DIGITAL SYSTEMS LABORATORY

STANFORD ELECTRONICS LABORATORIES
DEPARTMENT OF ELECTRICAL ENGINEERING
STANFORD UNIVERSITY · STANFORD, CA 94305

SU-SEL 77-02 I

PERFORMANCE ANALYSIS OF COMPUTER COMMUNICATION NETWORKS VIA RANDOM ACCESS CHANNELS

by

Philip S. Yu

April 1977

Technical Report No. 137

The work described herein was supported in part by the Ballistic Missile Defense Systems Command under contract no. DASG60-77-C-0073. Computer time was made available by the Stanford Linear Accelerator Center.

SU-SEL 77-021

PERFORMANCE ANALYSIS OF COMPUTER COMMUNICATION NETWORKS VIA RANDOM ACCESS CHANNELS

by
Philip S. Yu

April 1977

Technical Report No. 137

Digital Systems Laboratory
Departments of Electrical Engineering and Computer Science
Stanford University
Stanford, CA 94305

The work described herein was supported in part by the Ballistic Missile Defense Systems Command under contract no. DASG60-77-C-0073. Computer time was made available by the Stanford Linear Accelerator Center.

Digital Systems Laboratory
Departments of Electrical Engineering and Computer Science
Stanford University
Stanford, CA 94305

Technical Report No. 137

April 1977

PERFORMANCE ANALYSIS OF COMPUTER COMMUNICATION NETWORKS VIA RANDOM ACCESS CHANNELS

by

Philip S. Yu

ABSTRACT

The field of computer communication networks has grown very rapidly in the past few years. One way to communicate is viamultiple access broadcast channels. A new class of random access schemes referred to as the M_p-persistent CSMA scheme is proposed. It incorporates the nonpersistent CSMA scheme and l-persistent CSMA scheme, both slotted and unslotted versions, as its special cases with $p=0$ and 1, respectively. The performance of the M_p-persistent CSMA scheme under packet switching is analyzed and compared with other random access schemes. By dynamically adjusting p, the unslotted version can achieve better performance in both throughput and delay than the currently available unslotted CSMA schemes under packet switching. Furthermore, the performance of various random access shcemes under message switching is analyzed and compared with that under packet switching. In both slotted and unslotted versions of the MC-persistent CSMA scheme, the performance under message switching is superior to that under packet switching in the sense that not only the channel capacity is larger but also the average number of retransmissions per successful message under message switching is smaller than that per successful packet under packet switching. In dynamic reservation schemes, message switching leads to larger channel capacity. However, in both slotted and unslotted versions of the ALOHA scheme, the channel capacity is reduced when message switching is used instead of packet switching. This phenomenon may also happen in the M_p-persistent CSMA scheme as p deviates from 0 to 1 for certain distributions of message length. Hence, the performance under message switching may be superior to or inferior to that under packet switching

depending upon the random access scheme being used and the distribution of message length (usually a large coefficient of variation of message length implies a large degradation of channel capacity in this case) for certain random access schemes. Nevertheless, for radio channels, message switching can achieve larger channel capacity if appropriate CSMA schemes are used. A mixed strategy which is a combination of message switching and packet switching is proposed to improve the performance of a point to point computer communication network when its terminal access networks communicate via highly utilized radio channels.

Acknowledgement: The author would like to thank Professor Michael J. Flynn, Dr. Hisashi Kobayashi and Professor Andrew C. Yao for their helpful comments and suggestions.

The work described herein was supported in part by the Ballistic Missile Defense Systems Command under contract no. DASG60-77-C-0073. Computer time was made available by the Stanford Linear Accelerator Center.

TABLE OF CONTENTS

1. INTRODUCTION

The field of computer communication networks has grown very rapidly in the past few years. The advantage of computer communication networks is that the special resources and capabilities built up at each of the many separate computer facilities can profit by resource and load sharing. However, the complexities of the issues one faces in creating a network are staggering. Hence modeling and performance evaluation have become one of the most crucial issues in the design and operation of a network.

Computer communication networks may be conveniently partitioned into two separate subnetworks. The communication subnetwork which consists of long haul facilities for communication among geographically scattered computers and resources, and the terminal access networks which provide local distribution and terminal computer communications. Communication channels can be classified into dedicated channels and shared or multiple access channels. The store and forward communication subnetworks of ARPANET [1] and CYCLADES [20] via terrestrial links are typical examples using dedicated channels. Communications via multiaccess satellite channels provide an alternative solution to the implementation of communication subnetworks. In fact, the ARPANET has now been extended by satellite to Hawaii and to a few nodes in Europe. Although satellite communication is subject to an intrinsic propagation delay of about 0.26 seconds, its low cost compared with that of terrestrial links makes it still attractive. For terminal access networks, comparing with communications via a dedicated channel for each terminal, communications via a multiaccess radio channel by all terminals provide not only a solution which handles geographical dispersion of terminals and reduces the cost at the same time, but also an effective solution when terminals are mobile [1].

Various access schemes can be employed to handle transmissions via multiaccess channels. They can be completely random, [e.g. ALOHA or carrier sense multiple access (CSMA) scheme], randomly requesting only, [e.g. dynamic reservation scheme], fixed [e.g. time-division multiple access (TDMA) or frequency-division multiple access (FDMA) scheme] or completely centrally controlled [e.g. polling scheme]. As pointed out by Tobagi and Kleinrock [13], random access schemes lead to better performance for bursty users.

The difficulty encountered in analyzing the performance of computer communication networks via random access channels is that two or more messages overlapping in transmission time will collide with each other and lead to mutual destruction. This "collision" phenomenon makes the traditional queueing theory not applicable. While communicating via random access channels, two different strategies can be used, namely message switching and packet switching. Packet switching is basically the same as message switching except that each message is decomposed into smaller pieces called packets and then transmitted one by one instead of transmitting the message as a whole as in message switching. The two strategies are also used in store and forward computer communication networks via terrestrial links. For store and forward computer communication networks via terrestrial links, the major advantage of packet switching is the "pipelining" effect, i.e. different packets of the same message may be in transmission at different channels simultaneously, if the transmission requires multiple hops through the network. Hence the transmission delay under packet switching may be greatly reduced and the transmission overhead due to the extra header information contained in each packet will be overcome [1]. When communications are via multiaccess radio channels, the pipelining advantage of packet switching no more exists when all users are in line of sight. The relative performance of packet switching and message switching depends upon their susceptibilities to collision. In the following comparison of packet switching and message switching, we will neglect the overhead of the header information contained in each packet under packet switching.

In section 2, we survey the characteristic of various random access schemes, namely the various ALOHA and CSMA schemes and the analytic results on their performance. In section 3, we propose a new class of carrier sense multiple access schemes referred to as the Mp-persistent CSMA scheme. It incorporates the nonpersistent CSMA and l-persistent CSMA schemes, both slotted and unslotted versions, as its special cases with p = 0 and 1 respectively. Both the slotted and unslotted versions are analyzed and compared with other schemes. By varying the p dynamically, we can obtain the optimum Mp-persistent CSMA scheme.

In section 4, the performance of various random access schemes under message switching is examined and compared with that under packet switching. We first analyze the performance of the M_0-persistent CSMA scheme under message switching. Although the exact analysis is hard to conduct in general, we do obtain the upper bound and lower bound of throughput under the distribution free assumption on the number of packets contained in each message for both slotted and unslotted versions. Since the bounds are very close to each other as long as the channel is not nearly saturated, this gives-us a good estimation of the throughput. For both slotted and unslotted versions of the M_0-persistent CSMA scheme, message switching not only leads to larger channel capacity but also smaller average number of retransmissions per successful message compared with that per successful packet under packet switching. The analysis of the Mp-persistent CSMA scheme for $p \neq 0$ is hard to conduct. Nevertheless, by examining the simulation results on the Ml-persistent CSMA scheme, we find that as p goes from 0 to 1, the performance under message switching depends on the distribution of message length and may become inferior to that under packet switching.

We then examine the ALOHA scheme and show that the channel capacity of unslotted and slotted ALOHA scheme is generally reduced when message switching is used instead of packet switching. Finally, we summarize the various dynamic reservation schemes which generally lead to better performance under message switching.

In section 5, we try to combine various techniques to evaluate the performance of a computer communication network whose terrestrial network has the same topology as the communication network, CIGALE, within CYCLADES. Communications between terminals and the network are through a multiaccess broadcast-channel. Due to the insight obtained from performance analysis, we find that a mixed strategy which is a combination of packet switching and message switching will lead to better performance when the utilization of the radio channel is high and packet switching will be preferable when the utilization of the radio channel is low. In section 6, we draw the conclusion.

2. SURVEY OF VARIOUS RANDOM ACCESS SCHEMES

In this section, we examine the characteristic of various random access schemes proposed in the past and summarize the analytic results on their performance. Although all the access schemes considered in this section can be applied to ground radio communication, only those variations of the ALOHA scheme can be applied to satellite communication. This is due to the fact that the propagation delay of the satellite communication, 0.26 sec, is much larger than the packet transmission time. Hence, any access scheme which tries to detect whether the channel is busy before making transmissions does not make sense under satellite communications. In the discussion of communications via ground radio channels, we only consider the case where all terminals are in line of sight. For more elaborate cases see [12].

The common property shared by all random access schemes is that they take advantage of the "law of large number" [1] and improve the performance under bursty data from users which tend to generate demands at a very low duty cycle. As it is well known, the "law of large number" states that the collective demand of a large population of random users is very well approximated by the sum of average demands required by that population. That is to say the statistical fluctuations in any individual's demands are smoothed out in the large population case so that the total demand process appears to be a more deterministic demand process. Kleinrock [18] [1] shows that the concept of large shared single resources, e.g. random access channel, leads to improvements in mean response time due to the law of large number.

2.A The ALOHA Scheme

The ALOHA scheme [6], [1] appears to have been the first random access scheme to employ wireless communications. There are three different approaches to the solution of packet switching problem under ALOHA scheme. The first one has come to be known as the pure ALOHA or unslotted ALOHA scheme [10] in which users can transmit at any time they wish. If, after one propagation delay, they fail to hear their successful transmissions, they know a collision must have occurred and retransmit the packets after random retransmission delays. The reason for random retransmission delays is that if all users retransmit immediately upon hearing a conflict or after a fixed amount of delay, they are sure to conflict again.

The second scheme is referred to as slotted ALOHA scheme [3], [1]. In this scheme, we slot the time into segments whose duration is exactly equal to the transmission time of a single packet. All packets are required to begin their transmissions only at the beginning of a slot. The advantage of the slotted ALOHA scheme over the pure ALOHA scheme is that collisions are restricted to a single slot duration, hence, the probability of collisions is reduced.

The third scheme is referred to as reserved ALOHA scheme and will be discussed later with the other dynamic reservation schemes together.

To simplify the notation, from now on we will assume that the transmission time of each packet takes one unit of time. Let S be the average number of packets generated per unit time or per packet transmission time, where the arrival process is assumed to be Poisson. Under steady state conditions, S can also be referred to as the channel throughput rate. Furthermore, it can also be viewed as the channel utilization. The maximum achievable throughput for an access scheme is called the channel capacity of that scheme. Since conflicts do occur, the traffic offered to the channel from our collection of users consists of not only new packets, but also previously collided packets. We will use G to denote the mean offered traffic rate. Clearly $G \geq S$.

The following two assumptions have been made for analytic tractability [11].

Assumption 1. The average retransmission delay is large compared to the transmission time.

Assumption 2. The interevent times of the point process defined by both the start time of all the packets and the retransmission times are independent and exponentially distributed.

Clearly, the second assumption does not quite hold. However, the simulation results in [11] show that performance results based on this assumption are excellent approximations. Moreover, in the context of slotted ALOHA, assumption 2 has been shown to be satisfied as mean retranmission delay approaches infinite [16].

We now summarize the basic results on throughput for the ALOHA scheme under the above assumptions:

in the pure ALOHA scheme:

$$S = G e^{-2G} \tag{2.1}$$

in the slotted ALOHA scheme:

$$S = G e^{-G} \tag{2.2}$$

As we can see from equations (2.1) and (2.2), the channel capacity of the slotted ALOHA scheme is $^1/e = 0.368$ (at $G = 1$) which is twice as large as that of the pure ALOHA scheme.

We introduce at this point the expected packet delay D defined to be the average time from when a packet is generated until it is successfully received [11]. We assume that acknowledgement packets are always correctly received. Let α denote the transmission time of the acknowledgement packet which is usually transmitted through a separate channel and let a denote the one way propagation delay. If we neglect the small processing time required to perform the sum check and to generate the acknowledgement packet, the time out for receiving a positive acknowledgement is $1 + a + \alpha + a$.

Let δ be the average retransmission delay. Define

$$R \overset{\Delta}{=} 1 + 2a + \delta + \alpha \qquad (2.3)$$

Since $(G/S - 1)$ is the average number of retransmissions required, the average delay D is given by

$$D = (G/S - 1)(R + \bar{d}) + \bar{d} + 1 + a \qquad (2.4)$$

where $\bar{d}$ is 1/2 for the slotted version and 0 otherwise.

