2}]$. Using the values of P_{n_1,n_2} we then determine the steady state probabilities π_n which satisfy the following matrix equation:

$$\pi_n = \pi_n B \tag{4.19}$$

where B is the transition probability matrix with elements $b_{ij} = B_i(j)$.

The solution of n equations in n unknowns in equation (4.19) is straightforward using matrix inversion* and equations (4.9) through (4.18). Then using Wegestein's iteration scheme [47], we obtain the new F_0 and F_1 and repeat it until it converges to a solution for π_n. A FORTRAN program was written to solve π_n. The analytical results are compared with the simulation results in the next chapter.

4.2.4 Delay Analysis

The average total delay per packet, D, is the sum of the average waiting time in the buffer, W, plus the average time from the beginning of the frame till the packet is transmitted in its assigned slot. We refer to the latter as the service time, S. Hence,

$$D = W + S \tag{4.20}$$

Using Little's result [48], we can write W as;

$$W = \frac{Q}{\sigma} \tag{4.21}$$

where,

$$Q = \text{Average Queue Size}$$
$$= \sum_{n=0}^{\infty} n \cdot \pi_n$$

Assuming cyclic assignment discussed earlier in Section 3.2, the average service time is given by

---

* When n is very large, it might be difficult to invert the matrix B. In that case, other standard methods should be used to solve equation (4.19).

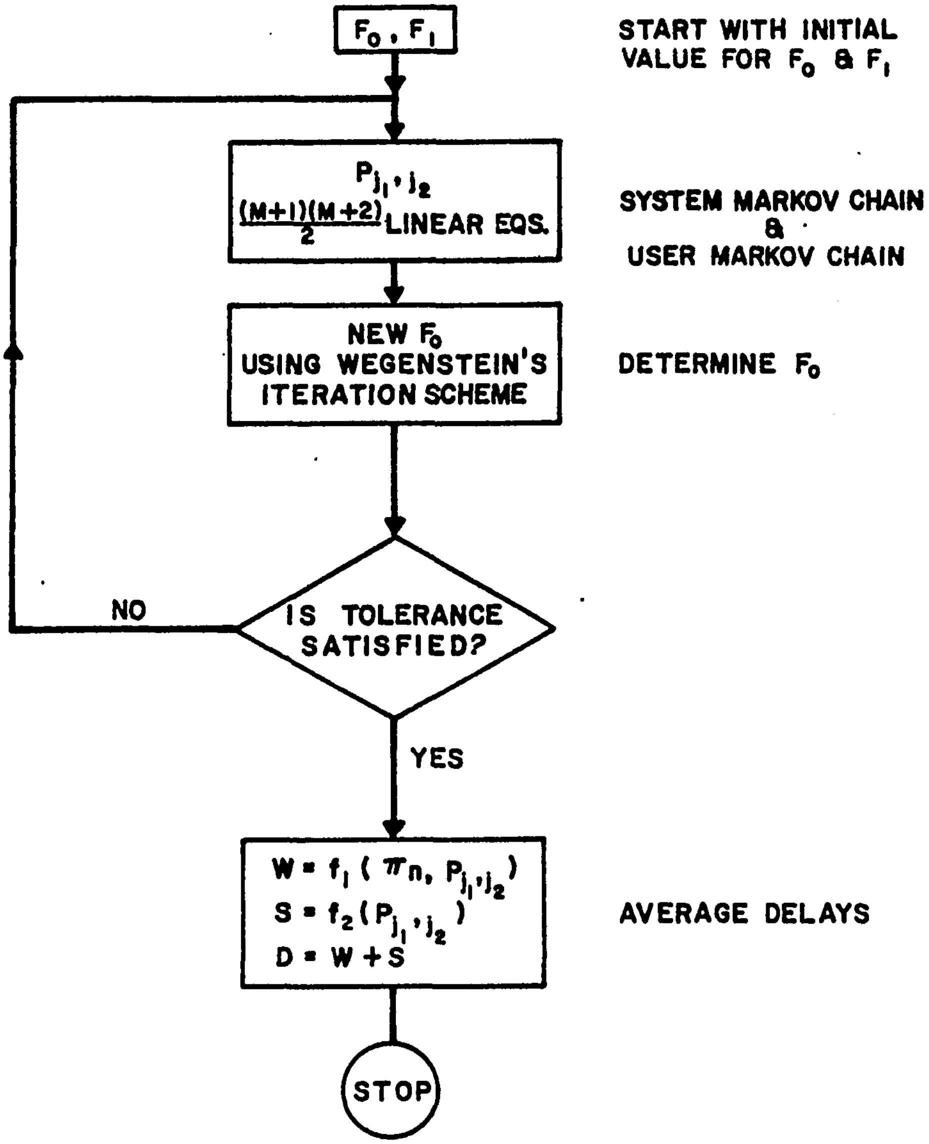

FIGURE 4·2. FLOW CHART TO DETERMINE AVERAGE DELAYS FOR THE PROPOSED SCHEME

$$S = \pi_1 \cdot \frac{M}{2} + \frac{(1-\pi_0-\pi_1)}{2} \left[\frac{M}{2} + \frac{1}{2}(S_f+M)\right] \tag{4.22}$$

The average frame size, another parameter of our interest, is given by

$$S_f = \sum_{j_1=0}^{M} \sum_{j_2=0}^{M-j_1} (M+j_2)P_{j_1 j_2} \tag{4.23}$$

Finally, we define stability as,

"The system is said to be stable if the steady state probability of having a finite number of packets in the buffer exists and is non-zero."

CHAPTER 5

SIMULATION RESULTS

A FORTRAN simulation program was written to find the throughput-delay performance of the proposed model described in the last chapter, under the following traffic conditions:

- Equal traffic distribution among nodes in which traffic was equally distributed among the nodes
- Unequal traffic distribution among nodes in which the arrival rate of packets for one "big" node is higher than that of other "small" nodes.

It will be assumed that the satellite channel is a highly reliable data link, in which noise may be neglected (a typical value for the capacity of the channel is 50,000 bits/sec). Packets of fixed length are used and time is slotted so that one packet is transmitted per slot. The slot size and the number of nodes is such that the duration of these M fixed slots is one round-trip delay to the satellite i.e., .27 sec.

5.1 Equal Traffic Among Nodes

In this case, the arrival process at each of the M terminals is assumed to be a Bernoulli process with rate σ, i.e., the probability of arrival of a (single packet) message at any terminal in each slot is σ. The total arrival rate is, therefore, M σ packets per slot, which is same as the system throughput or the system utilization factor. The Burnoulli process is well suited to the discrete time slot structure and is a discrete-time analog of the Poisson process. For varying M, the number of nodes and $M\sigma$, the system utilization factor, the following statistics have been collected from the simulation model. In a finite run-length simulation, the statistics are an arithmetic average over the M queues. Also, a comparison between these system performance parameters obtained by the mathematical model and those obtained by the simulation model has been made for M = 4 for the situation where each node is allowed to transmit maximum two packets in a time frame (i.e., $P = Q = R = 1$) for

(1) Mean and standard deviation of the queue sizes at the nodes at the beginning of a frame (TABLE 5-1).

(2) Mean and standard deviation of the frame length expressed in number of time slots (TABLE 5-2).

(3) Average and maximum delay for a packet at various loads (Figures 5.2 - 5.7). The analytical result for the average delay for $M=4$ and $P=Q=R=1$ is also shown in Figure 5.2 to compare with the simulation result.

Figure 5.1 compares the average delay for the proposed scheme with that for other protocols discussed in Chapter 3. It shows that the proposed scheme gives a significant improvement for delay over the other reservation schemes. It is also seen from Figures (5.2 - 5.7) that, for the traffic situation discussed above, each node needs approximately two (2) slots at most to transmit packets in a time frame to achieve the best performance. This result also agrees with the results of Balgangadhar and Pickholtz's scheme [37] where they have found out that the average number of frame size is approximately 1 slot per node at low throughput and less than 2 slots per node at high throughput (See Table 5.2). In other words, the knowledge of the second preceding reservation vector for a node is not necessary to have two slots at most in a time frame.

The agreement between the analytical results and the simulation results is seen to be excellent over a wide range of $M\sigma$, the system utilization factor. For example, for $M = 4$ and the utilization factor of 0.4, the mean queue size at the epochs and the mean frame size, computed analytically, are 0.43 and 4.24 respectively (see Tables 5.1 and 5.2). For the same value of $M\sigma$, simulation results give the mean queue size at the epoch as 0.44 and the mean queue size at the epochs as 0.44 and the mean frame size as 4.29. The disagreement between these and the analytical values is about 2 percent and 1 percent, respectively.

The agreement between the analytical and the simulation results (Figure 5.2) is seen to be particularly good for low values of system utilization factor. It is difficult to argue why the independence assumption made in the mathematical analysis should hold for low values of the utilization factor. On the other hand, for large values of M, one might argue that because of the large number of queues, the different queues become decoupled and hence makes the independence assumption valid. But further analysis of this should be made.

5.2 Unequal Traffic Among Nodes

In this case, it is considered that a "big" node consists of (single packet) messages whose arrival rate (σ_b) is eight times the arrival rate (σ_a) of (single packet) messages at other "small" nodes. Total small node traffic was divided equally among themselves. The total arrival rate is, therefore, ($\sigma_b + (M-1)\sigma_a$), which is equal to the system throughput.

For varying M, the number of nodes, and the system throughput, the average and maximum delay for a packet at "small" and "big" nodes at various loads have been obtained from the simulation model (Figures 5.8 - 5.13). It is seen from these figures that by assigning variable slots to the "big" node we improve the average and maximum delay characteristics of the system. The knowledge of preceding reservation vectors, is, therefore, necessary to assign variable slots to a node in a time frame.

The same algorithm for reservation of slots is applied to all nodes (Table 3.1). It is seen that the best performance is obtained for $P=3$, $Q=6$, and $R=11$ for the traffic situation discussed above. However, optimum values of P, Q and R and their relationships will depend on the ratio of the arrival rates at "big" and "small" nodes. Further analysis of these should be made. It is important to note here that an adaptive algorithm (similar to the algorithm discussed in this dissertation), where a node asking for more than two slots [by sending reservation vector (1.1)] is assigned some additional slots in addition to the slots assigned to it in the previous frame, is also studied and found not suitable in this case.