2.B The CSMA Scheme [11]

The CSMA scheme is proposed by Kleinrock and Tobagi [11] to be used in communications via radio channels. The fundamental difference between ground radio channel and satellite channel is the propagation delay, In ground radio channel, the propagation delay is usually much smaller than the packet transmission time. Hence, we can permit the user to listen to the channel and if the carrier signal is heard, then the user realizes that the channel is in use by some other users and will postpone its transmission until the channel is sensed to be idle. This is referred to as the carrier sense multiple access scheme (CSMA). Depending upon the action taken by the users after sensing the channel, two kinds of access schemes have been introduced by Kleinrock and Tobagi [11], namely, the nonpersistent CSMA scheme and the p-persistent CSMA scheme.

We first consider the nonpersistent CSMA. Here, the idea is to reduce the interference among packets by always rescheduling any packet finding the channel in busy state upon its arrival. Specifically, a ready user senses the channel and operates as follows:

(1) If the channel is sensed idle, it transmits the packet immediately.

(2) If the channel is sensed busy, it schedules the retransmission of the packet according to the retransmission delay distribution. At this new point in time, it senses the channel and repeats the same procedure again.

We next consider the l-persistent CSMA scheme, which is a special case of the p-persistent CSMA scheme and achieves reasonable throughput by never letting the channel idle if some ready user is available. To be more precise, a ready user senses the channel and operates as follows:

(1) If the channel is sensed idle, it transmits the packet with probability 1.

(2) If the channel is sensed busy, it waits until the channel goes idle, (i.e. persisting on sensing) and only then transmits the packet with probability 1.

In both the nonpersistent CSMA and l-persistent CSMA schemes, we can also have slotted versions which will be referred to as the slotted nonpersistent CSMA scheme and the slotted l-persistent CSMA scheme, respectively. In the slotted versions, the time axis is slotted and the slot size is set to a, the propagation delay. Recall in the slotted ALOHA scheme the slot size is the packet size not the propagation delay. All ready users are synchronized and forced to sense the channel at the beginning of the next time slot and take the appropriate action according to the access scheme.

The above l-persistent and nonpersistent CSMA schemes differ mainly in the way to handle ready users who sense the channel busy. In the case of l-persistent CSMA, users will persist on waiting and then transmit after the channel becomes idle. If there are more than one user waiting for transmission, with probability one conflict will occur. Hence, for interference reduction and throughput improvement, p-persistent CSMA is proposed which generalizes the l-persistent CSMA in the sense that after sensing the channel idle, users will transmit with probability p. The parameter p will be chosen so as to reduce the level of interference while keeping the gap between two consecutive nonoverlapped transmissions as small as possible. To be more precise, the time axis is slotted with slot size a and the system is synchronized such that all packets are required to start their transmissions at the beginning of a slot as in slotted l-persistent or nonpersistent CSMA. A ready user senses the channel and operates as follows:

1. If the channel is sensed idle, it transmits the packet with probability p. With probability l-p, the user delays the transmission of the packet by one slot time, a. If at this new point in time, the channel is still sensed idle, the same procedure is repeated. Otherwise, the user will schedule the retransmission of the packet according to the retransmission delay distribution since some packet has already started transmission.
2. If the channel is sensed busy, it waits until the channel becomes idle

and then operates as above.

We now summarize the basic results on throughput for various CSMA schemes under packet switching [11]. Let G be the offered packet traffic and S be the packet arrival rate as before. Under steady state, S also represents the throughput of the system. Under assumptions 1 and 2 mentioned before, we get

nonpersistent CSMA

$$S = \frac{Ge^{-aG}}{G(1 + 2a) + e^{-aG}} \tag{2.5}$$

slotted nonpersistent CSMA

$$S = \frac{aGe^{-aG}}{1 - e^{-aG} + a} \tag{2.6}$$

l-persistent CSMA

$$S = \frac{G\left[1 + G + aG\left(1 + G + {}^{aG}/2\right)\right] e^{-G(1 + 2a)}}{G(1 + 2a) - (1 - e^{-aG}) + (1 + aG)\, e^{-G(1 + a)}} \tag{2.7}$$

slotted l-persistent CSMA

$$S = \frac{Ge^{-G(1 + a)}\,(1 + a - e^{-aG})}{(1 + a)(1 - e^{-aG}) + ae^{-G(1 + a)}} \tag{2.8}$$

(slotted) p-persistent CSMA

$$S = \frac{(1 - e^{-aG})\,(P_S'\Pi_0 + P_S(1 - \Pi_0))}{(1 - e^{-aG})\,(a\bar{t}'\Pi_0 + a\bar{t}(1 - \Pi_0) + 1 + a) + a\Pi_0} \tag{2.9}$$

where P_S', P_S, $\bar{t}'$, $\bar{t}$ and Π_0 are defined in [11] . Since the P_S', P_S, $\bar{t}'$ and $\bar{t}$ are not in closed forms, it is too complicated to reproduce them here.

The formula for average delay time can be simplified by making the following assumption [11]. When a packet is blocked, it behaves as if it could transmit and learned about its blocking only a units of time after the end of its virtual transmission where α is the transmission time of the acknowledgement packet. The average delay time for various CSMA schemes has the following form

$$D= \left(\frac{G}{S} - 1\right) (R + \bar{d}) + \bar{d} + 1 + a \qquad (2.10)$$

where $\bar{d}$ is the mean pretransmission delay and R is defined in (2.3) to be (a + 2a + 1). The exact expressions for mean pretransmission delays of various CSMA schemes can be found in [11].

From now on, we will refer to the crucial factor ($\frac{G}{S}$ - 1) in the delay equation as the number of rctransmissions per successful transmission although it actually represents the total number of retransmissions and schedulings for convenience. In fact, scheduling may be viewed as virtual transmission. Similarly, we will refer to $\frac{G}{S}$ as the number of transmissions per successful transmission.

2.C The Dynamic Reservation Scheme

The dynamic reservation scheme has larger channel capacity under message switching. Most of the analytic results available on dynamic reservation scheme is under message switching. Since the emphasis in this section is primarily on packet switching, we will delay the discussion of dynamic reservation scheme until section 4.

3. THE Mp-PERSISTENT CSMA SCHEME

3.A Definition

In this section, we propose a new class of carrier sense multiple access schemes which incorporate the nonpersistent CSMA and l-persistent CSMA schemes, both slotted and unslotted versions proposed by Kleinrock and Tobagi [11] as its special cases. Although it does not incorporate the p-persistent CSMA schemewhen $p \neq 1$, they are similar in the sense that bothschemes are trying to reduce the interference among users sensing a channel busy by approximately $1 - p$. Hence, we refer to this new class of CSMA schemes as the modified p-persistent CSMA scheme or the M_p-persistent CSMA scheme. It has the following characteristics:

(i) Both slotted and unslotted versions of the Mp-persistent CSMA scheme lead to closed form expressions for throughput equations. This makes the determination of optimum p to operate under a given load a much simpler task as compared with that of the p-persistent CSMA scheme.

(ii) The optimum unslotted Mp-persistent CSMA scheme achieves larger channel capacity and smaller transmission delay than both nonpersistent CSMA scheme and l-persistent CSMA scheme, the two currently available unslotted CSMA schemes. Even without varying p dynamically, we can choose appropriate p to achieve larger channel capacity and smaller transmission delay than the nonpersistent CSMA scheme.

(iii) The channel capacity of the optimum slotted Mp-persistent CSMA is larger than that of the optimum (slotted) p-persistent CSMA.

However, its transmission delay may be inferior to that of the optimum (slotted) p-persistent scheme.

Under the nonslotted Mp-persistent CSMA scheme, a ready user senses the channel and operates as follows:

1. If the channel is sensed idle, it transmits the packet.
2. If the channel is sensed busy, then with probability (l-p), it schedules the retransmission of the packet according to the retransmission delay distribution and with probability p, it waits (i.e. persists on sensing) until the channel goes idle and only then transmits the packet.

Again, wc can have slotted version. In the slotted version, the time axis is slotted and the slot size is taken to be a, the propagation delay, as in the slotted nonpersistent or slotted l-persistent CSMA schemes.

From the above definition, nonpersistent CSMA is a special case of the Mp-persistent CSMA scheme when P = 0 and l-persistent CSMA is a special case of the Mp-persistent CSMA scheme when P = 1, for both slotted and unslotted versions. Furthermore, since channel capacity of the optimum p-persistent CSMA scheme is smaller than that of the non-persistent CSMA scheme in the normal operation range of the CSMA scheme [1], this explains the first part of the third characteristic of the Mp-persistent CSMA scheme.

3.B Performance Analysis of the Slotted Mp-persistent CSMA Scheme Under Packet Switching

In this section we examine the performance of the slotted Mp-persistent CSMA scheme. The performance of the unslotted version will be examined in the next subsection.

Let S and G be the steady state throughput and offered traffic as before. The following theorem expresses the throughput in terms of offered traffic for the slotted Mp-persistent CSMA scheme.

Theorem 3-1

For a given offered packet traffic G and a given value of the parameter p, the throughput equation for slotted Mp-persistent CSMA under packet switching is:

$$S = \frac{pG + aG - pGe^{-aG}}{a + (1 + a)(e^{(a+p)G} - e^{pG})} \qquad (3.1)$$

Proof

G denotes the arrival rate of new and rescheduled packets. Only a fraction of it constitutes the channel traffic, since a packet which finds the channel busy is rescheduled with probability 1-p without being transmitted. In this slotted version, if two packets conflict, they will overlap completely. Consider the time axis in Fig. 3.1 and let t be the start of a time slot. Assume packets arrive during its preceeding slot which is in an idle period.

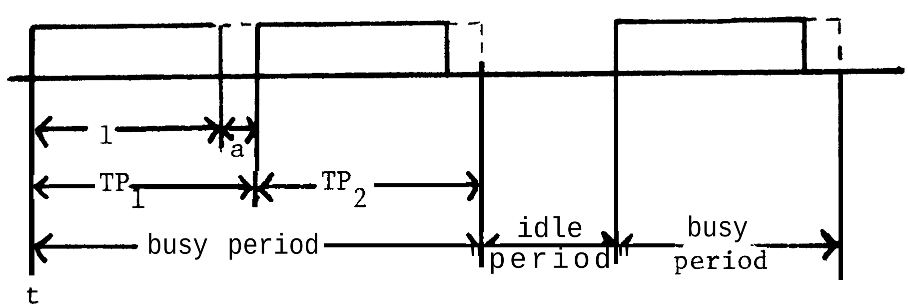


Fig. 3.1 SLOTTED Mp PERSISTENT CSMA: BUSY AND IDLE PERIODS

An idle period is the period of time the channel is idle and consists of at least one slot time, i.e. a units of time. A busy period is defined to be the time between two successive idle periods. The new arrivals will start transmissions at time t and a new busy period begins- At the end of a transmission period, there might be some packets pending for transmission if

(1) some packets arrive during $(t + 1, t + 1 + a]$

or

(2) some packets arrive during $(t, t + 1]$ and haven't been scheduled for retransmission.

The pending packets will immediately begin another transmission period at the end of the current transmission period. If there is no pending packet, a new idle period will begin. As we shall see the assumption that with Probability' p the arrival packets which sense the channel busy will persist on sensing is equivalent to replacing the arrival process in $(t,t+1]$ by another Poisson process with rate pG. Hence, the total arrival rate per transmission period is equal to $(p+a)G$. Similar result is derived in theorem 3.3. We postpone the derivation since this general result is not needed in the proof of this theorem. Let N denote the number of packets accumulated at the end of a transmission period and let $q_n \triangleq Pr\{N=n\}$ be the distrkbution of the number of packets accumulated. The probability that the busy period will terminate at the end of the transmission period is given by

$$q_0 = Pr\{\text{no arrival during } (t+1, t+1+a]\}\ Pr\{\text{all packets arriving during } (t,t+1] \text{ are rescheduled}\}$$

$$= e^{-aG} \sum_{n=0}^{\infty} (1-p)^n \frac{G^n}{n!} e^{-G}$$

Using the fact that $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$, after simplification, we get

$$q_0 = e^{-(a+p)G} \tag{3.2}$$

In order to find the probability of success over a transmission period, Π_i, one has to distinguish between two cases: i = 1 and i ≠ 1. We first look at the case where i ≠ 1. Clearly, the successful probability is the probability of exactly one packet pending for transmission and is given by

Π_1 = (Pr {the 0n1y pending packet arrives before the last time slot of the TP*}

+ Pr {the only pending packet arrives during the last time slot of the TP}) / Pr {at least one packet pending at the end of the TP}

$$= \frac{e^{-aG}\sum_{n=1}^{\infty}\binom{n}{1}p(1-p)^{n-1}\frac{G^n}{n!}e^{-G}+aGe^{-aG}\sum_{n=0}^{\infty}(1-p)^n\frac{G^n}{n!}e^{-G}}{1-\Pi_0}$$

Changing the index of summation and using the Taylor's series expansion of e^x, we get

$$\Pi_1 = \frac{(p+a)G\,e^{-(a+p)G}}{1-e^{-(a+p)G}} \tag{3.3}$$

Furthermore, in the first transmission period, the successful probability is given by

$$\Pi_1' = \frac{\text{Pr (exactly one arrival during the last time slot}\}}{\text{Pr (at least one arrival during the last time slot)}}$$

$$= \frac{aGe^{-aG}}{1-e^{-aG}} \tag{3.4}$$

The length of each transmission period is always equal to 1 + a.

*transmission period

Since the traffic process is an independent one, the number of transmission periods in a busy period is geometrically distributed with a mean equal to $\frac{1}{q_0}$. Similarly, the number of slots in an idle period is also geometrically distributed with a mean equal to $\frac{1}{1 - e^{-aG}}$. Thus the average idle period, $\bar{I}$, is given by

$$\bar{I} = \frac{a}{1 - e^{-aG}} \tag{3.5}$$

and the average busy period, $\bar{B}$, is given by

$$\bar{B} = (1 + a)e^{(a + p)G} \tag{3.6}$$

Let $\bar{U}$ be the average time during a busy period that the channel is used without conflicts, then

$$\bar{U} = \Pi_1' + (\frac{1}{q_0} - 1)\,\Pi_1 \tag{3.7}$$

Substituting (3.2), (3.3), and (3.4) into (3.7) we get

$$\bar{U} = pG + \frac{aG}{1 - e^{-aG}} \tag{3.8}$$

Using renewal theory arguments, the average channel utilization is given by [11]

$$\bar{S} = \frac{\bar{U}}{\bar{B} + \bar{I}} \tag{3.9}$$

Finally, sutstituting (3.5), (3.6), and (3.8) into (3.9), we obtain (3.i) after simplification.

By setting p = 1 and 0 in (3.1) we obtain equation (2.8) and (2.6), the throughput equations for the slotted l-persistent CSMA scheme and the nonpersistent CSMA scheme, respectively, as expected.

At this point, we proceed to investigate the average delay. Recall our definition of R given in (2.3). The following theorem gives the delay equation of the slotted Mp-persistent CSMA scheme under packet switching.

Theorem 3.2

For a given offered traffic G and a given value of parameter P, the delay equation for slotted Mp-persistent CSMA under packet switching is given by

$$D = \left(\frac{G(P_I + P_W)}{S} - 1 \right) (R + \bar{d}_s) + \frac{G(1 - P_I - P_W)}{S} \delta + \bar{d}_s + 1 + a \qquad (3.10)$$

Where

$$\bar{d}_s = \frac{a^2 + (e^{aG}-1)\, e^{pG}(a^2 + (1+2a)p)}{2(a + (p+a)e^{pG}(e^{aG}-1)} \qquad (3.11)$$

$$P_I = \frac{a + ae^{pG}(e^{aG}-1)}{a + (1+a)e^{pG}(e^{aG}-1)} \qquad (3.12)$$

and

$$P_W = \frac{p\ e^{pG}(e^{aG}-1)}{a + (1+a)e^{pG}(e^{aG}-1)} \qquad (3.13)$$

Proof

Consider the time axix in Fig. 3.1. Each transmission period has the same length (1+a). An arrival packet may either be transmitted after a pre-transmission delay or be scheduled for retransmission if the channel is busy, and will be transmitted after a pretransmission delay if the channel is idle. The probability that an arrival packet will detect the channel idle is

$$P_I = \frac{\bar{I} + \frac{1}{q_0} a}{\bar{I} + \bar{B}}$$

From (3.2),(3.5), (3.6) and the above equation, we obtain (3.12). The corresponding mean pretransmission delay is $\frac{a}{2}$.

The probability that an arrival packet will detect the channel busy and persist on waiting is

$$P_W = \frac{p/q_0}{\bar{I} + \bar{B}}$$

From (3.2). (3.5), (3.6) and the above equation, we obtain (3.13).

The corresponding mean pretransmission delay is $(\frac{1}{2} + a)$. Hence the overall mean pretransmission delay is

$$\bar{d}_s = \frac{P_I}{P_I+P_W} \frac{a}{2} + \frac{P_W}{P_I+P_W} (\frac{1}{2} + a)$$

Combining (3.12), (3.13) and the above equation together, we get (3.11). Since the delay before next channel sensing is

$$R^* = \begin{cases} R + \bar{d}_s, & \text{if the packet is transmitted unsuccessfully} \\ \delta & \text{if the packet is resecheduled upon arrival} \end{cases}$$

the retransmission delay is given by (3.10). Furthermore, if we choose to treat all packet arrivals in a uniform way, we may assume when a packet is schedule for retransmission upon arrival, it behaves as if it could transmit and learned about its rescheduling only α units of time after the end of its virtual transmission. With this simplification, the delay equation is given by the following corollary.

Corollary 3.2:

Under the above assumption, the delay equation for slotted M_p-persistent CSMA under packet switching is given by

$$D = (\frac{G}{S} - 1)(R + \bar{d}'_s) + \bar{d}'_s + 1 + a \tag{3.14}$$

where

$$\bar{d}'_s = \frac{a^2+(e^{aG}-1)\, e^{pG}(a^2+(1+2a)p)}{2(a+(1+a)e^{pG}(e^{aG}-1))} \tag{3.15}$$

In Fig. 3.2a and 3.3a, we plot, versus G, the channel throughput, S, of-the slotted M_p-persistnet CSMA scheme for various values of p when a = 0.01 and 0.1, respectively. As we can see that the channel capacity achieves its maximum value when p equals to zero and deteriorates as p increases. In Fig. 3.2b and 3.3b, we plot the number of transmissions per successful transmission, G/S, versus the arrival rate for various values of p when a = 0.01 and 0.1, respectively. As we can see, for a given traffic load

FIG. 3.2.A THRØUGHPUT IN SLØT. VERSIØN (A=0.01)

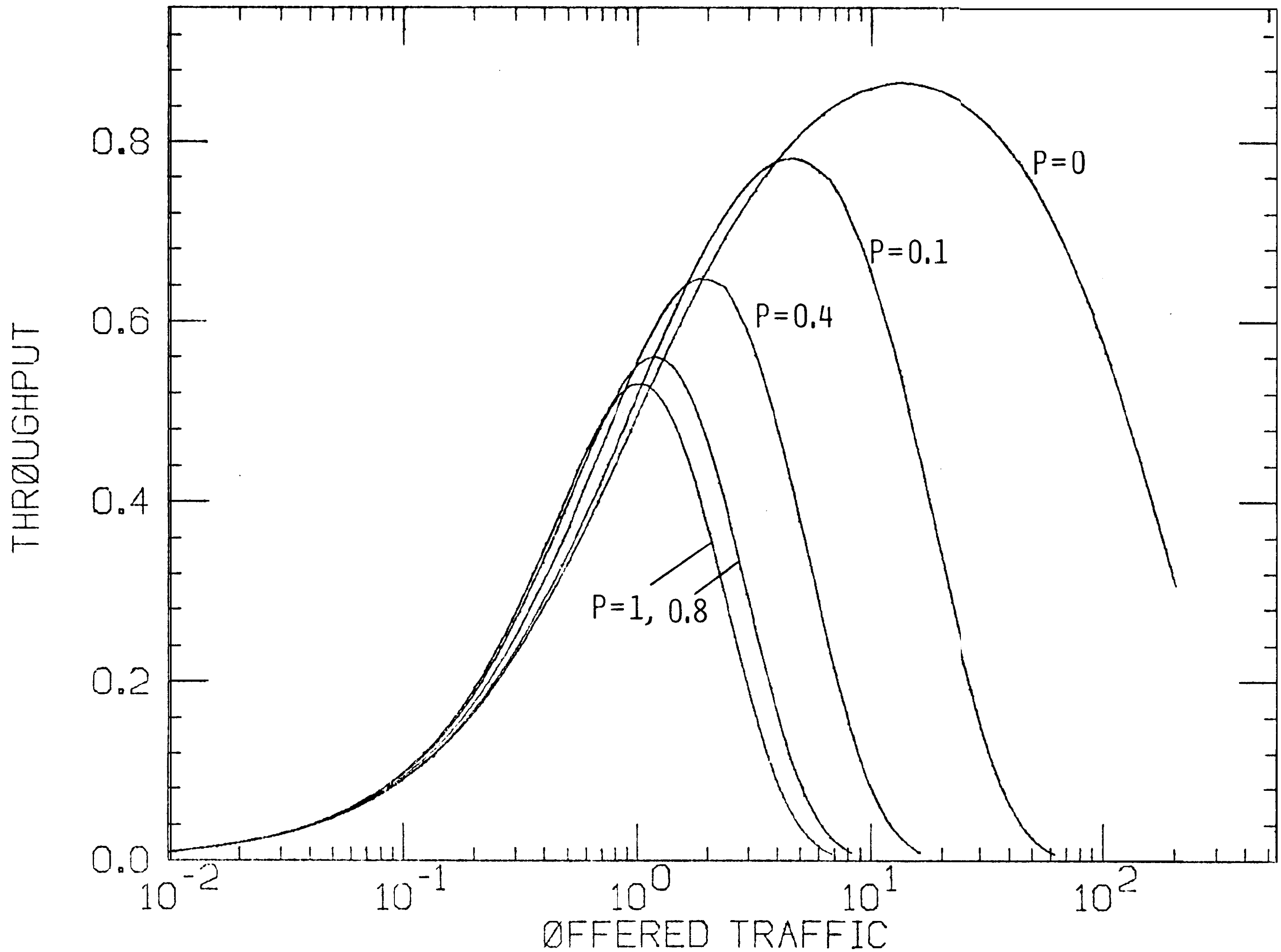


FIG.3.2.B NØ. ØF TRANS. IN SLØT. VERSIØN(A=0.01)

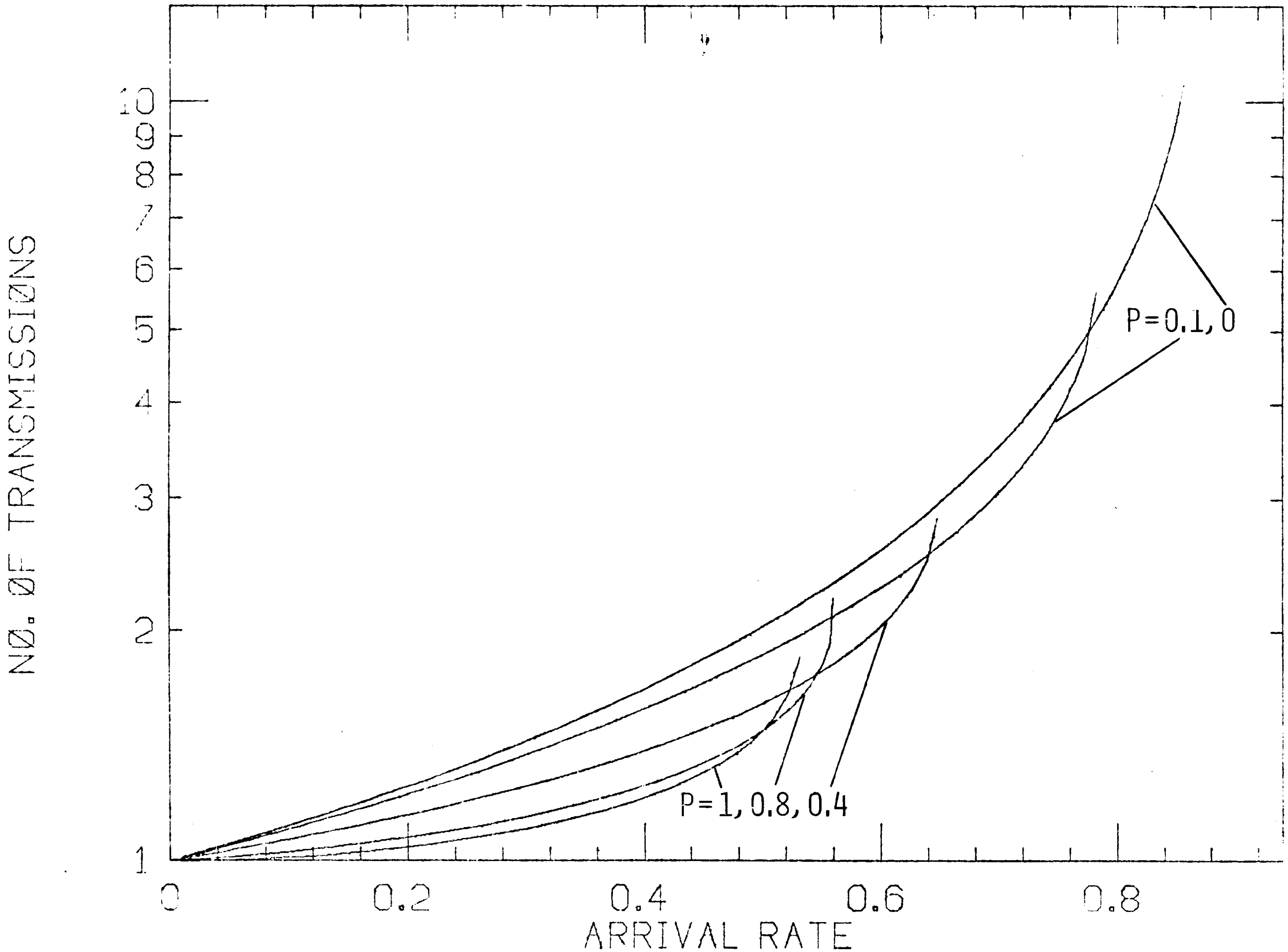


FIG. 3.3.A THRØUGHPUT IN SLØT. VERSIØN (A=0.1)