TABLE 5.1

COMPARISON OF COMPUTED (ANALYTICAL) AND SIMULATED VALUES FOR THE MEAN AND STANDARD DEVIATION (IN PARENTHESIS, BELOW THE MEAN) OF QUEUE SIZES AT THE BEGINNING OF A FRAME, FOR VARIOUS VALUES OF N AND Nσ (THE NUMBER OF NODES AND THE SYSTEM UTILIZATION FACTOR, RESPECTIVELY)

	Nσ = System Utilization							
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
ANALYTICAL	.1	.2	.31	.43	.57	.72	.91	1.15
(M = 4)	(.3)	(.45)	(.54)	(.63)	(.7)	(.77)	(.86)	(0.88)
Simulated								
M = 4	.1	.2	.32	.44	.57	.72	.93	1.2
	(.3)	(.45)	(.54)	(.64)	(.72)	(.79)	(.9)	(1.1)
M = 12	.1	.2	.31	.43	.56	.72	.91	1.17
	(.3)	(.45)	(.55)	(.65)	(.75)	(.85)	(.96)	(1.09)
M = 20	.1	.2	.31	.42	.55	.74	.92	1.21
	(.31)	(.44)	(.56)	(.65)	(.74)	(.85)	(.97)	(1.12)

TABLE 5.2

COMPARISON OF COMPUTED (ANALYTICAL) AND SIMULATED VALUES FOR THE MEAN AND STANDARD DEVIATION (IN PARENTHESES, BELOW THE MEAN) OF FRAME SIZE IN NUMBER OF SLOTS, FOR VARIOUS VALUES OF N AND Nσ (THE NUMBER OF NODES AND THE SYSTEM UTILIZATION FACTOR, RESPECTIVELY)

	Nσ = System Utilization							
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
Analytical	4.02	4.08	4.13	4.24	4.4	4.6	5.0	5.5
(M = 4)	(.15)	(.3)	(.44)	(.55)	(.73)	(.9)	(.98)	(1.3)
Simulated								
M = 4	4.02	4.08	4.17	4.29	4.47	4.72	5.17	5.91
	(.15)	(.3)	(.45)	(.57)	(.76)	(.93)	(1.22)	(1.69)
M = 12	12.05	12.24	12.5	12.97	13.57	14.45	15.76	17.7
	(.24)	(.49)	(.74)	(1.02)	(1.41)	(1.82)	(2.37)	(3.24)
M = 20	20.08	20.34	20.88	21.59	22.53	24.08	26.09	29.37
	(.28)	(.62)	(.96)	(1.37)	(1.74)	(2.35)	(2.95)	(3.89)

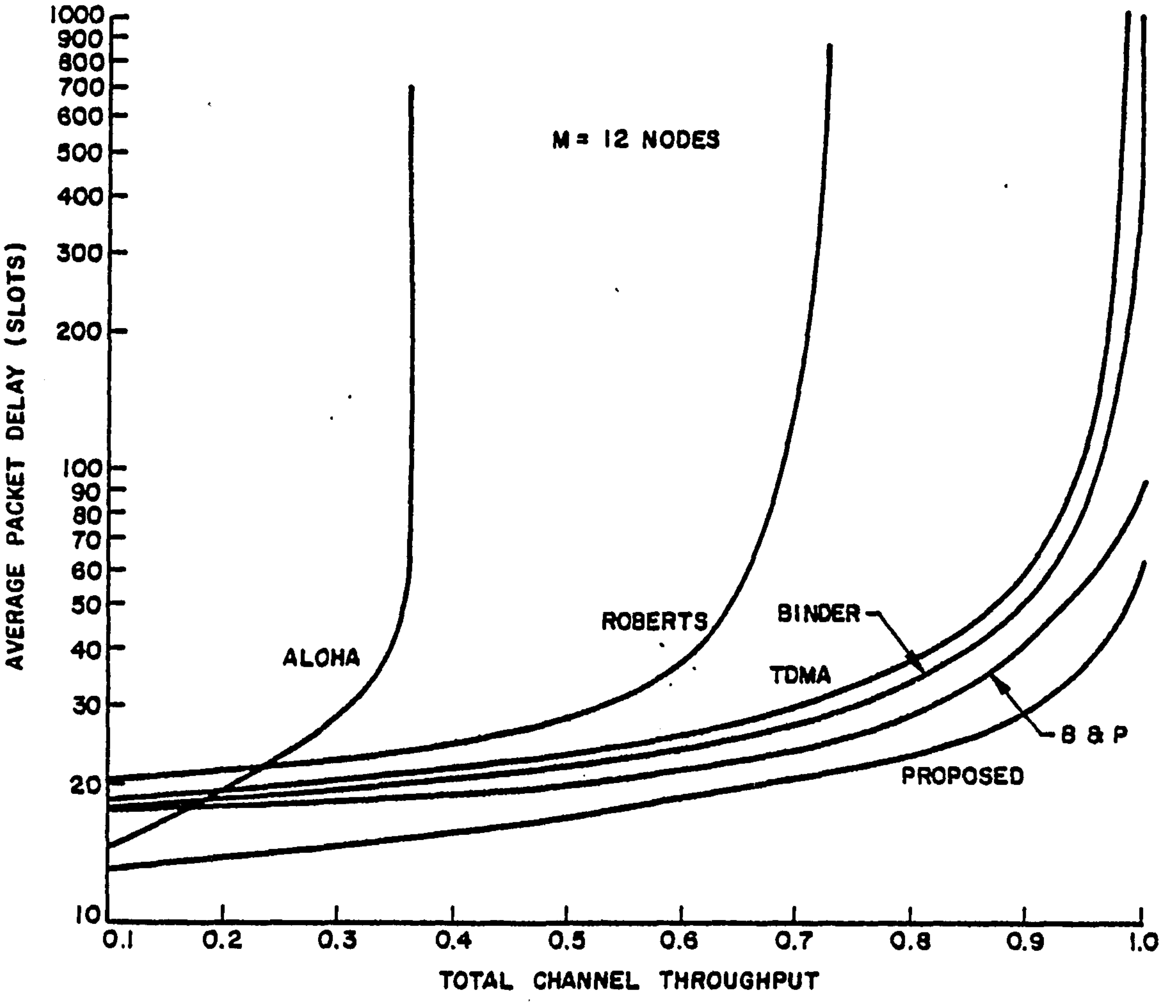

FIGURE 5·1. COMPARISON OF VARIOUS SCHEMES (12 NODES): AVERAGE PACKET DELAY vs THROUGHPUT

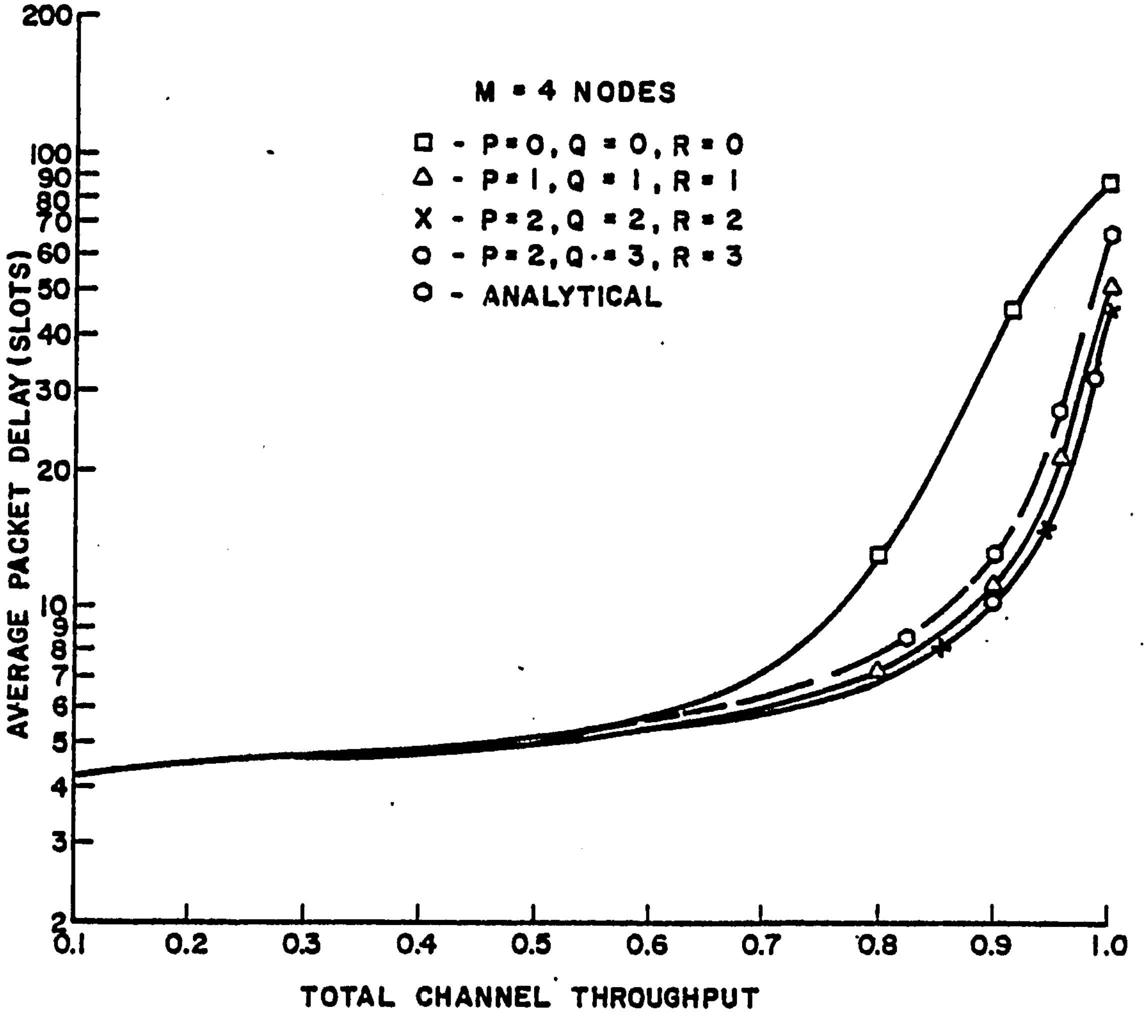

FIGURE 5·2. AVERAGE PACKET DELAY vs. THROUGHPUT (4 NODES)