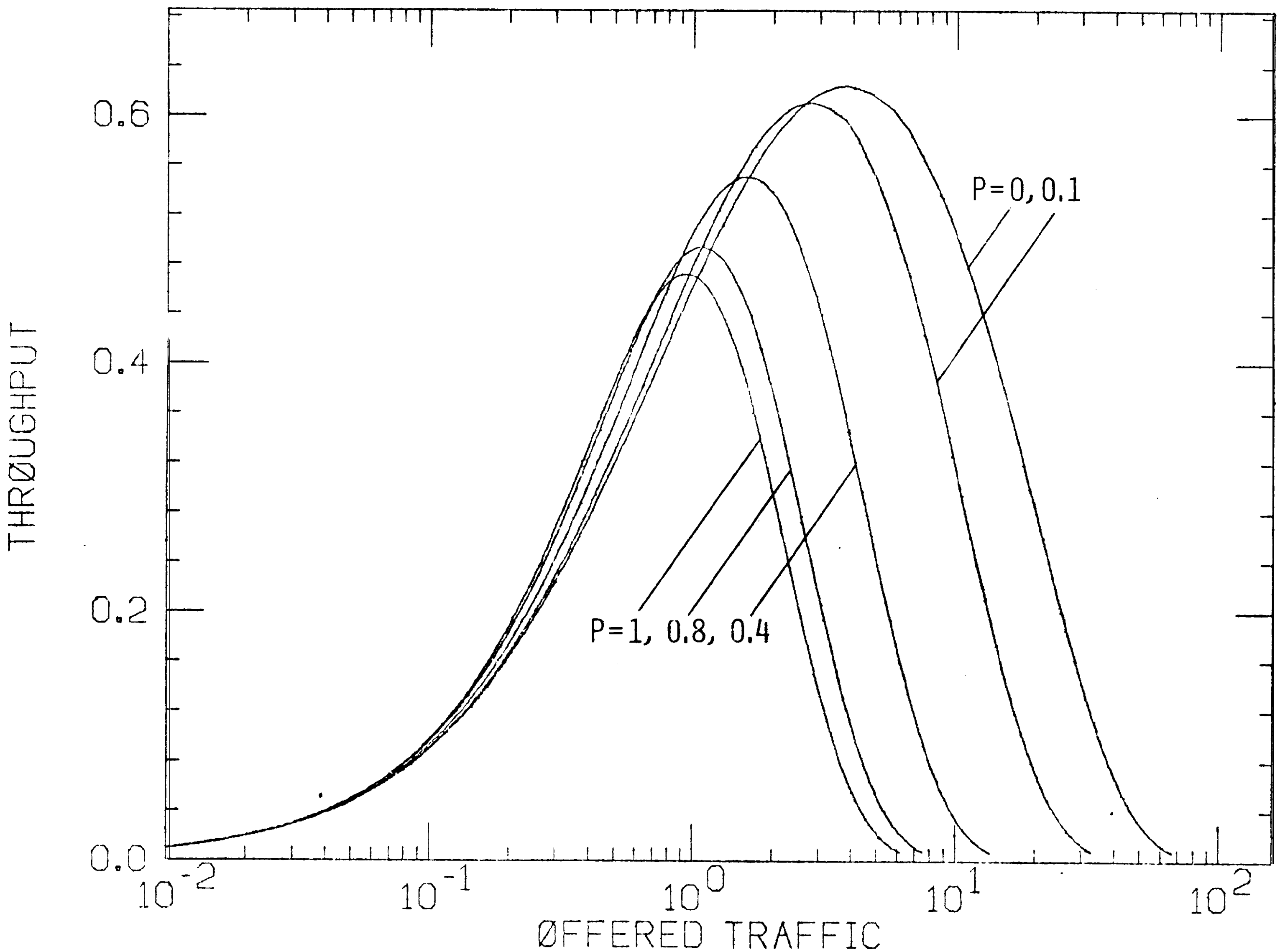


FIG. 3.3.m NØ. ØF TRANS. IN SLØT. VERSIØN(A=0.1)

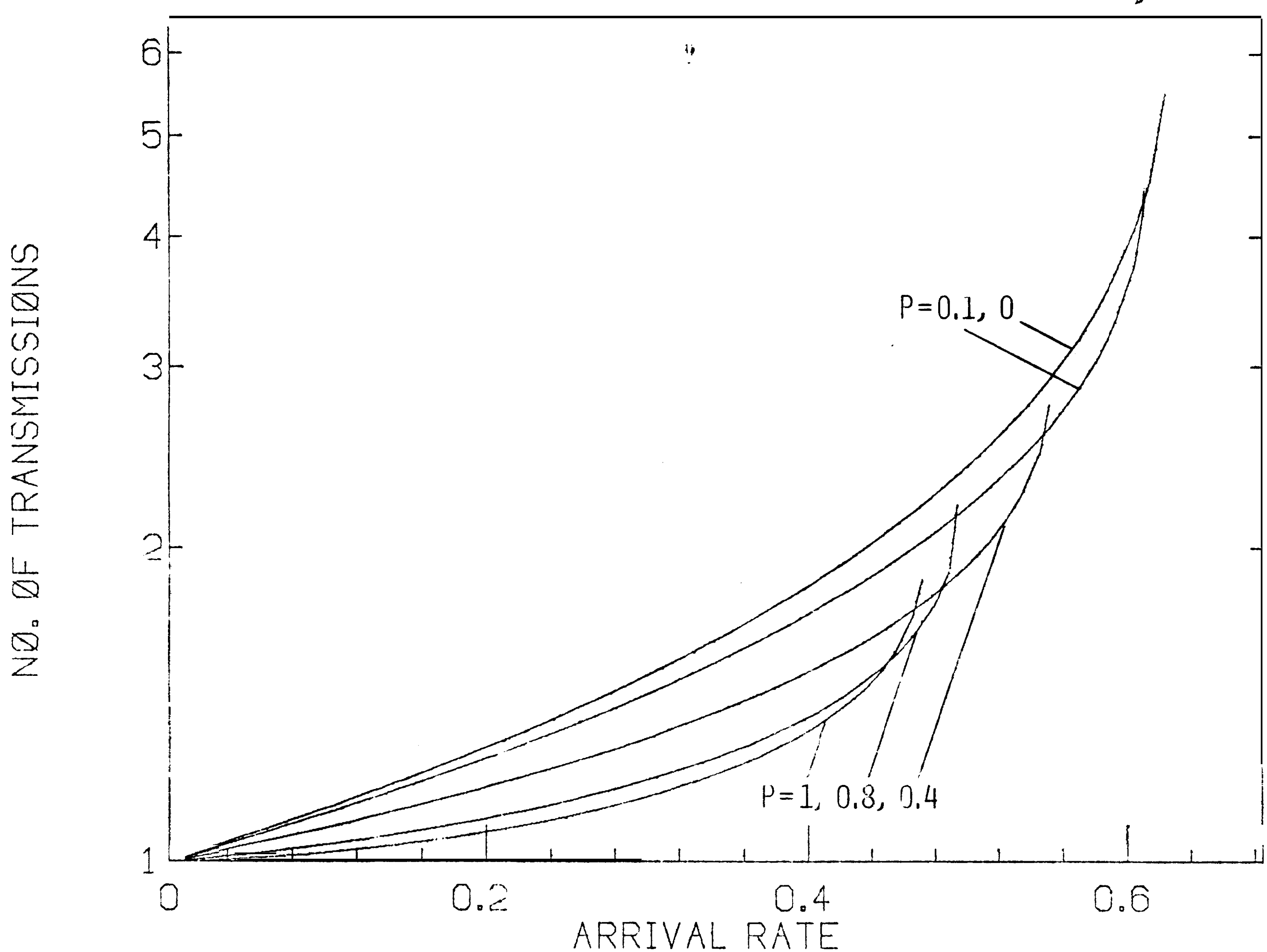


FIG. 3.4.A ØPT. SLØT. Mp PERS. CSMA (A=0.01)

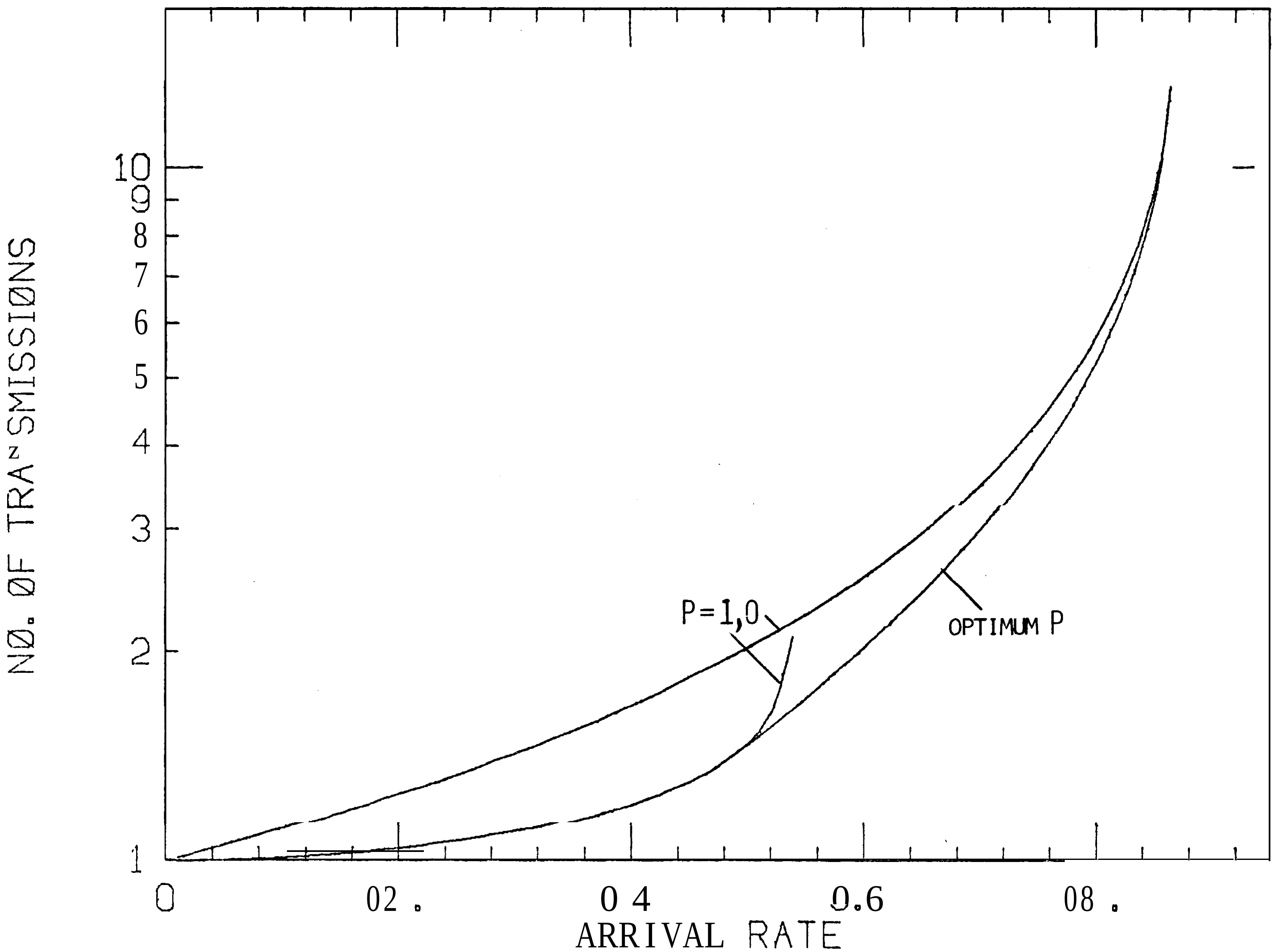
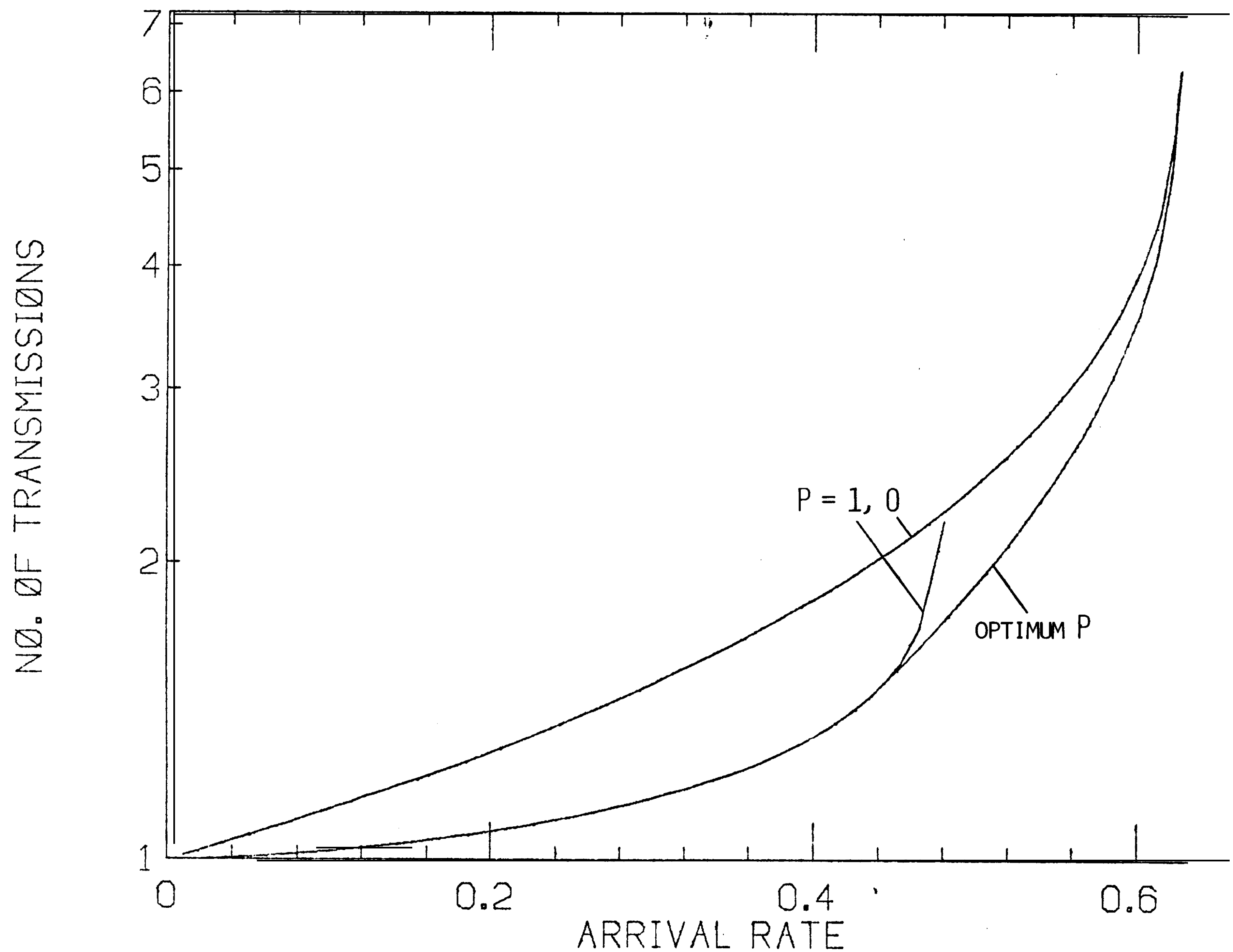


FIG. 3.4.B ØPT. SLØT. Mp PERS. CSMA (A=0.1)



the number of retransmissions can be reduced by properly choosing p. By varying p dynamically according to the channel traffic to minimize the delay or number of retransmissions, we obtain the optimum slotted M_p-persistent CSMA scheme.

In Fig. 3.4a and 3.4b, we plot the number of transmissions per successful packet of the optimum slotted M_p-persistent CSMA scheme with that of slotted MO-persistent CSMA and slotted M_1-persistent CSMA, for $a=0.01$ and 0.1, respectively. As we can see the optimum slotted M_p-persistent CSMA scheme achieves the same channel capacity as M_0-persistent CSMA but with smaller transmission delay or number of retransmissions: per successful packet. Although the reduction in delay may be less than that by optimum p persistent CSMA, optimum slotted M_p-persistent CSMA does achieve largerchannel capacity hence better stability.

3.C Performance Analysis of the Unslotted M_p-persistent CSMA Scheme Under Packet Switching

In this subsection, we analyze the performance of the unslotted M_p-persistent CSMA scheme under packet switching. The following theorem expresses the throughput in terms of offered traffic for this scheme.

<u>Theorem 3.3</u>:

For a given offered packet traffic G and a given value of the parameter p, the throughput equation for unslotted M_p-persistent CSMA under packet -switching is

$$S = \begin{cases} \dfrac{Gpe^{-(2a+p)G}(1-GP(1-p)+(G(a+1)(1-p)-1)e^{Ga(1-p)})+G(1-p)e^{-G(p+a)}(e^{-aGp}-pe^{-aG})}{(1-p)^2((1+2a)G-(1-e^{-aG}))+(1-p)e^{-Gp}(e^{-aGp}-pe^{-aG})} & \text{if } p \neq 1 \\[2ex] \dfrac{G[1+G+aG(1+G+aG/2)]e^{-G(1+2a)}}{G(1+2a)-(1-e^{-aG})+(1+aG)e^{-G(1+a)}} & \text{if } p = 1 \end{cases} \qquad (3.16)$$

Proof:

Consider Fig. 3.5 and let t be the arrival time of a packet. The channel is assumed to be idle at time t. The idle and busy periods are defined as before. Since there is a propagation delay of length a, any other packets arriving during [t,t+a] will sense the channel idle and start&o transmit. If no other transmissions occur during [t,t+a], the first packet will be successful. Let t+Y be the time of occurence of the last packet arriving between t and t+a. Any packet arrives during (t+a,t+Y+1+a] will sense the channel busy and persist on waiting with probability p. We now proceed to evaluate the probability distribution of the number of packets pending for transmission at the end of each transmission period, which is equal to that of the number of arrivals which persist on waiting in 1+Y units of time. We will show the above distribution is a Possion distribution. Let $q_k(y)$ be the probability of exactly k packets pending at the end of a transmission period whose duration is $1+y$. Since the arrival process is a Poisson process with arrival rate G,

$$q_k(y) = \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \frac{(G(1+y))^n e^{-G(1+y)}}{n!}$$

Changing the index of summation

$$q_k(y) = \frac{(pG(1+y))^k e^{-G(1+y)}}{k!} \sum_{n=0}^{\infty} \frac{(G(1+y))^n (1-p)^n}{n!}$$

Using the fact that $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$, we get

$$q_k(y) = \frac{(pG(1+y))^k e^{-pG(1+y)}}{k!}$$

This is exactly the Poisson distribution with parameter $pG(1+y)$. That is to say, the arrival process of the pending packets is a Poisson process with rate pG.

The distribution function for Y is

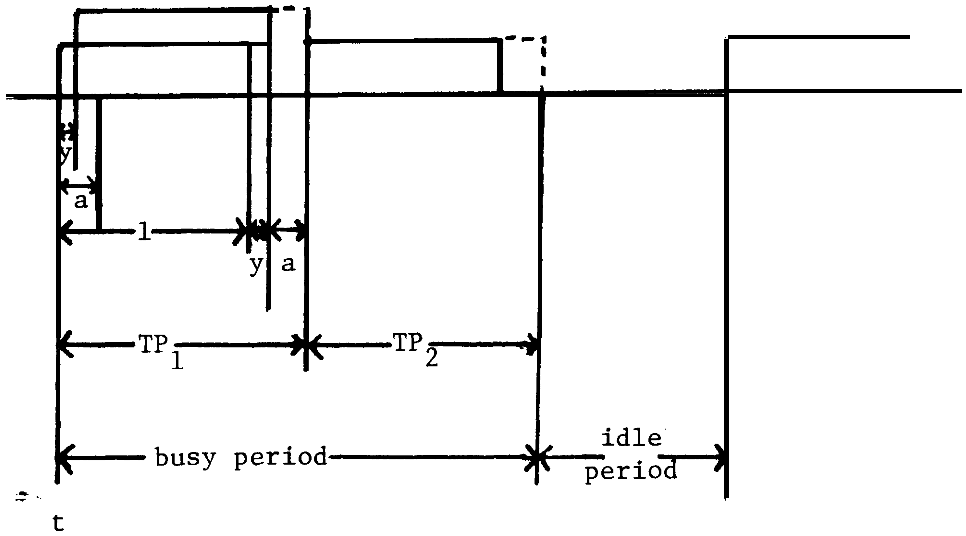


Figure 3.5: UNSLOTTED M_p-PERSISTENT CSMA, BUSY AND IDLE PERIODS

$$F_Y(y) \triangleq Pr\{Y \le y$$

$$= Pr\{\text{no arrival occurs in an interval of length } a-y)$$

$$= e^{-G(a-y)} \qquad (y \le a) \qquad (3.17)$$

Let $\overline{Y}$ be the average of Y, then

$$\overline{Y} = \int_0^a y \, dF_Y(y)$$

$$= a - \frac{1}{G}(1-e^{-aG}) \qquad (3.18)$$

The Laplace transform of the probability density function of Y is given by

$$F_Y^*(s) \triangleq \int_0^\infty e^{-sy} \, dF_Y(y)$$

$$= e^{-aG} + \frac{G(e^{-as}-e^{-aG})}{G-s} \qquad (3.19)$$

Since the Laplace transform of the density function of the sum of two independent random variables is equal to the product of the Laplace transforms of the density functions of the individual random variables, we obtain the Laplace transform of $Z \triangleq 1+Y$ as

$$F_Z^*(s) = e^{-s}\left(e^{-aG} + \frac{G(e^{-as}-e^{-aG})}{G-s}\right) \qquad (3.20)$$

Let q_m be the probability that m packets accumulated at the end of a transmission period and Q(Z) be the generatine function of q_m defined by

$$Q(Z) \triangleq \sum_{m=0}^{\infty} q_m Z^m$$

From [1], we know if the arrival process is a Poisson process with rate Gp, the generating function, Q(Z), of the number of arrivals and the Laplace transform, F*(Z), of that observed period have the following relationship:

$$Q(Z) = F_Z^*(Gp(1-Z)); \qquad (3.21)$$

From (3.20) and (3.21), we get

$$Q(z) = e^{Gp(z-1)}e^{-aG}\left(1 + \frac{e^{aG(1-p+pz)}-1}{1\ p+pz}\right)$$

The probability of having no packet accumulated at the end of a transmission period is given by

if $p \neq 1$

$$q_0 = Q(z)|_{z=0}$$

$$= \frac{e^{-Gp}(e^{-aGp}-pe^{-aG})}{1-P} \qquad (3.22)$$

if $p = 1$, by L'Hospital rule

$$q_0 = \lim_{z \to 0} Q(z)$$

$$= e^{-G(1+a)}(1+aG) \qquad (3.23)$$

Let $\overline{B}$ and $\overline{I}$ denote the expected duration of the busy period and idle period, respectively. We now consider the probability of success of an arbitrary packet. If the packet arrives during an idle period it will be successful if no other packets arrive during the next a units of time. Hence its probability of success is e^{-aG}. If the packet arrives during the middle of a busy period excluding the first a units of time of each transmission period, it will be successfully transmitted if it is the only persistently waiting packet during the transmission period and no packet arrives during its first a units of transmission . Let B' denote the time during a cycle that the channel is in its busy period excluding the first a units of time of each transmission period. From [1], we know that conditioning on the fact that a packet arrives in B' this packet is more likely to arrive in a longer transmission period than a shorter one. Let $\hat{z}$ denote the length of the transmission period in which the assumed arrival occured, and $\hat{q}_0$ be the probability that no pending arrival occurs in $\hat{z}$. Then the probability of success of the packet is $p\hat{q}_0e^{-aG}$. If the packet arrives during the first a units of time of a transmission period, it definitely will not succeed. Hence,

$$p_s = \Pr\{\text{success}\}$$

$$= \frac{\overline{I}}{\overline{B} + \overline{I}} e^{-aG} + \frac{\overline{B'}}{\overline{B} + \overline{I}} p\hat{q}_0 e^{-aG} \tag{3.24}$$

It is clear that the average idle period is given by

$$\overline{I} = 1/G \tag{3.25}$$

since the arrival process is Poisson.

To find the average busy period, recall the number of transmission periods in a busy period is geometrically distributed with mean $1/q_0$. Hence, the average busy period is given by

$$\overline{B} = \frac{1 + a + \overline{Y}}{q_0} \tag{3.26}$$

Similarly

$$\overline{B'} = (1 + \overline{Y})/q_0 \tag{3.27}$$

Finally, we proceed to evaluate $\hat{q}_0$. From (3.20), the density function of Z is given by

$$f_Z(x) = e^{-aG} \delta(x-1) + G e^{-aG} e^{G(x-1)} \quad 1 \le x \le 1 + a \tag{3.28}$$

where $\delta(x)$ is the impulse function.

The probability density function of $\hat{Z}$ is given by [1]

$$f_{\hat{Z}}(x) = \frac{x f_Z(x)}{\overline{Z}}$$

From (3.28),

$$f_{\hat{Z}}(x) = \frac{e^{-aG}}{1 + \overline{Y}} \delta(x-1) + \frac{Gxe^{-aG} e^{G(x-1)}}{1 + \overline{Y}} \quad 1 \le x \le 1 + a \tag{3.29}$$

and the probibility of no pending arrivals during the interval $\hat{Z}$ is

$$\hat{q}_0 = \int_1^{1+a} e^{-Gpx} f_{\hat{Z}}(x)\, dx$$

Substituting (3.29) into the above equation, after simplification, we obtain

$$\hat{q}_0 = \begin{cases} \dfrac{e^{-G(a+p)}}{(1+\overline{Y})G(1-p)^2} \left(1-Gp(1-p) + (G(a+1)(1-p) -1)\, e^{Ga(1-p)}\right) & \text{if } p \neq 1 \\ \dfrac{e^{-G(1+a)}}{1 + \overline{Y}} \left(1 + aG(1 + a/2)\right) & \text{if } p = 1 \end{cases} \qquad (3.30)$$

Combining (3.25), (3.26), (3.27) and (3.24), we get

$$p_s = \frac{(1 + \overline{Y})\, p\, \hat{q}_0\, e^{-aG}/q_0 + e^{-aG}/G}{(1 + a + \overline{Y})/q_0 + 1/G} \qquad (3.31)$$

Finally, substituting the equations (3.18),(3.22) or (3.23) and (3.30) into (3.31) and recalling that $S = G\, p_s$, we obtain (3.16).

The following theorem gives the delay equation of unslotted M_p-persistent CSMA scheme under packet switching.