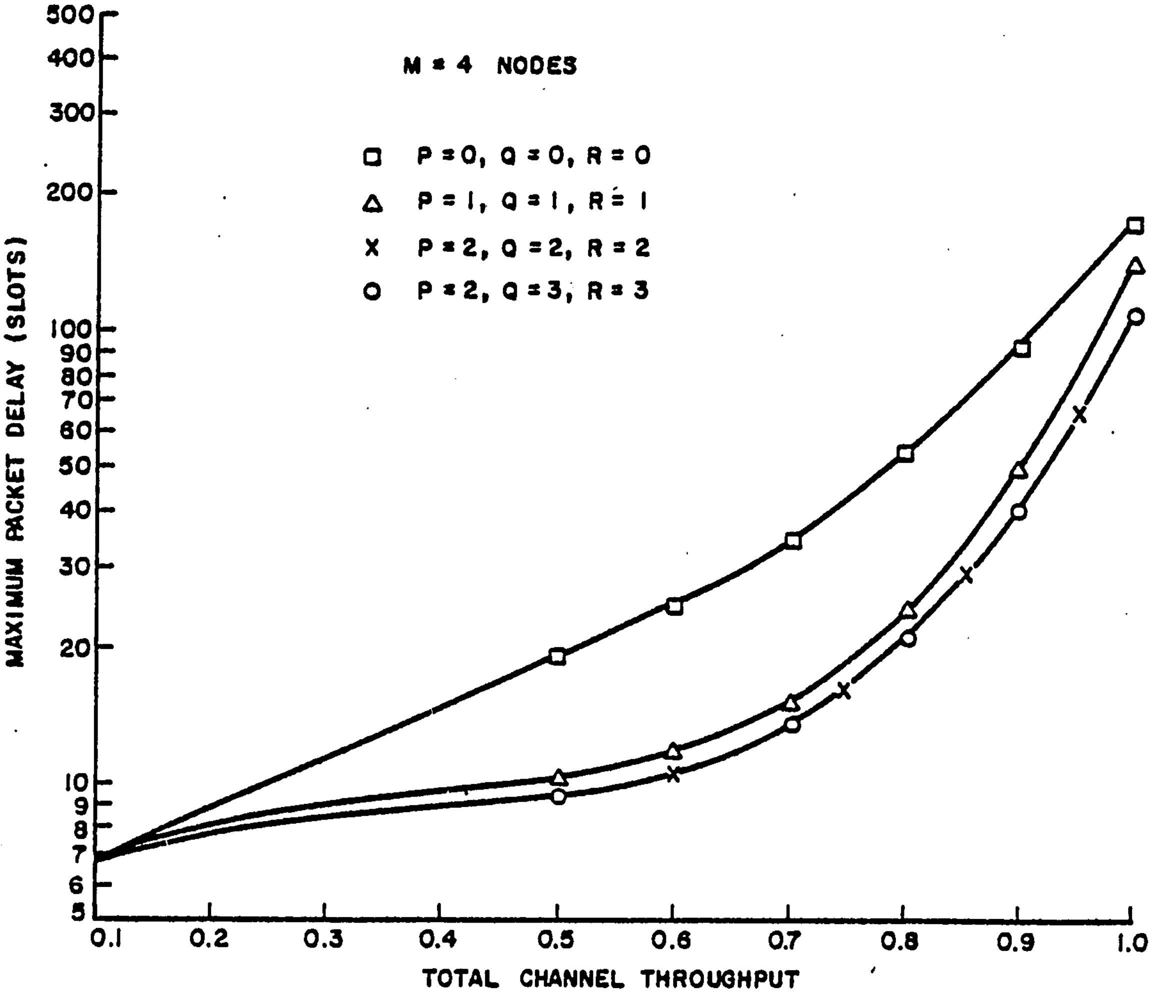

FIGURE 5·3. MAXIMUM PACKET DELAY vs THROUGHPUT
(4 NODES)

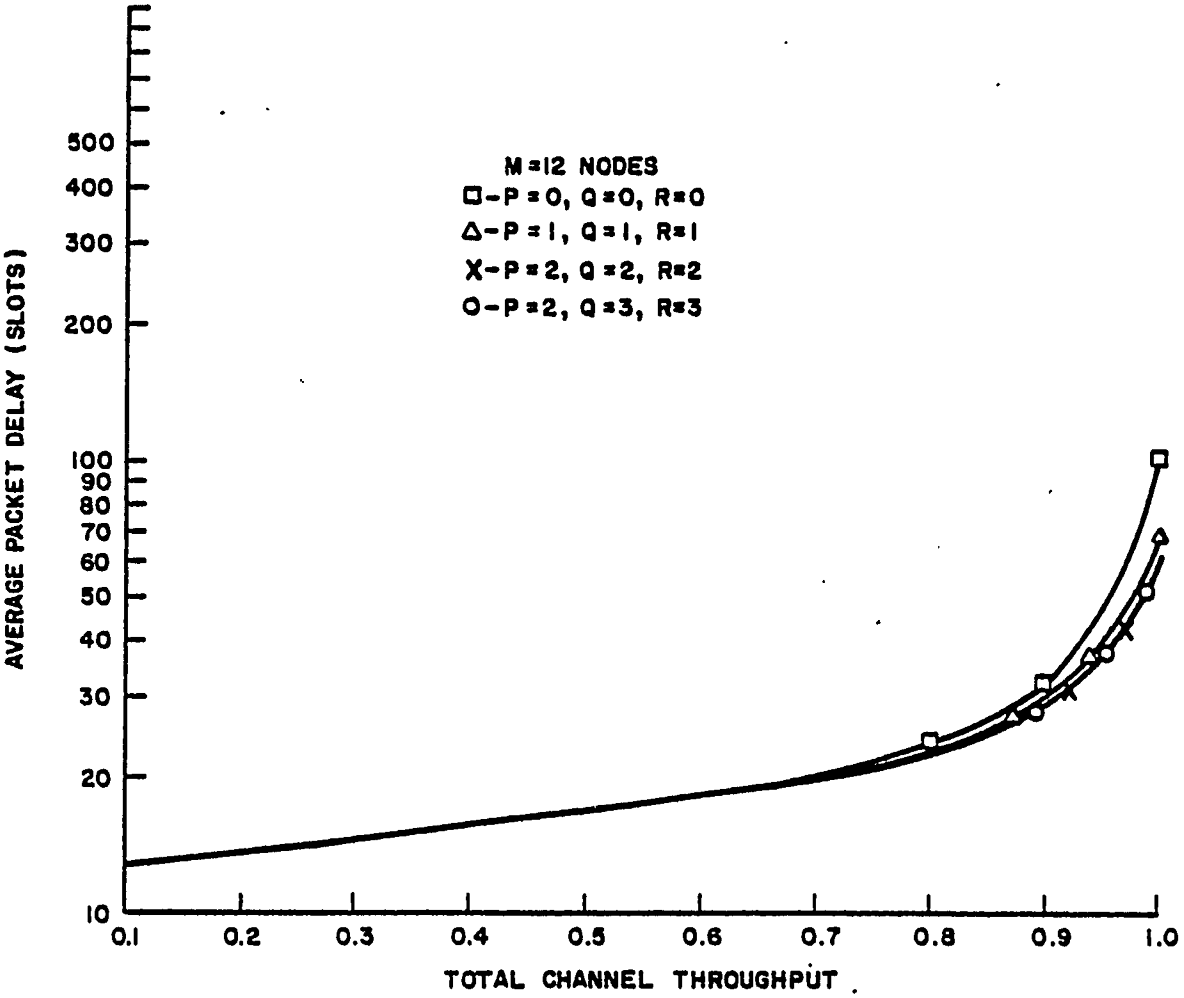

FIGURE 5·4. AVERAGE PACKET DELAY VS. THROUGHPUT (12 NODES)

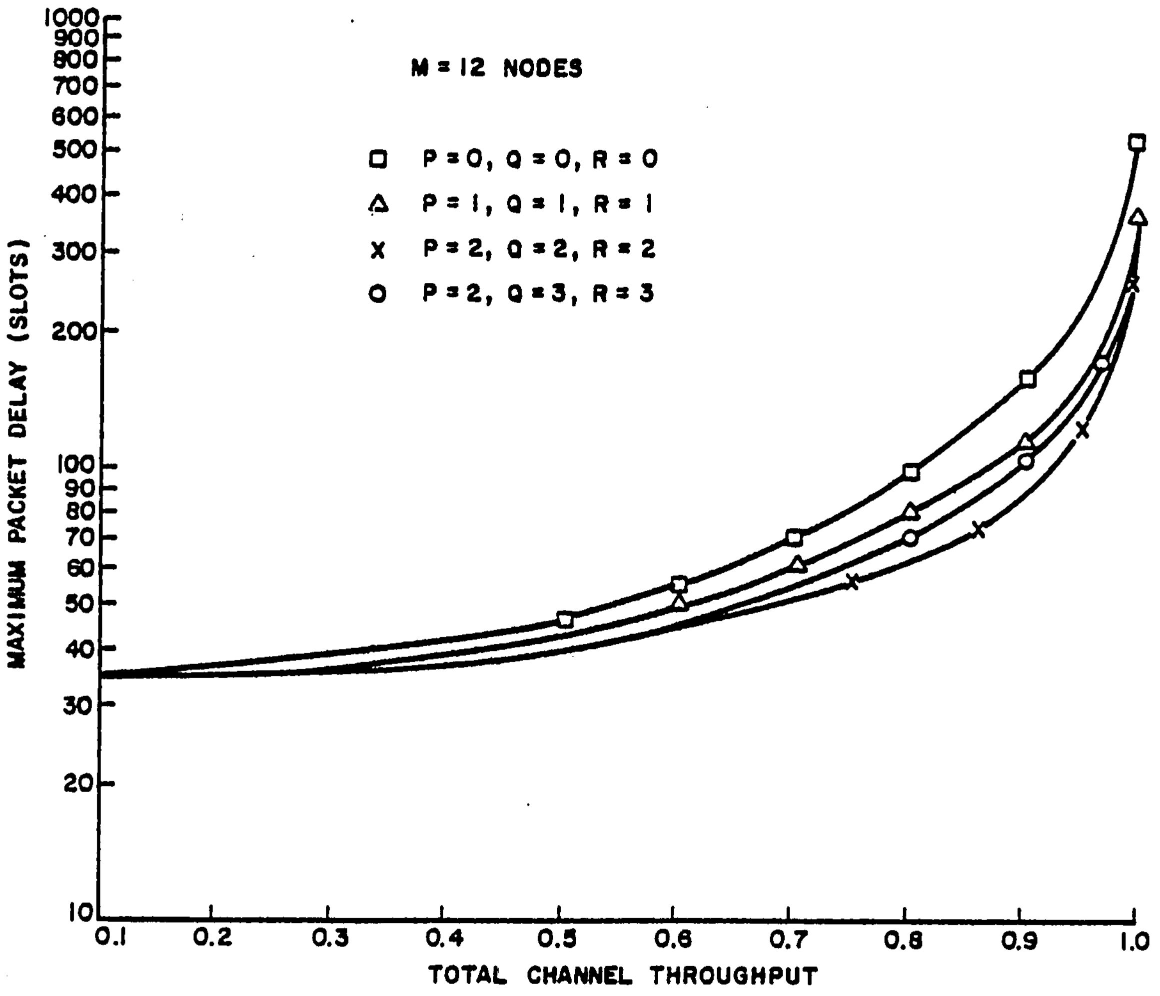

FIGURE 5·5. MAXIMUM PACKET DELAY vs THROUGHPUT (12 NODES)