Theorem 3.4:

For a given offered traffic G and a given value of the parameter p, the delay equation for the unslotted M_p-persistent CSMA scheme under packet switching is given by

$$D = \left(\frac{G(p_I + p_w)}{S} - 1\right)(R + \overline{d}) + \frac{G(1-p_I - p_w)}{S}\,\delta + \overline{d} + 1 + a \qquad (3.32)$$

where

$$\overline{d} = \frac{p(G^2(1 + a^2) + 2\,(G - 1)(aG - (1 - e^{-aG})))}{2G(q_0 + aG + Gp\,(1 + a) - p(1 - e^{-aG}))} \qquad (3.33)$$

$$p_I = \frac{q_0 + aG}{q_0 + G(1 + 2a) - (1 - e^{-aG})} \qquad (3.34)$$

$$p_w = \frac{p(G(1 + a) - (1 - e^{-aG}))}{q_0 + G(1 + 2a) - (1 - e^{-aG})} \qquad (3.35)$$

and q_0 is given in (3.22) and (3.23)

Proof:

Consider the time axis in Fig. 3.5. An arrival packet may either be transmitted after a pretransmission delay or be scheduled for retransmission if the channel is busy, and will be transmitted immediately without delay if the channel is idle. The probability that an arrival packet will detect the channel idle is

$$p_I = \frac{\overline{I} + a/q_0}{\overline{I} + \overline{B}}$$

Combining (3.25), (3.26), (3.18) and the above equation together, we obtain (3.34). The probability that an arrival packet will detect the channel busy and persist on waiting is

$$p_w = \frac{p\,\overline{B}'}{\overline{I} + \overline{B}}$$

Combining (3.25), (3.26), (3.27), (3.18) and the above equation together, we obtain (3.35). Since the length of a transmission period is a random variable, the mean pretransmission delay in this case should be equal to the residue life of a transmission period excluding the first a units of time and is given by $\overline{Z^2}/2\overline{Z}$ under the Poisson assumption where $Z \stackrel{\Delta}{=} 1 + Y$. From the distribution function (3.20) of Z, we get

$$\overline{Z^2} = 1 + a^2 + 2\left(1 - \frac{1}{G}\right)\overline{Y}$$

Hence the overall mean pretransmission delay is,

$$\overline{d} = \frac{p_w}{p_I + p_w} \cdot \frac{\overline{Z^2}}{2\,\overline{Z}}$$

$$= \frac{p_w}{p_I + p_w} \frac{1 + a^2 + 2\left(1 - \frac{1}{G}\right)\overline{Y}}{2\,(1 + \overline{Y})} \qquad (3.36)$$

Combining (3.18), (3.34), (3.35) and (3.36) together, we obtain (3.33)

Since the delay before next channel sensing is

$$R^* = \begin{cases} R + \bar{d} & \text{if the packet is transmitted unsuccessfully} \\ \delta & \text{if the packet is rescheduled upon arrival} \end{cases}$$

the retransmission delay is given by (3.32). Again, if we can assume when a packet is scheduled for retransmission upon arrival, it behaves as if it went through a virtual transmission. The simplified delay equation is given by the following corollary.

<u>Corollary 3.4</u>:

Under the above assumption, the delay equation for unslotted M_p-persistent CSMA under packet switching is given by

$$D = (\frac{G}{S} - 1)\ (R + \bar{d}') + \bar{d}' + 1 + a \tag{3.37}$$

where

$$\bar{d}' = \frac{p(G^2(1 + a^2) + 2(G-1)(aG - (1 - e^{-aG})))}{2G(q_0 + G(1 + 2a) - (1 - e^{-aG}))} \tag{3.38}$$

In Fig. 3.6, 3.7 and 3.8, we plot, versus G, the channel throughput, S, of the unslotted M_p-persistent CSMA scheme for p = 0, 0.1, 0.4, 0.8 and 1 when a = 0.01, 0.05 and 0.1, respectively. In Fig. 3.9, we plot versus G the channel throughput, S, of the unslotted M_p-persistent CSMA scheme for p = 0, 0.001, 0.005, 0.01 when a = 0.01. From the above figures, we observe that we can either maximize the channel capacity or minimize the number of transmissions per successful packet under a specific traffic level by appropriately choosing p. Furthermore, setting p equal to zero does not achieve maximum channel capacity as in the slotted case. In fact, if we want to select a fixed p algorithm to operate for simplicity, we seem to be able to choose a p which can lead to not only larger channel capacity but also smaller number of transmissions per successful packet than M_0-persistent CSMA. For example, when a = 0.05 or 0.1, setting p = 0.1 will improve the performance over M_0-persistent CSMA.

In Fig. 3.10a and 3.10b, we plot the number of transmissions per successful packet of the optimum unslotted M_p-persistent CSMA with that of unslotted M_0-persistent CSMA and unslotted M_1-persistent CSMA, for a = 0.01 and 0.1, respectively. As we can see the optimum unslotted M_p-persistent CSMA shceme achieves better performance than the two currently available CSMA schemes in both throughput and delay.

FIG. 3.6 THRØUGHPUT IN UNSLØT. VERSIØN (A=0.01)

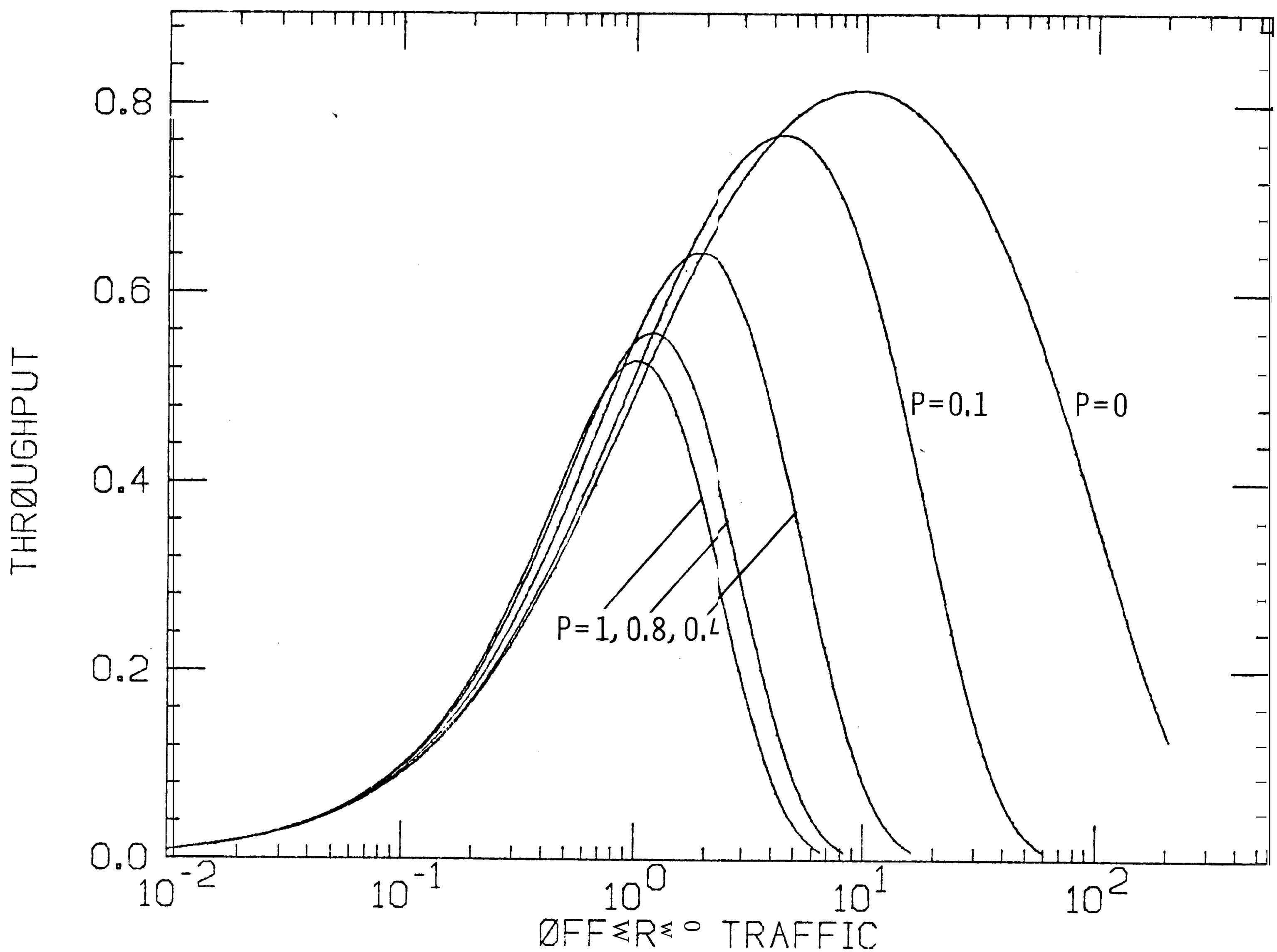


FIG. 3.7 THRØUGHPUT IN UNSLØT. VERSIØN (A=0.05)

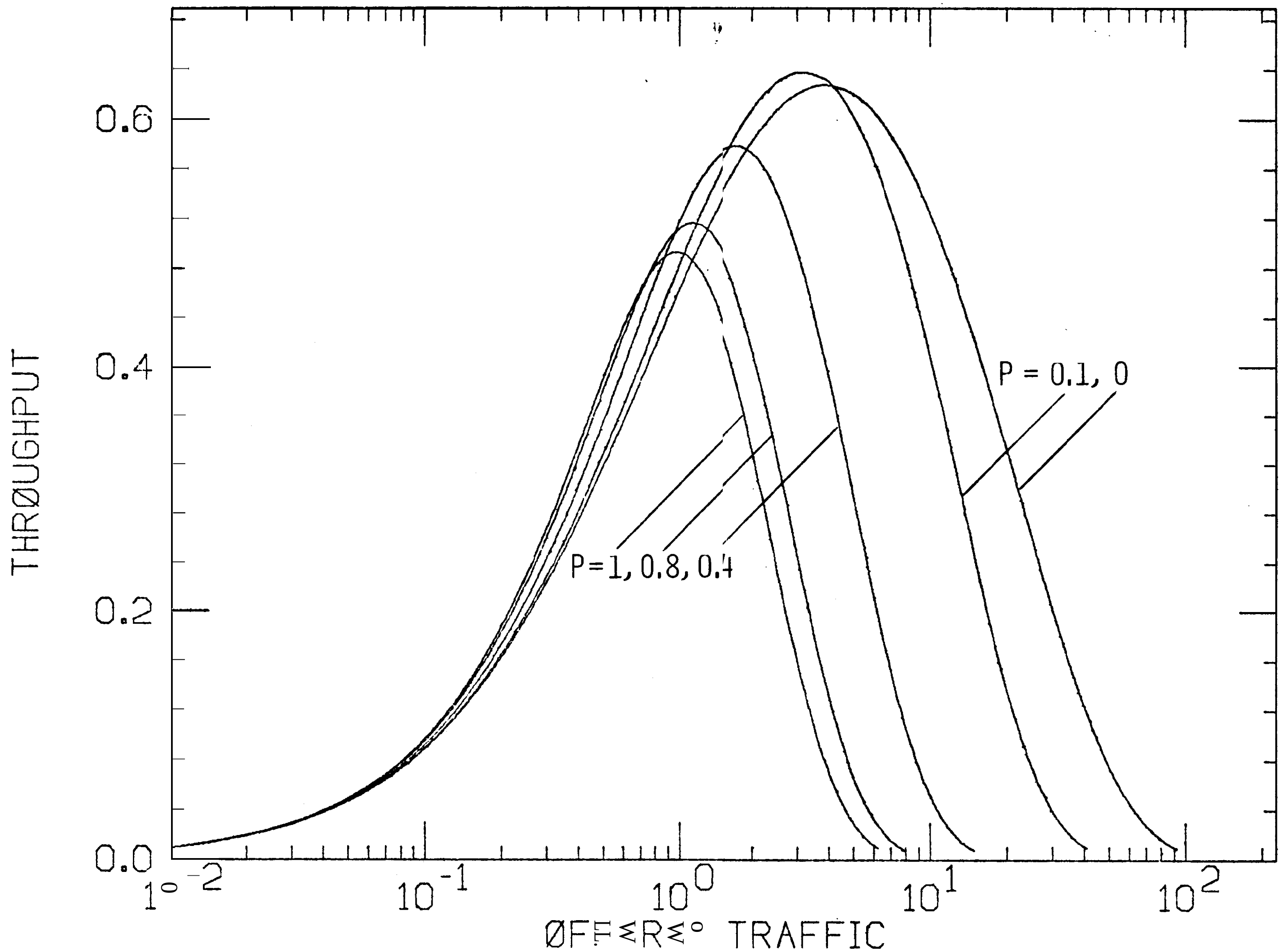


FIG. 3.8 THRØUGHPUT IN UNSLØT. VERSIØN (A=0.1)