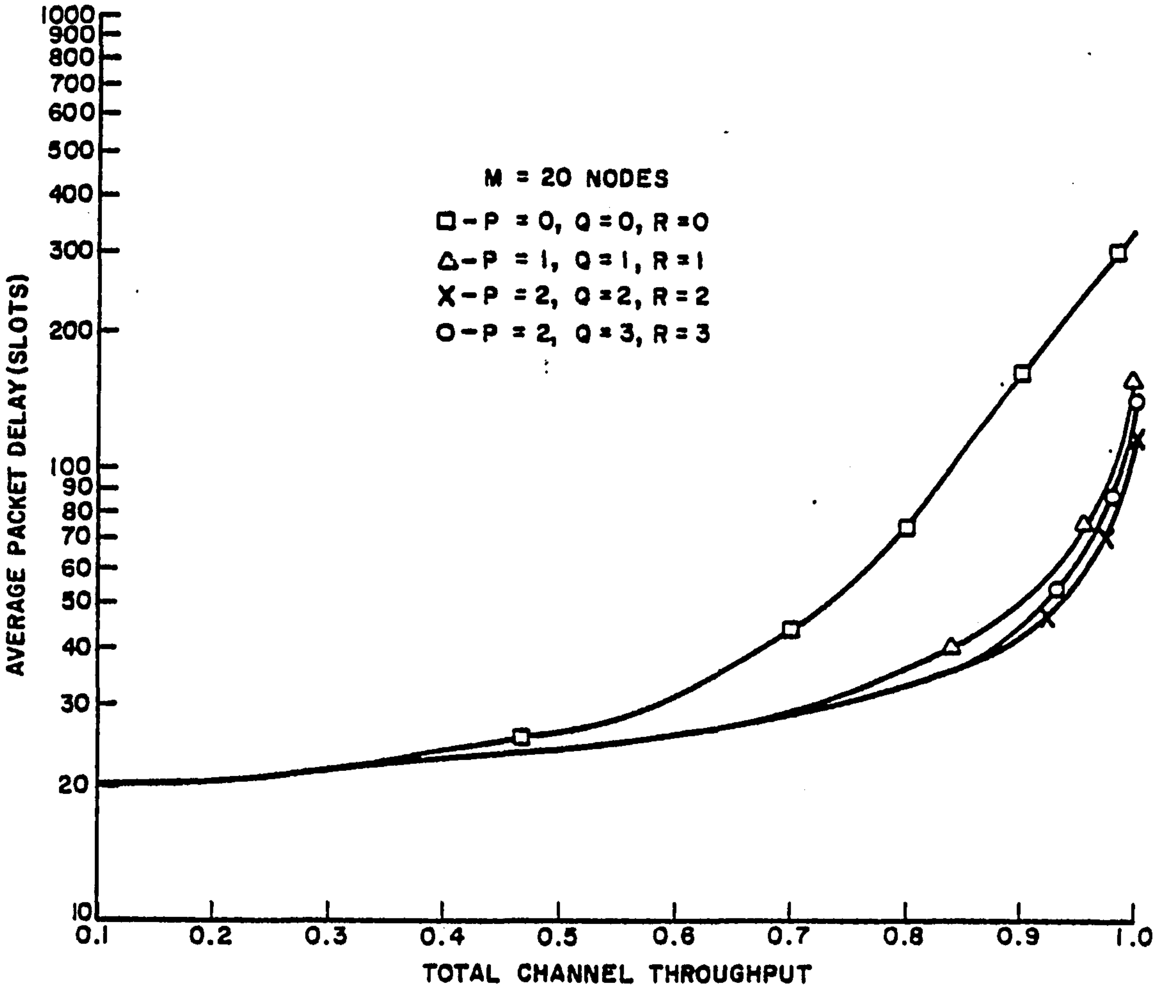

FIGURE 5-6. AVERAGE PACKET DELAY vs. THROUGHPUT (20 NODES)

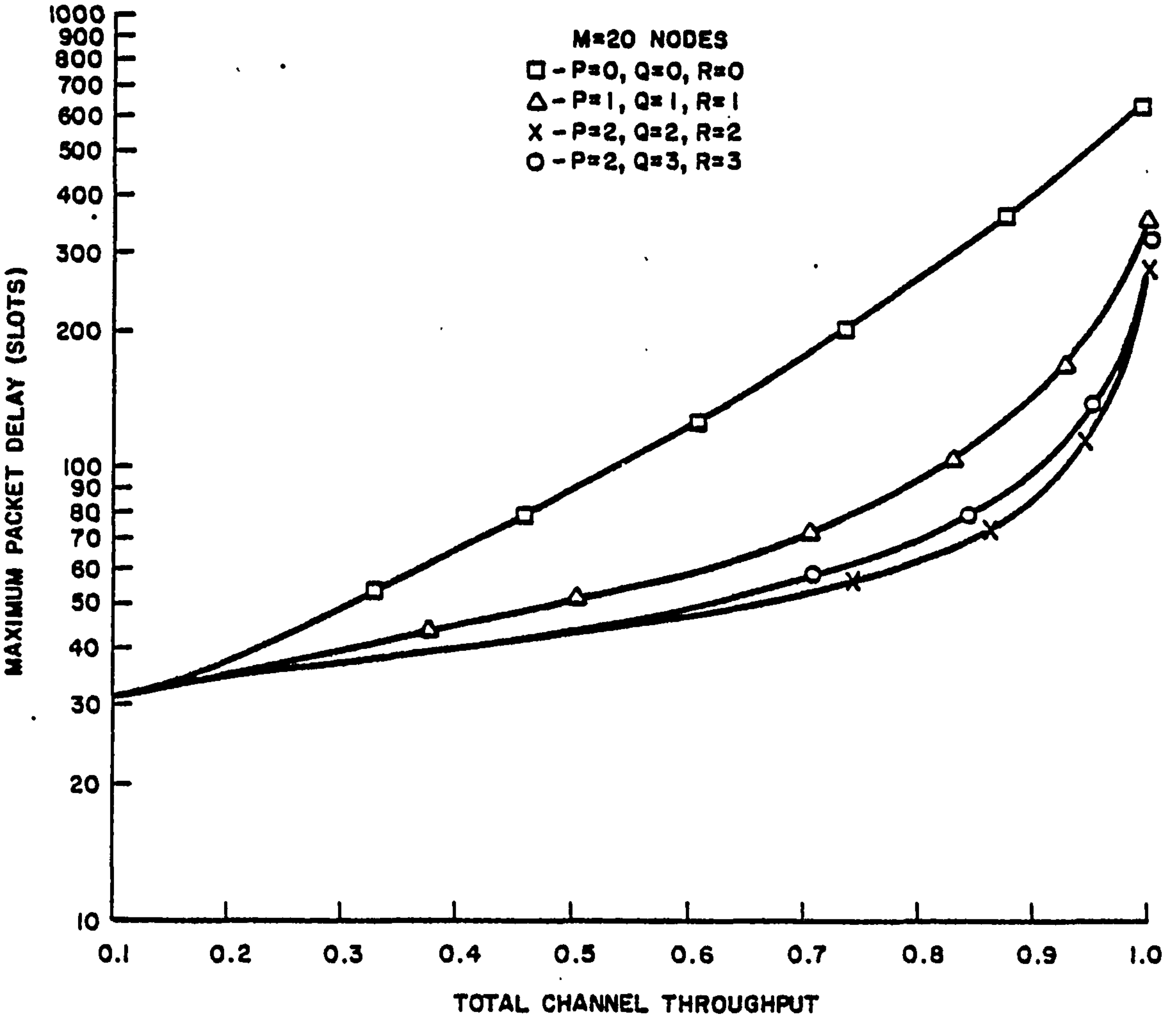

FIGURE 5·7. MAXIMUM PACKET DELAY VS. THROUGHPUT (20 NODES)

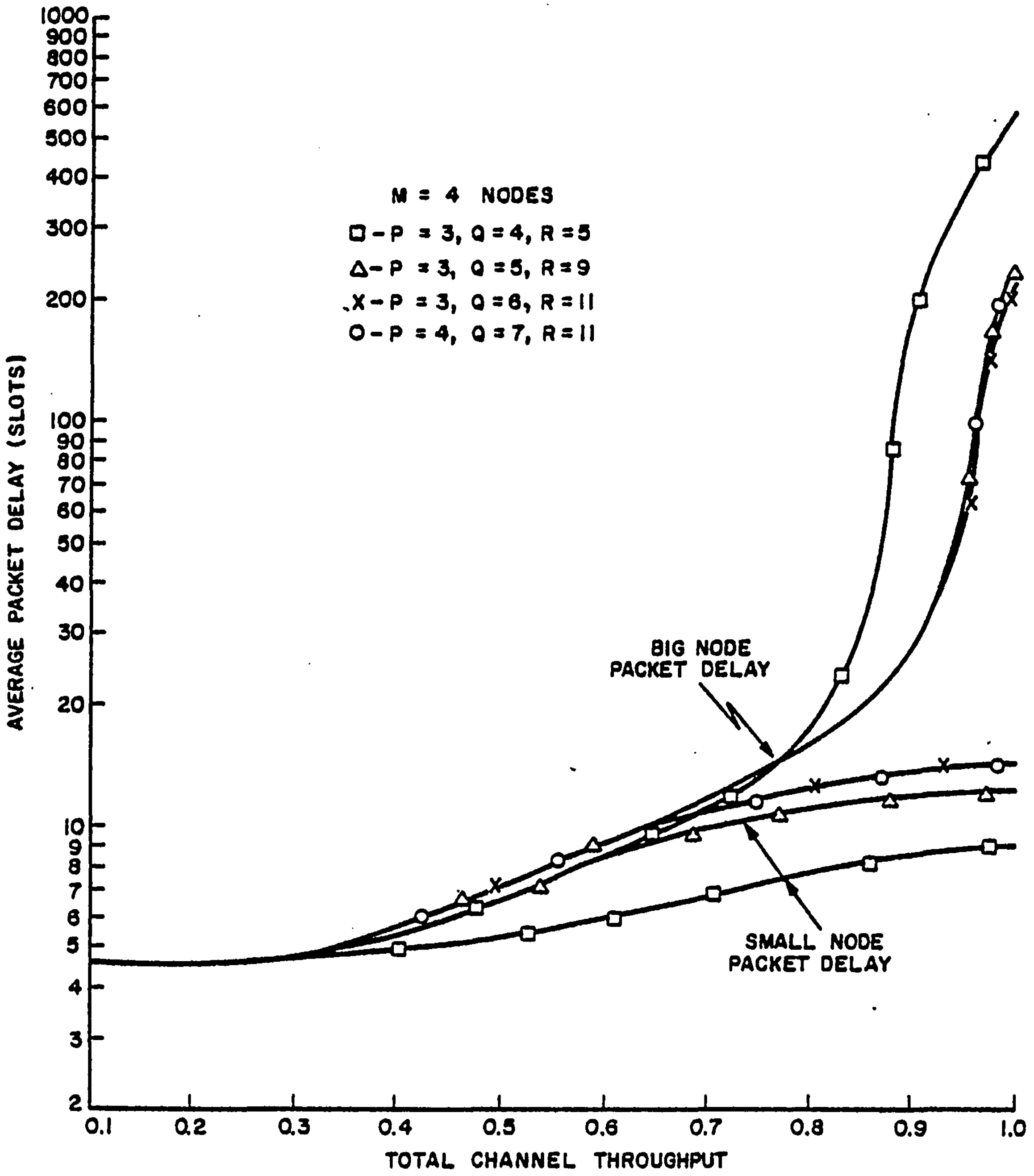

FIGURE 5-8. AVERAGE PACKET DELAY vs. THROUGHPUT (4 NODES): PACKET ARRIVAL RATE OF 1 BIG NODE EIGHT TIMES HIGHER THAN THAT OF EACH OF 3 SMALL NODES.

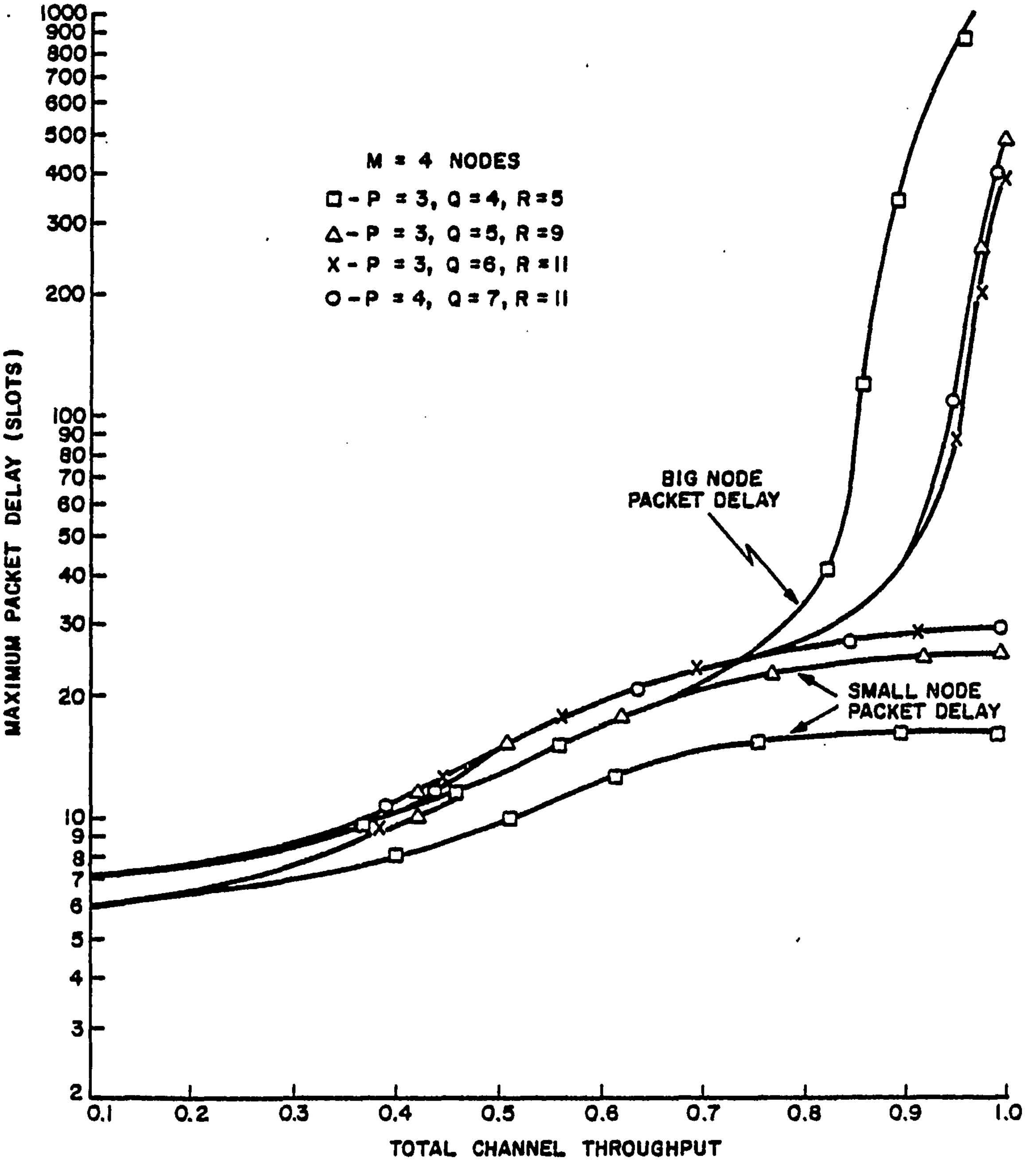

FIGURE 5·9. MAXIMUM PACKET DELAY vs. THROUGHPUT (4 NODES): PACKET ARRIVAL RATE OF 1 BIG NODE EIGHT TIMES HIGHER THAN THAT OF EACH OF 3 SMALL NODES.

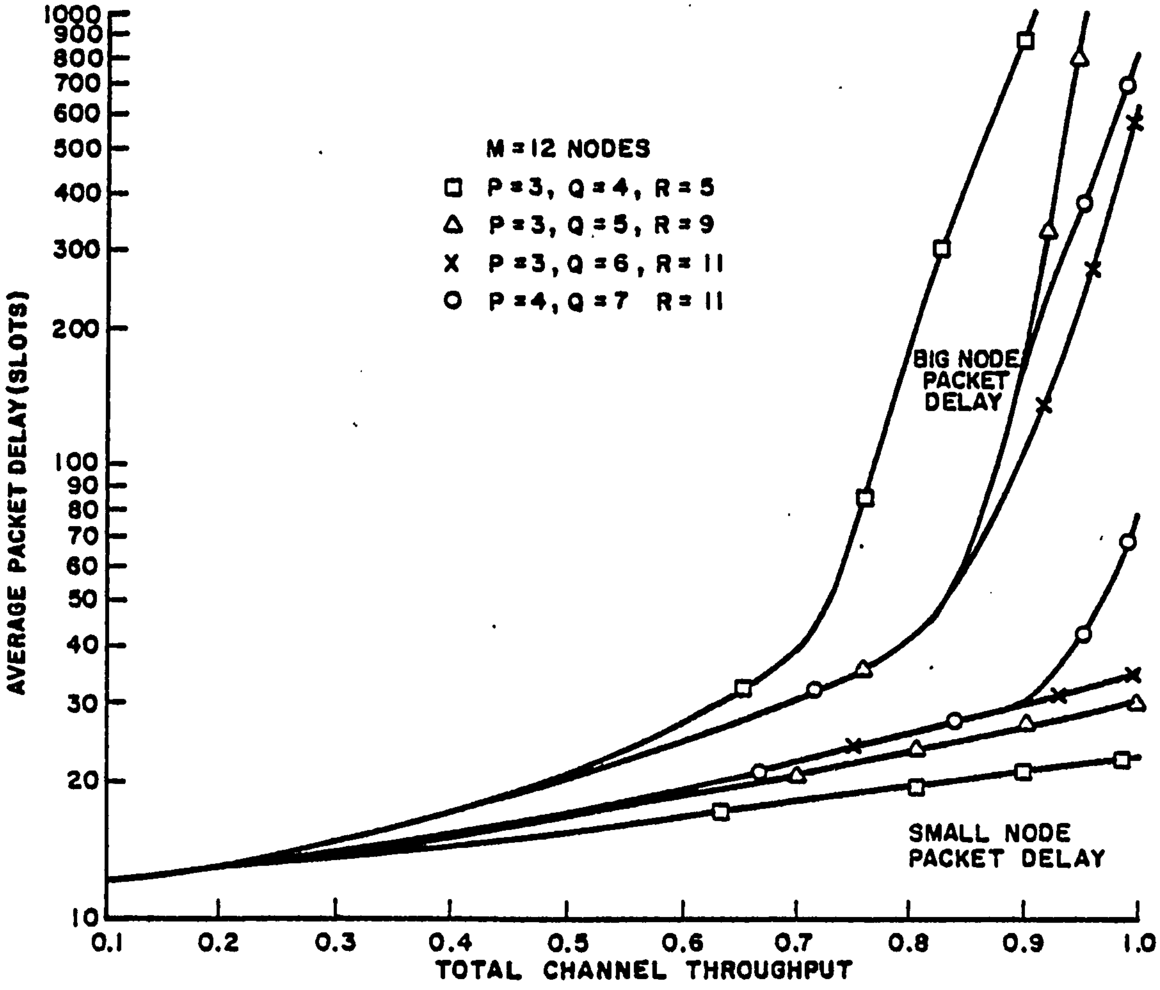

FIGURE 5·10. AVERAGE PACKET DELAY VS THROUGHPUT (12 NODES): PACKET ARRIVAL RATE OF 1 BIG NODE EIGHT TIMES HIGHER THAN THAT OF EACH OF 11 SMALL NODES.