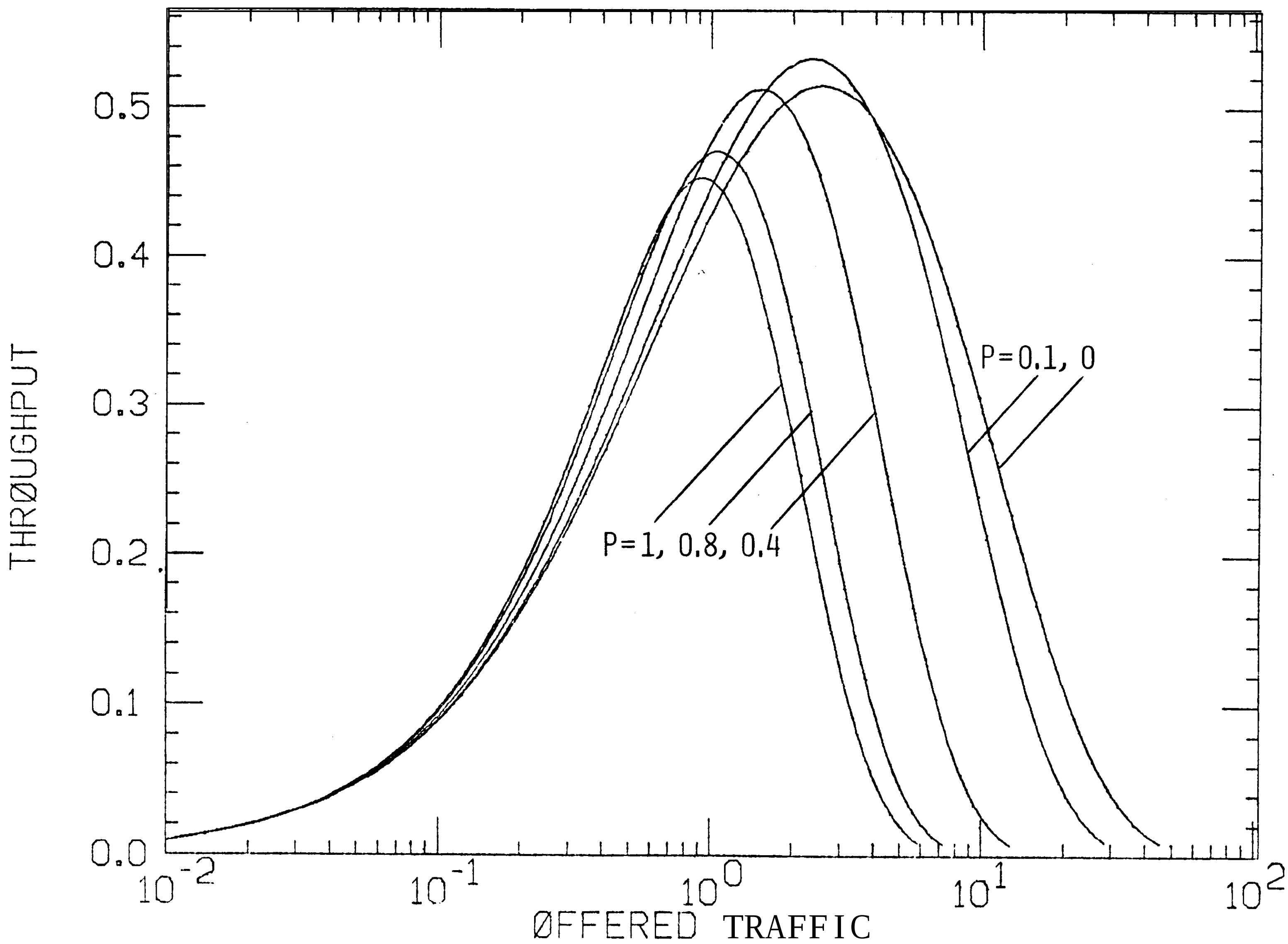


FIG. 3.9 THRØUGHPUT IN UNSLØT. VERSIØN (A=0.01)

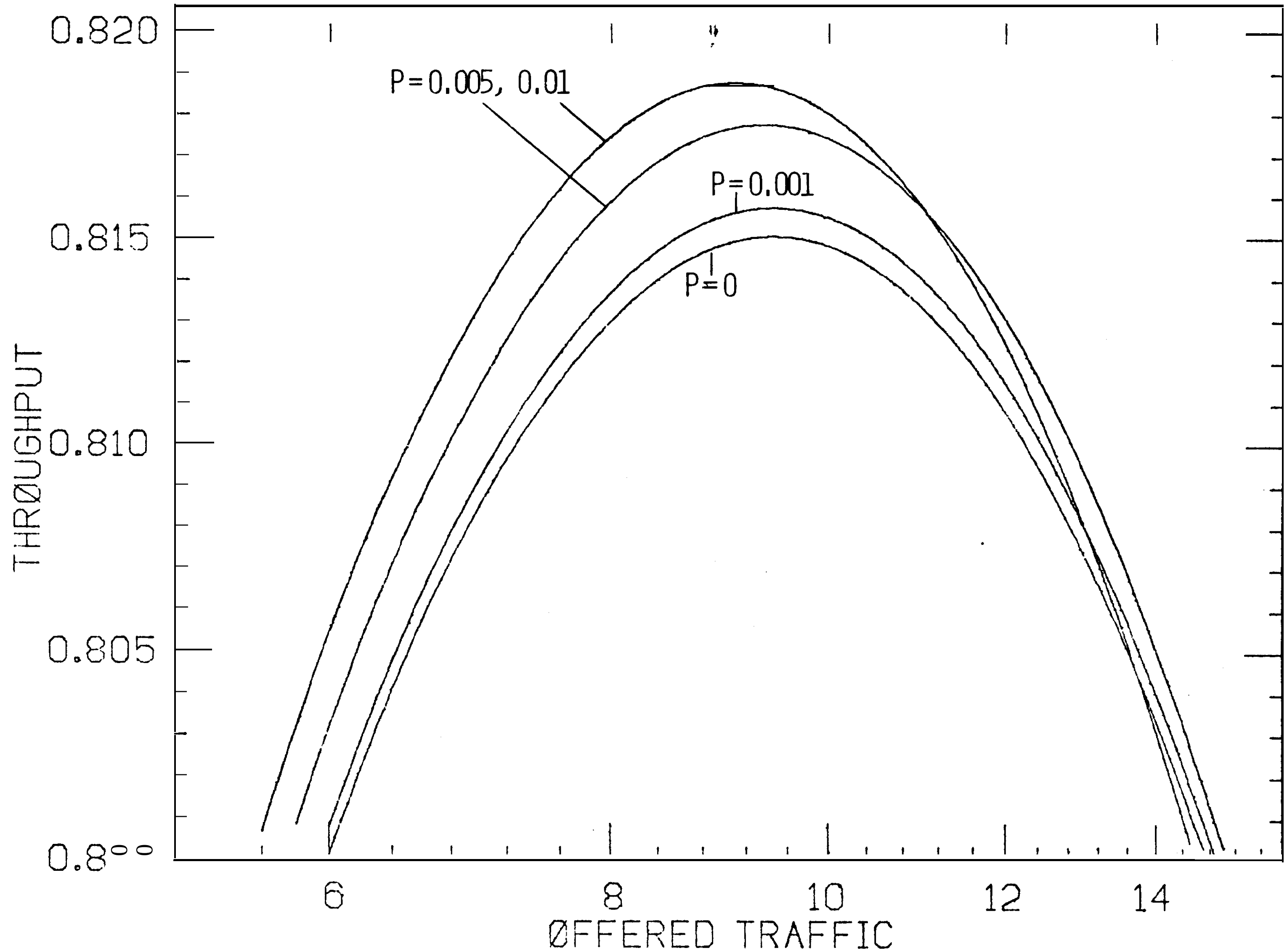


FIG. 3.10.A ØPT. UNSLØT. Mp PERS. CSMA (A=0.01)

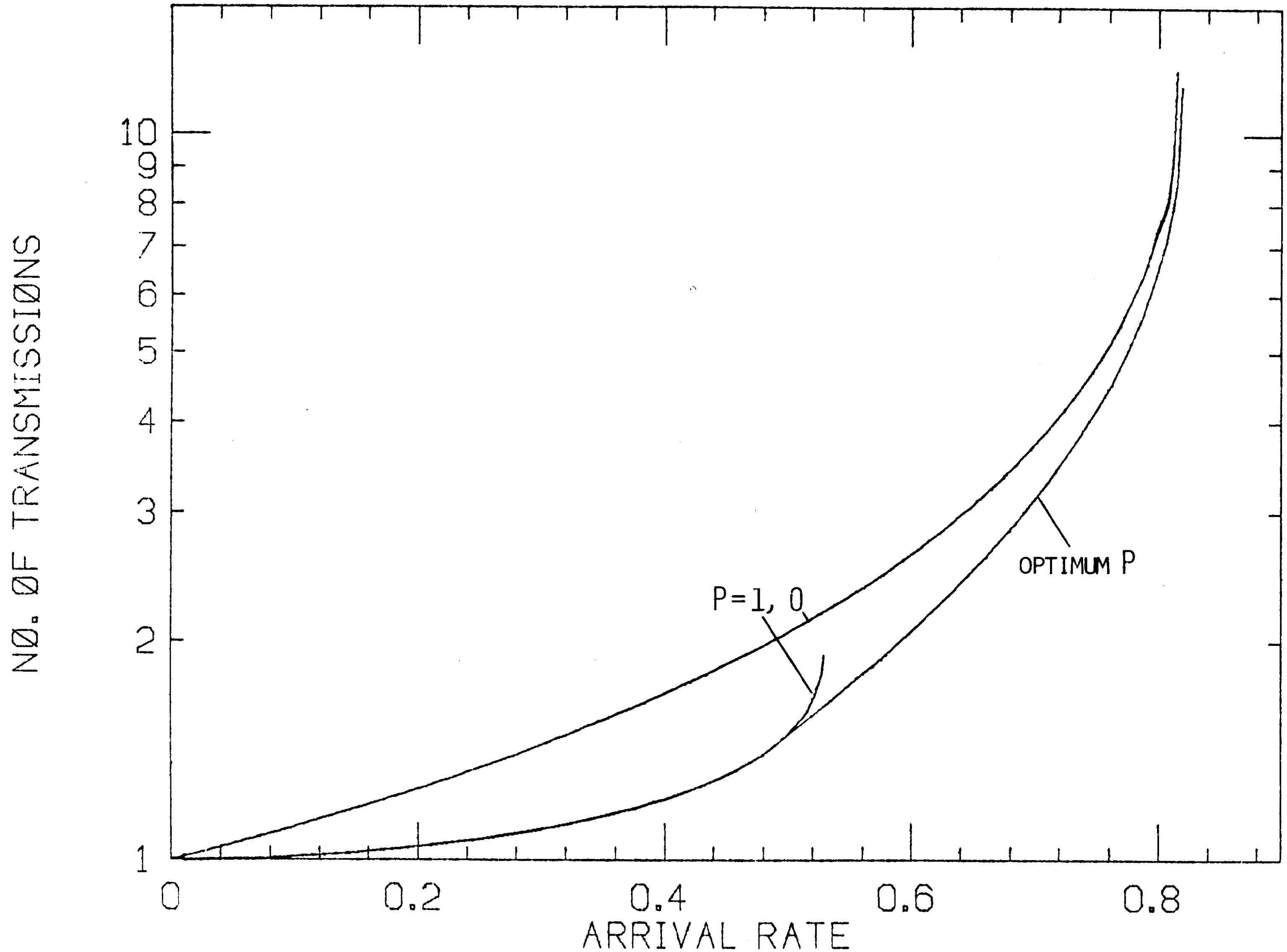
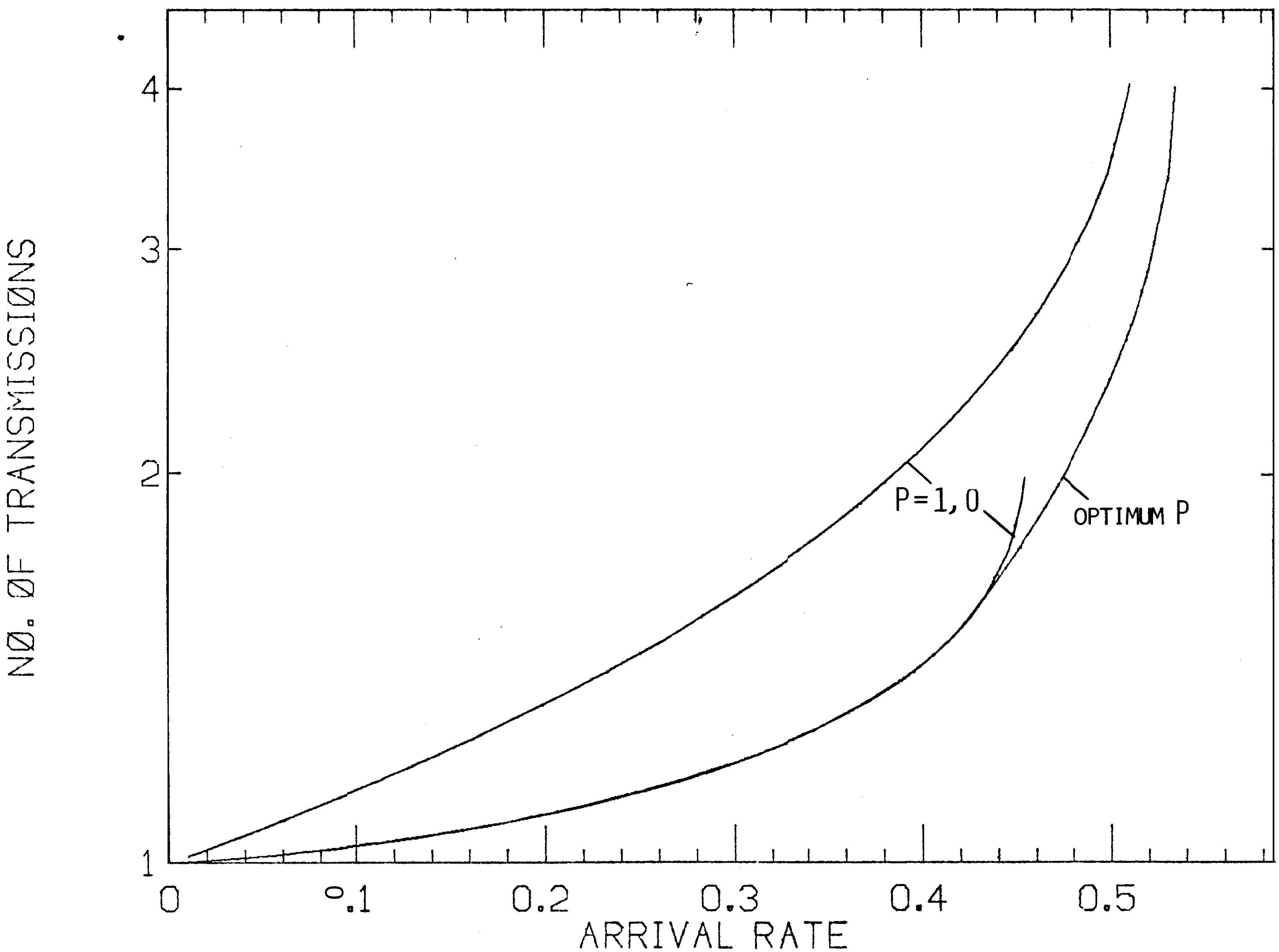


FIG. 3.10.B ØPT. UNSLØT. Mp PERS. CSMA (A=0.1)



4. PERFORMANCE ANALYSIS UNDER MESSAGE SWITCHING

4.A Basic Assumptions

In this section, we analyze the performance of various random access schemes under message switching and then compare the performance with that under packet switching. As we shall see, the performance under message switching may be superior to or inferior to that under packet switching depending upon the random access scheme being used and the distribution of message length for certain random access schemes. Random access schemes being examined include M_p-persistent CSMA, ALOHA and dynamic reservation schemes.