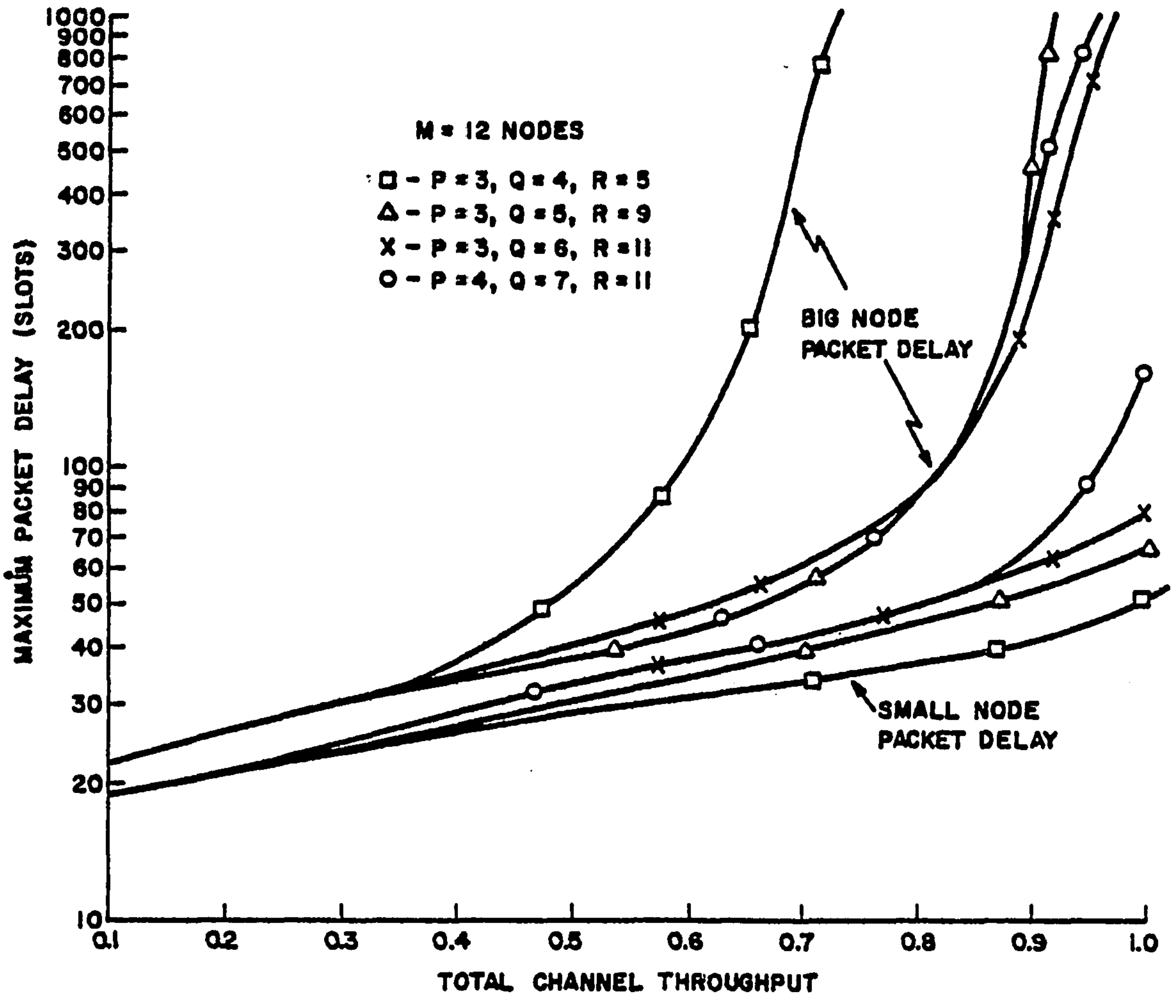

FIGURE 5-11. MAXIMUM PACKET DELAY VS. THROUGHPUT (12 NODES): PACKET ARRIVAL RATE OF 1 BIG NODE EIGHT TIMES HIGHER THAN THAT OF EACH OF 11 SMALL NODES.

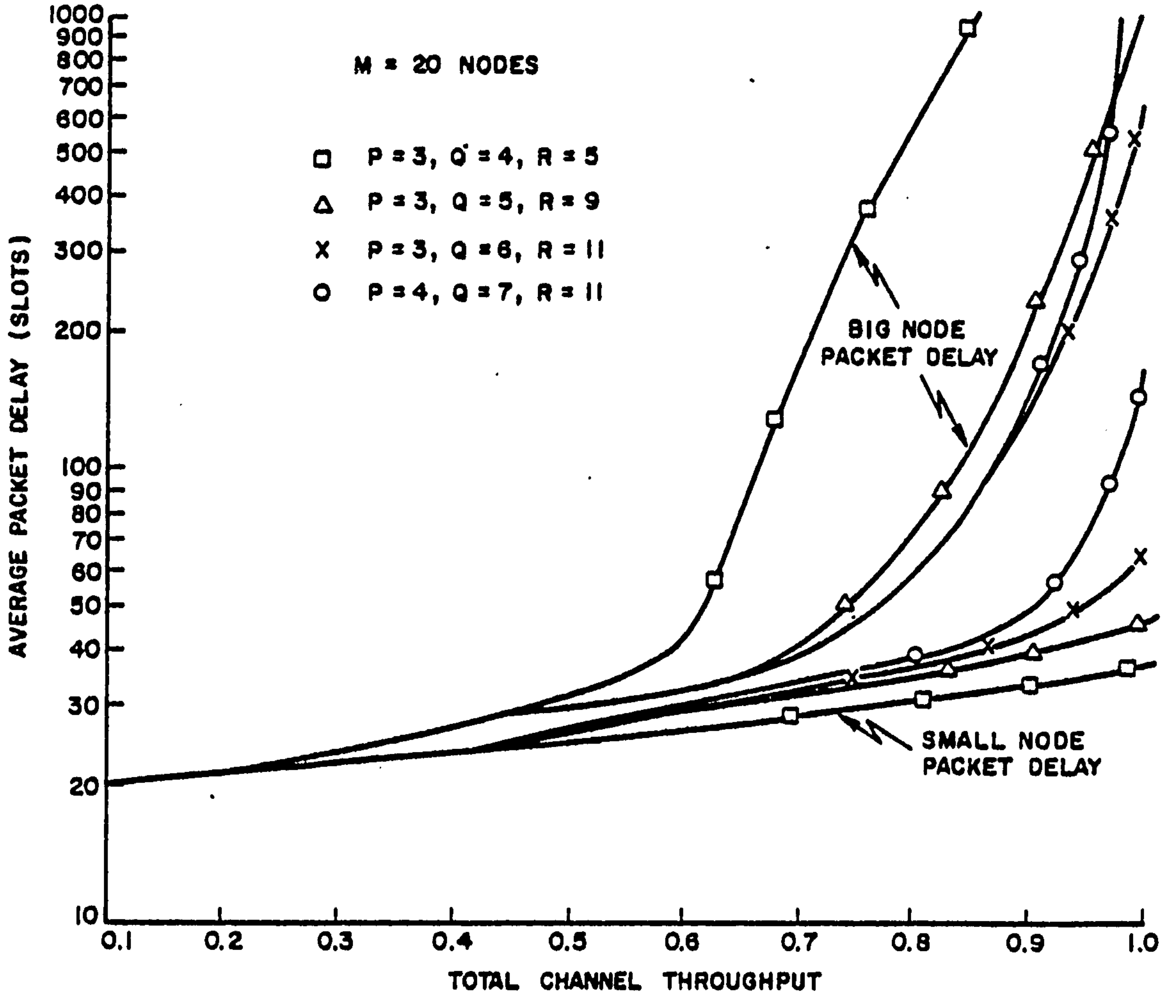

FIGURE 5·12. AVERAGE PACKET DELAY vs THROUGHPUT (20 NODES): PACKET ARRIVAL RATE OF 1 BIG NODE EIGHT TIMES HIGHER THAN THAT OF EACH OF 19 SMALL NODES.

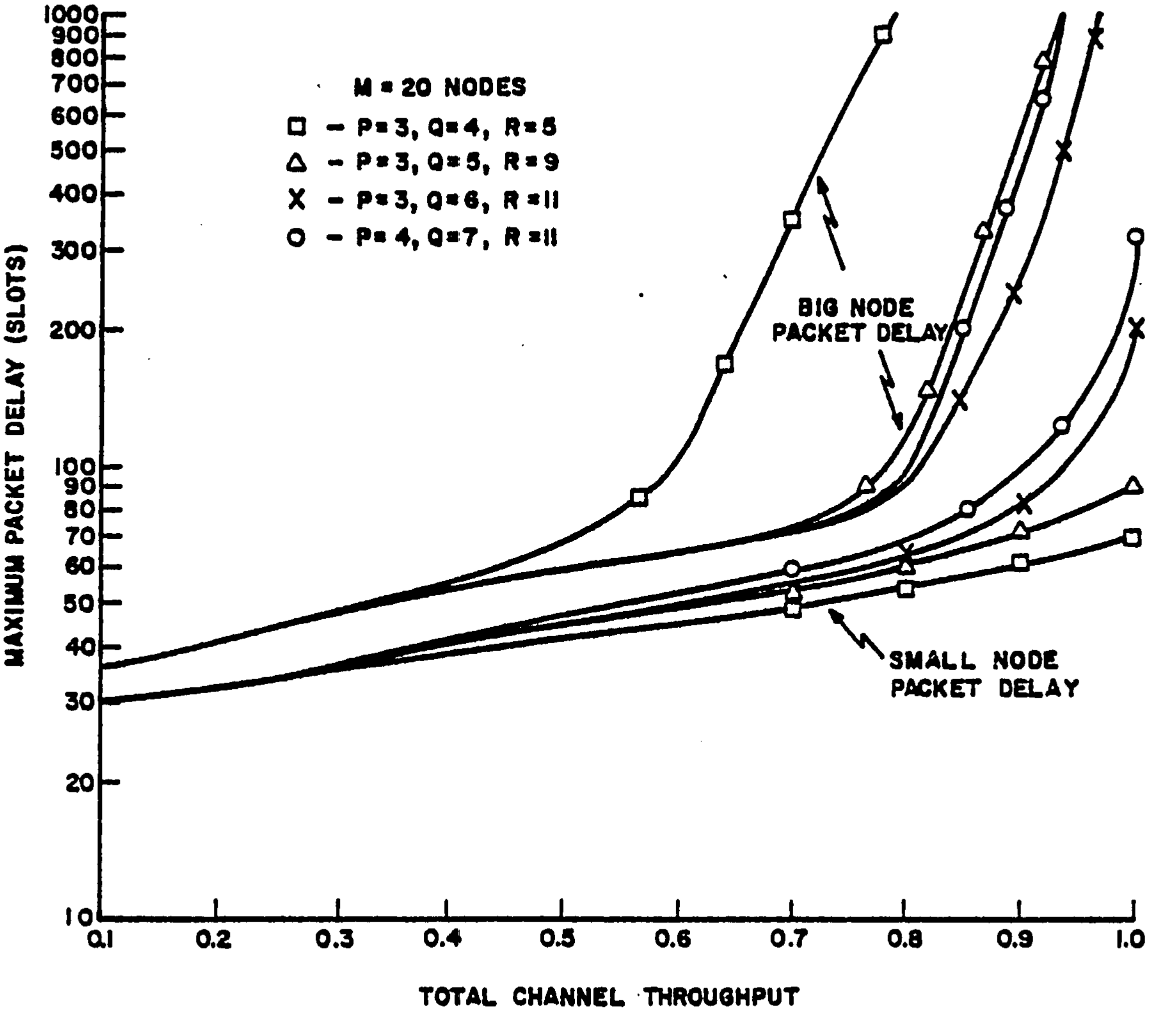

FIGURE 5·13 MAXIMUM PACKET DELAY VS. THROUGHPUT (20 NODES): PACKET ARRIVAL RATE OF 1 BIG NODE EIGHT TIMES HIGHER THAN THAT OF EACH OF 19 SMALL NODES.

CHAPTER 6

CONCLUSIONS AND SUGGESTIONS FOR FURTHER WORK

The field of large data networks has seen a tremendous growth in the last decade. In the subarea of multiple access schemes for neworks that include broadcast (and particularly satellite) links although many time-division multiplexed schemes have been proposed for the shared use of a common channel by geographically separated users, very few of the schemes have been modeled analytically. Also, the problems of choosing the optimum access protocol (in the sense of minimizing average delay per message for a given throughput) have not been formulated yet. The basic goal of this dissertation is to develop optimum control schemes that can be employed when using a geostationary satellite channel for intercommunications between a set of geographically distributed nodes and, more importantly, modeling and analysis of such schemes.

Several existing and proposed satellite multiple access techniques are examined. In this dissertation, we develop and analyze a dynamic reservation-based multiple access packet switching technique, which minimizes both average and maximum delays and maximizes traffic throughput. The analysis is based on a combination of two Markov chain model; one Markov chain describes the status of the buffer contents of a typical user, we refer to it as the User Markov Chain and one that describes the status of all the users of the channel, we refer to it as the System Markov Chain.

Simulation results have been obtained under various traffic distributions among nodes. These results show markedly improved delay-throughput characteristics over other existing and proposed multiple access techniques with especially significant performance gains at high traffic and also as the traffic imbalance among the users increases. A good agreement between the simulation results and the analytical results is obtained.

For equal traffic distribution among nodes, each node needs approximately two (2) slots at most to transmit packets in a time frame to achieve the best performance. It is found that the average number of frame size is approximately 1 slot per node at low throughput and less than 2 slots per node at high throughput. In other words, the knowledge of the second preceding reservation vector for a node is not necessary to assign two slots at most in a time frame.

For unequal traffic distribution among nodes, on the other hand, the delay-throughput characteristic is improved by assigning variable slots to the nodes, in a time frame. The knowledge of preceding reservation vectors is, therefore, necessary to assign variable slots to a node in a time frame. The optimum values of variable slots to be assigned to the nodes are obtained for unequal traffic among nodes in which the arrival rate of packets for one node is eight times higher than that of other nodes. However, these values will depend on the traffic distribution among nodes. Further analysis of these should be made.

The proposed packet switching technique based upon dynamic reservation multiple access concept may also be applied to the use of ground radio channels for terminal-computer communications. In this case the analysis will be the same, except the time frame will consist of only reservation-based variable time slots but no preassigned fixed slots because there is no round-trip delay.

REFERENCES

[1] Jackson, P. E. and Stubbs, C. D., "A Study of Multiaccess Computer Communications," Spring Joint Computer Conference, AFIPS S Conference Proceedings, 1969, Vol. 34, pp. 491-504.

[2] Fuchs, E. and Jackson, P. E., "Estimates of Distribution of Random Variables for Certain Computer Communications Traffic Models," Communications of the ACM, Vol. 13, No. 12, pp. 752-757, December 1970.

[3] Abramson, N., "Packet Switching with Satellites," National Computer Conference, New York, June 4-8, 1973, AFIPS Conference Proceedings, 1973, Vol. 42, pp. 695-702.

[4] Feller, W., "An Introduction to Probability Theory and its Applications," 3rd ed., Vol. I, Wiley, New York, 1968.

[5] Martin, J., "Systems Analysis for Data Transmission," Prentice-Hall, Englewood Cliffs, N. J., 1972.

[6] Hayes, J. and Sherman, D., "Traffic Analysis of a Ring Switched Data Transmission System", Bell System Technical Journal, Vol. 50, No. 9, pp. 2947-2978, November 1971.

[7] Pierce, J. R., "Network for Block Switching of Data," IEEE Convention Record, New York, March 1971.

[8] Chu, W. W., "A Study of Asynchronous Time Division Multiplexing for Time-Sharing Computer Systems," Spring Jint Computer Conference, AFIPS Conference Proceedings, 1969, Vol. 35, pp. 669-678.

[9] Baran, P., "On Distributed Communications: XI. Summary Overview," Rand Corporation, Santa Monica, California, Memorandum RM-3767-PR, August 1964.

[10] Kleinrock, L., "Communication Nets: Stochastic Message Flow and Delay", McGraw Hill, New York, 1964, out of print (Reprinted by Dover Publications, New York, 1972).

[11] Roberts, L. G., "Data by the Packet," IEEE Spectrum, Vol. 11, No. 2, pp. 46-51, February 1974.

[12] Network Analysis Corporation., "The Practical Impact of Recent Computer Advances on the Analysis and Design of Large Scale Networks," First Semiannual Technical Report, Glen Cove, New York, May 1973.

[13] Frank, H., Gerla, M., and Chou, W., "Issues in the Design of Large Distributed Computer Communication Networks," Proceedings of the National Telecommunication Conference Atlanta, November 26-28, 1973.

[14] Frank, H., and Chou, W., "Topological Optimization of Computer Networks," Proceedings of the IEEE, Vol. 60, No. 11, pp. 1385 - 1396, November 1972.

[15] Gerla, M., Chou, W., and Frank, H., "Computational Considerations and Routing Problems for Large Computer Communication Networks," Proceedings of the National Telecommunication Conference Atlanta, November 26 - 28, 1973.

[16] Abramson, N., "THE ALOHA SYSTEM--Another Alternative for Computer Communications," Fall Joint Computer Conference, AFIPS Conference Proceedings, 1970, Vol. 37, pp. 281-285.

[17] Martin, J., "Future Developments in Telecommunications" Prentice-Hall, Englewood Cliffs, N. J. 1971.

[18] Roberts, L. G. "Extensions of Packet Communication Technology to a Hand Held Personal Terminal," Spring Joint Computer Conference, AFIPS Conference Proceedings, 1972, Vol. 40, pp. 295-298.

[19] Fultz, G. L., "Adaptive Routing Techniques for Message Switching Computer-Communication Networks," School of Engineering and Applied Science, University of California, Los Angeles, UCLA-ENG-7252, July 1972.

[20] Cacciamani, E., "Data Services--American Satellite Corporation," Seventh Hawaii International Conference on System Sciences, University of Hawaii, Honolulu, January 8 - 10, 1974, Proceedings of the Special Subconference on Computer Nets, 1974.

[21] Gray, E. M., "ANIK," Communications Society, Vol. 12, No. 1, pp. 3 - 5, January 1974.

[22] Gabbard, O., "Design of a Satellite Time-Division Multiple-Access Burst Synchronizer," IEEE Transactions on Communications Technology, Vol. COM-16, No. 4, pp. 589-596, August 1968.

[23] Rettberg, R., "A Brief Simulation of the Dynamics of an ALOHA System with Slots," ARPA Networks Information Center, Stanford Research Institute, Menlo Park, California, ASS Note 11 (NIC 11293), July 1972.

[24] Cacciamani, E., "The SPADE System as Applied to Data Communications and Small Earth Station Operation," COMSAT Technical Review, Vol. 1, No. 1, pp. 171-182, 1971.

[25] Dunn, D. and Eric, M., "The Economics of Packet Switching," Seventh Hawaii International Conference on System Sciences, University of Hawaii, Honolulu, January 8 - 10, 1974, Proceedings of the Special Subconference on Computer Nets, 1974.

[26] Tobagi, F. A., "Multi-access Protocols in Packet Communication Systems," IEEE Trans. Commun., Vol. 28, pp. 468-488, April 1980.

[27] Kahn, R. E., Gronemeyer, S. A., Burchfiel, J. and Kunzelman, R. C., "Advances in Packet Radio Technology", Proc. IEEE, Vol. 66, Nov. 1978.

[28] Rubin, I., "Message Delays in FDMA and TDMA Communication Channels", IEEE Trans. Commun., Vol. 27, pp. 769 - 777, May 1979.

[29] Lam, S., "Delay Analysis of Time-Division Multiple Access (TDMA) Channel," IEEE Trans. Commun., Vol. COM-25, Dec. 1977.

[30] Kosovych, O., "Fixed Assignment Access Technique," IEEE Trans. Commun., Vol. COM-26, Sept. 1978.

[31] Abramson, N., "The Aloha System," in Computer Communication Networks, N. Abramson and F. Kuo, Eds. Englewood Cliffs, N. J., Prentice-Hall, 1973.

[32] Roberts, L. G., "ALOHA Packet System With and Without Slots and Capture," Comput. Commun. Rev., Vol. 5, pp. 28 - 42, April 1975.

[33] Kleinrock, L. and Lam, S., "Packet-Switching in a Slotted Satellite Channel," Nat. Computer Conf. AFIPS Conf. Proc., Vol. 42, Montvale, N. J. AFIPS Press, 1973, pp. 703-710.

[34] Crowther, W. R., Rettbert, R., Walden, D., Ornstein, S., and Heart, F., "A System for Broadcast Communication, Reservation-ALOHA," in Proc. 6th Hawaii Int. Syst. Sci. Conf., Jan. 1973.

[35] Roberts, L., "Dynamic Allocation of Satellite Capacity Through Packet Reservation," in Proc. AFIPS Conf., Vol. 42, June 1973.

[36] Binder, R., "A Dynamic Packet Switching System for Satellite Broadcast Channels," in Proc. ICC'75, San Francisco, CA, June 1975.

[37] Balgangadhar, M. M. and Pickholtz, L., "Analysis of Reservation Multiple Access Technique for Data Transmission Via Satellites", IEEE Trans. Commun. Vol. COM-27, Oct. 1979.

[38] Spragins, John D., "Loop Transmission Systems-Mean Value Analysis," IEEE Trans. on Comm., Vol. COM-20, No. 3, June 1972, pp. 592-602.

[39] Konheim, A. G. & Meister, "Service in a Loop System," JACM, Vol. 19, No. 1, Jan. 1972, pp. 92-108.

[40] Konheim, A. G., "Service Epochs in a Loop System," Symposium on Computers and Communications Networks and Teletraffic, Polytechnic Institute of Brooklyn, April 4-6, 1972, pp. 125-143.

[41] Spragins, John D., "Loops Used for Data Collection," Symposium on Computers and Communications Networks and Teletraffic, Polytechnic Institute of Brooklyn, April 4-6, 1972, pp. 59-76.

[42] Chu, W. W. and Konheim, A. G., "On the Analysis and Modeling of a Class of Computer Communication Systems," IEEE Trans. on Comm. Vol. COM-20, No. 3, June 1972, pp. 645-660.

[43] Fayolle, G. and Iasnogorodski, "Two Coupled Processors: The Reduction to a Riemann-Hilbert Problem" Zeilschriftfur Wahrscheinlichkeits Theory and Venwandte Gebiete, by Springer-Verlag 1979.

[44] Leibowitz, M. A., "An Approximate Method for Treating a Class of Multiqueue Problems," IBM Journal of Research and Development, Vol. 5, No. 3, pp. 204-209, July 1961.

[45] Hashida, O., "Analysis of Multiqueue," Review of the Electrical Comm. Lab., Nippon Telegraph and Telephone Public Corporation, Vol. 20, No. 3-4, March-April, 1972.

[46] Saadawi, T. N. and Ephremides, A., "Stability and Optimization of Slotted ALOHA with a Finite Number of Buffered Users", presented at the 19th CDC Conference, Albuquerque, New Mexico, Dec., 1980.

[47] Lance, G. N., "Numerical Methods for High Speed Computers", Iliffe, London, 1960, pp. 134-138.

[48] Little, D. C., "A Proof of the Queuing Formula: $L = \lambda W$", Operations Research, Vol. 9, 1961, pp. 383-387.

APPENDIX

We have shown here the analysis for the same system as described in the dissertation, except in this case we have assigned maximum one slot to a busy user (i.e. $P = Q = R = 0$). At the beginning of a frame, the small segments where the reservations are sent by the nodes, are assumed to be of zero duration. This analysis is valid to analyze the packet switching technique using ground radio channels for terminal-computer communications, except in this case, the round-trip delay will be of zero duration.

1. The User Model

The system consists of M terminals, each of which has an infinite buffer. The arrival process at each of the M terminals is assumed to be a Bernoulli process with rate σ. The user may be in one of two states; an idle user where his buffer is empty and an active user, where his buffer is non-empty.

Whenever the user has a packet in his buffer, it sends a reservation request at the beginning of the frame, and at the same time the packet at the head of line (HOL) is moved to the transmitter and waits there until it is transmitted in its assigned slot.

We need the following definitions; Let

π_i = steady state probability of having i packets in the buffer at the beginning of a frame

Hence

π_0 = probability of an empty buffer

= probability of an idle user

$1-\pi_0$ = probability of an active user

2. THE SYSTEM MARKOV CHAIN

The state of the system is described by the number of users requesting transmission at the beginning of the frame. Let us first define the following:

$j =$ Number of users requesting transmission at the beginning of a frame

$L_j =$ Length of a frame having j users requesting transmission at the beginning of a frame

$=$ M slots, where M is equal to or greater than one round-trip delay in time slots

$P_j =$ Steady state probability of having j users requesting transmission at the beginning of a frame

$\sigma_j =$ Probability of an idle user generating one packet or more during the frame of length L_j

$= 1-(1-\sigma)^M$

$F =$ Probability of having one packet only in the buffer at the beginning of a frame given user is active

$$= \frac{\pi_1}{1-\pi_0} \tag{A.1}$$

$K =$ Number of users having more than one packet remained active out of j users requesting transmission at the beginning of a frame

$M - K =$ Number of idle users at the beginning of a frame

$C_K(j) =$ Pr [K users having more than one packet remains active out of j users requesting transmission at the beginning of a frame]

$$= \binom{j}{K}(1-F)^K(F)^{j-K} \tag{A.2}$$

In the above binomial distribution, it is assumed that the probability a user remain active is independent from the state of the other users. Similar assumption has been stated and justified in Reference 45, where the case of a single server serving N lines in cyclic order is studied. In that analysis, the number of lines is assumed constant during a cycle and is determined by a binomial distribution.

$q(i,K,j)$ = Pr [i idle users receive packet(s) out of (M−K) idle users during a frame given K users remain active out of j users requesting transmission at the beginning of a frame].

$$= \binom{M-K}{i}(\sigma_j)^i[1-\sigma_j]^{M-K-i} \tag{A.3}$$

$h(i,K,j)$ = Pr[K users remain active and i idle users receive packets during a frame, given j users requesting transmission at the beginning of a frame]

$$= \binom{M-K}{i}(\sigma_j)^i(1-\sigma_j)^{M-K-i} \times \binom{j}{K}(1-F)^K(F)^{j-K}$$

$$= q(i,K,j)C_K(j) \tag{A.4}$$

The System Markov chain is illustrated in Fig. A.1.a. Therefore, π_n, the steady state probability of having n users requesting transmission at the beginning of the frame is given by

$$P_n = \sum_{j=0}^{M} A_j(n)P_j \qquad 0 \leq n \leq M \tag{A.5}$$

where the transition probability $A_j(n)$ is the transition probability from state j to state n and is given by

$$A_j(n) = \sum_{K=0}^{\min(j,n)} \Pr\left[\begin{array}{l}\text{K users remain active and (n−k) idle}\\ \text{users receive packet(s) during a frame given j}\\ \text{users requesting transmission at the beginning}\\ \text{of a frame}\end{array}\right]$$

$$= \begin{cases} \sum_{k=0}^{\min(j,n)} h(n-k,k,j), & 1 \le j \le M,\ 0 \le n \le M \\ \begin{pmatrix} M \\ n \end{pmatrix} \sigma^n (1-\sigma)^{M-n}, & j = 0,\ 0 \le n \le M \end{cases} \tag{A.6}$$

3. The User Markov Chain

The state of the user is described by the number of packets in the buffer at the beginning of the frame. Let

n = number of packets in the buffer at the beginning of a frame

π_n = steady state probability of having n packets in the buffer at the beginning of a frame

$$g(i,L_j) = \Pr[\text{i packets arrive during the frame of length } L_j]$$

$$= \begin{pmatrix} M \\ i \end{pmatrix} \sigma^i (1-\sigma)^{M-i} \tag{A.7}$$

In Fig. A.2.b, we show the user Markov chain. Note that $B_{n-i}(n)$ is the transition probability from state (n-i) to state (n). Let us now determine the transition probabilities.

$$B_{n+1}(n) = \Pr[\text{One packet leaves and zero packet arrives during the frame of length } L_j]$$

$$= \sum_{j=1}^{M} g(0,L_j)P_j/(1-P_0)\ , \quad n = 0,1,2,\ldots \tag{A.8}$$

Notice that the initial value of the summation is j = 1, since at the beginning of the frame we have at least one user requesting transmission. For the same reason, we divide by $(1-P_0)$. Similarly,

$$B_n(n) = \Pr\,[\text{one packet leaves and one arrives during the frame of length } L_j]$$

$$= \begin{cases} = \left[\sum_{j=1}^{M} g(1,L_j)P_j\right]/(1-P_0)\,, & n \neq 0 \\ \left[\sum_{j=0}^{M-1} g(0,L_j)P_j\right]/(1-P_M)\,, & n = 0 \end{cases} \tag{A.9}$$

Note that for $n = 0$, we can have a maximum of (M-1) requesting users. Hence, we need to divide by the probability of having (M-1) or less requesting users. Similarly,

$$B_{n-i}(n) = \sum_{j} \Pr\,[\text{one packet leaves and } i+1 \text{ arrives during the frame length } L_j]$$

$$= \left[\sum_{j=i}^{M} g(i+1,L_j)P_j\right]/(1-P_0), \quad \begin{array}{l} n = 2,3,\ldots \\ i = 1,\min(n-1,M-1) \end{array} \tag{A.10}$$

and

$$B_0(n) = \left[\sum_{j=0}^{M-1} g(n,L_j)P_j\right]/(1-P_M), \quad n = 1,\ldots,M \tag{A.11}$$

Note that, we divide by $(1-P_M)$ for the same reasoning given for Eq. (A.9). Therefore, we have now a set of simultaneous coupled non-linear equations in π_0, π_1 and $[P_n]$. These equations can be solved numerically as described in the dissertation. Hence the average delay for a packet (D) which is the sum of the average time (W) in the buffer, W, plus the average service time(s) in the frame can be determined as the follows.

$$D = W + S \tag{A.12}$$

where

$$W = \frac{Q}{\sigma}$$

$$Q = \sum_{n=0}^{\infty} n \cdot \pi_n$$

$$S = \frac{1}{2} \sum_{j=0}^{M} j \cdot P_j$$

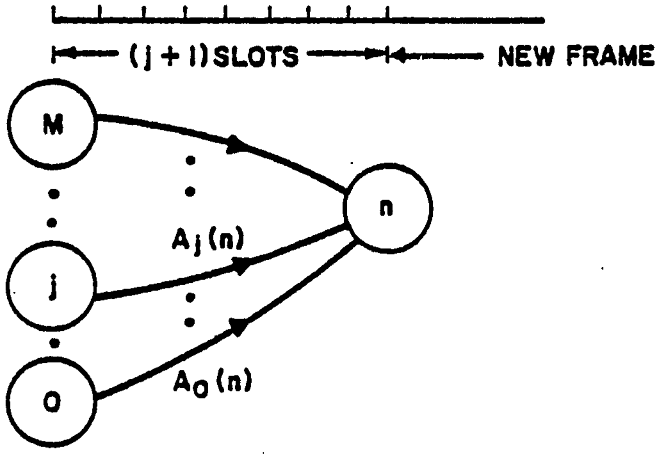

FIGURE A-1. SYSTEM MARKOV CHAIN

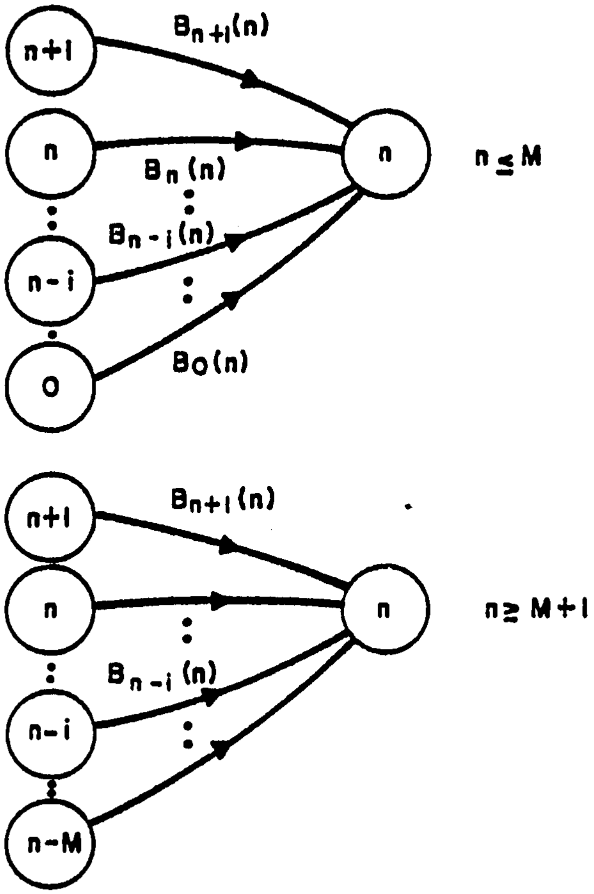

FIGURE A-2. USER MARKOV CHAIN