A message may consist of one or more packets. Let $\{F(n), n \geq 1\}$ be the distribution function of the number of packets contained in each message. Let S^* be the message arrival rate and assume that the arrival process is a Poisson process. Then the arrival rate of messages with length n will be $S^*(F(n)-F(n-1))$ and the arrival proceses of different message lengths form independent Poisson processes. Assume the total packet arrival rate is S' and the average number of packets contained in each message is L (>1), then we have

$$L = \sum_{n=1}^{\infty} n(F(n)-F(n-1))$$

and

$$S' = \sum_{n=1}^{\infty} nS^* (F(n)-F(n-1))$$

$$= LS^*$$

Let G^* be the mean offered message traffic which is the average traffic offered to the channel from our collection of users and consists of not only new messages but also previously collided messages.

To make the problem analytically tractable, we make the following assumptions.

Assumption 1: Each time a message is transmitted or retransmitted, it chooses its length independently from the distribution function F(n).

Assumption 2: The average retransmission delay is reasonably large, so the probability of successive collisions is small.

Assumption 3: The interevent times of the point process defined by both the start times of all the messages and the retransmission times are independent and exponentially distributed.

Evidently, the length of a message will not change upon retransmissions. Nevertheless, assumption 1 is just the message independent assumption [2] adopted in modeling a store and forward computer communication network via terrestrial links where message length is resampled independently at each node from a common distribution as the message hops through the network. As we shall see that in the CSMA scheme long messages are not discriminated. Hence, the message independent assumption is very reasonable. But in the ALOHA scheme, long messages are discriminated. So the message independent assumption will lead to more optimistic performance prediction and will be avoided if possible. The other assumptions have already been used in the analysis of packet switching. Comments on the validities of these assumptions can be found in [11] and [16], respectively. Some simulation results on the slotted M_0-persistent CSMA scheme under message switching with assumptions 1 and 3 released are also included in this section. The simulation results and analytic results are very close to each other as expected.

Furthermore, we will not assume any specific distribution for the number of packets contained in each message. That is to say {F(n)} can be any general discrete distribution function.

In the following discussion, unless specified as message throughput or message arrival rate, we will denote channel throughput in terms of equivalent packet throughput and arrival rate in terms of equivalent packet arrival rate. Similar remarks hold for channel capacity and offered traffic.

4.B The M_p-persistent CSMA Scheme

From the definition of the M_p-persistent CSMA scheme, it is clear that the probability of conflict is the probability that more than one message starts to transmit at the beginning of a transmission period for the slotted version or during the first a units of time in a transmission period for the unslotted version. Hence, the probability of conflict under message switching is independent of the length of a message. Furthermore, the offered message traffic of message with length n is equal to $G^*(F(n) - F(n-1))$. That is to say the percentage of messages with length n among the messages waiting for retransmissions is the same as that among the new arrivals. This property has another implication. It means that the average number of retransmissions per successful transmission is the same for all messages regardless of their length. Long messages are not discrimenated under the M_p-persistent CSMA scheme.

We now examine the case where p = 0. The slotted version will be considered first. As we shall see, the performance of the slotted M_0-persistent CSMA scheme under message switching is superior to that under packet switching. Not only the average number of retransmissions per successful message transmission is less than that per successful packet transmission under packetswitching, but also the lower bound of the channel capacity under message switching is larger than the channel capacity under packet switching. A similar remark holds for the unslotted version. The following theorem derives the lower bound and upper bound on the message throughput under message switching for the slotted M_p-persistent CSMA scheme. The upper bound can be achieved when every message contains exactly L packets.

Theorem 4.1:

For a given offered message traffic, G*, the bounds on message throughput under message swicching for the slotted M_0-persistent CSMA scheme is given by

$$\frac{G^* e^{-aG^*}}{1 + LG^*} \leq S^* \leq \frac{aG^* e^{-aG^*}}{a + L(1-e^{-aG^*})} \qquad (4.1)$$

Proof:

In this slotted version, if two messages conflict, their start times will coincide. Consider the time axis in Fig. 4.1 and let t be the start of a time slot. Assume packets arrive during the previous slot which is in an idle period or is the last slot of a busy period. A busy period is defined to be the time between t and the end of the transmission of the longest message starting at t. An idle period is defined to be the period of time between two successive busy periods and may be of zero duration.

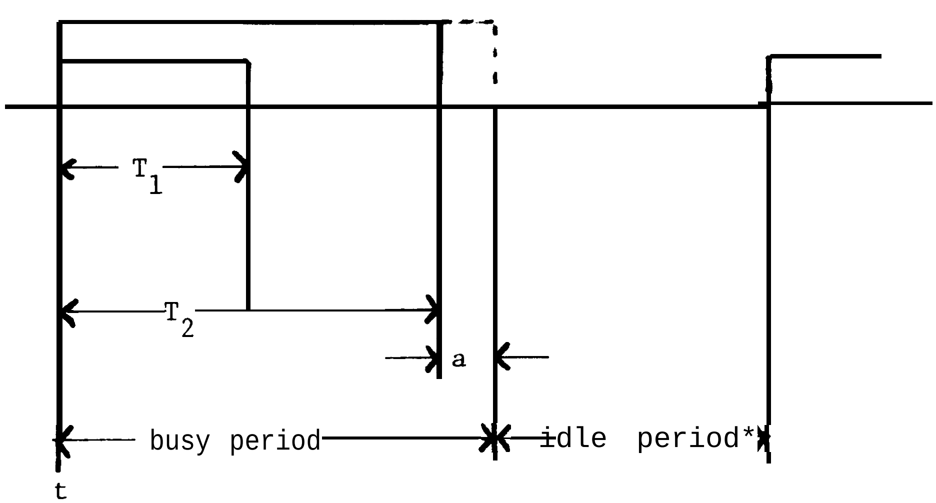


Fig. 4.1: Slotted M_0-persistent CSMA: Busy and Idle Periods

Let p_n denote the probability that exactly n messages start to transmit at the beginning of a busy period, then

$$p_n = \frac{(aG^*)^n e^{-aG^*}}{(1-e^{aG^*})n!} \tag{4.2}$$

Now let us consider the average length of a busy period conditioning on n messages start to transmit at the beginning of the busy period. Let B_n denote the length of this busy period, then

$$B_n = \max\,(T_1,\ T_2, \ldots .\ T_n) + a$$

where T_i is the length of the i-th message. The T_i's are assumed to be independently and identically distributed. Let

$$B_n = B'_n + a$$

i.e. $$B'_n = \max\,(T_1, T_2, \ldots .\ T_n)$$

then the distribution function of B'_n will be

$$F'_n(k) = (F(k))^n$$

where F(k) is the distribution function of T_i.

By assumption the expected value of T_i is L. Clearly, $\overline{B}'_n$, the expected value of B'_n, depends upon the distribution function of T_i and cannot be determined by L alone. Nevertheless, without specifying the distribution function of T_i explicitly, we can still obtain bounds on $\overline{B}'_n$ in terms of L. Since

$$T_1 \leq B'_n = \max\,(T_1,\ T_2, \ldots .\ T_n) \leq T_1 + T_2 + \ldots + T_n$$

we get

$$L \leq \overline{B}'_n \leq nL \tag{4.3}$$

We now proceed to evaluate bounds on average length of a busy period, by removing the condition on the number of messages transmitting at the beginning of a busy period. Let $\overline{B}$ be the average length of a busy period, then

$$\bar{B} = \sum_{n=1}^{\infty} p_n(\bar{B}'_n + a)$$

$$= \sum_{n=1}^{\infty} \bar{B}'_n p_n + a \qquad (4.4)$$

Combining (4.4), (4.2) and the right hand inequality of (4.3), we get

$$\bar{B} \le a + \sum_{n=1}^{\infty} nL \frac{(aG^*)^n e^{-aG^*}}{n!(1-e^{-aG^*})}$$

Using the fact that $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$, after simplification, we get

$$\bar{B} \le a + \frac{aG^*L}{1-e^{-aG^*}} \qquad (4.5)$$

Similarly, combining (4.4), (4.2) and the left hand inequality of (4.3), we get

$$\bar{B} \ge a + L \qquad (4.6)$$

Since the traffic process in each slot is an independent one, the average idle period, $\bar{I}$, is given by

$$\bar{I} = a \sum_{n=0}^{\infty} n(e^{-aG^*})^n (1-e^{-aG^*})$$

$$= \frac{a\, e^{-aG^*}}{1-e^{-aG^*}} \qquad (4.7)$$

Let $\bar{U}^*$ be the probability during a busy period that the channel is used without conflicts, clearly

$$\bar{U}^* = \frac{aG^* e^{-aG^*}}{1-e^{-aG^*}} \qquad (4.8)$$

Using renewal theory arguments, the average channel utilization is given by

$$S^* = \frac{\bar{U}^*}{\bar{B} + \bar{I}} \qquad (4.9)$$

Finally, substituting (4.5) or (4.6), (4.7) and (4.8) into (4.9), we get (4.1).

Be examining the first and second derivatives of the lower bound on S^*, we find that the lower bound achieves its maximum value at

$$G^* = \frac{-a + \sqrt{a^2 + 4aL}}{2aL}$$

Clearly the maximum of the lower bound on throughput is a lower bound on maximum achievable throughput. In Table 4.1, we compare the lower bound on channel capacity under message switching with the channel capacity under packet switching for various value of a and L. The lower bound on the channel capacity under message switching is always larger than the channel capacity under packet switching for all cases shown. The longer the average message length is, the larger the lower bound on the channel capacity will be. We can even achieve a reasonable performance when a is large, if the average message length is quite large. The upper bound on the channel capacity under message switching is also shown on Table 4.1 for comparison. Although the lower bound and upper bound are independent of the exact distribution of the number of packets in each message and depend only upon the mean number of packets in each message, they turn out to be quite close to each other, expecially when a is small.

Now let us examine the asymptotic behavior of the message throughput S^*, when aG^* is small. The upper bound and lower bound are extremely close to each other in this case.

<u>Theorem 4.2</u>:

When $g^* = aG^*$ is small

$$\frac{aG^*(1-aG^*) + O(g^{*3})}{a + LaG^*} \leq S^* \leq \frac{aG^*(1-aG^*) + O(g^{*3})}{a+LaG^* - \frac{L}{2}a^2G^{*2} + O(g^{*3})} \qquad (4.10)$$

under message switching.

The proof is straightforward and is omitted.

a	packet switching	Message Switching											
		L=2		L=4		L=8		L=12		L=16		L=20	
		upper bound	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound	lower bound
0.001	0.956	0.967	0.957	0.976	0.969	0.983	0.978	0.986	0.982	0.988	0.984	0.989	0.986
0.005	0.903	0.927	0.906	0.948	0.932	0.963	0.952	0.970	0.960	0.974	0.965	0.976	0.969
0.01	0.865	0.898	0.870	0.927	0.906	0.948	0.432	0.957	0.944	0.963	0.952	0.967	0.957
0.03	0.776	0.829	0.789	0.877	0.844	0.911	0.886	0.927	0.905	0.937	0.918	0.943	0.926
0.05	0.719	0.784	0.738	0.843	0.805	0.887	0.856	0.907	0.881	0.919	0.896	0.927	0.906
0.07	0.676	0.750	0.700	0.817	0.774	0.867	0.833	0.890	0.861	0.905	0.878	0.914	0.890
0.1	0.624	0.719	0.655	0.794	0.738	0.851	0.805	0.877	0.837	0.892	0.856	0.903	0.870
0.3	0.439	0.560	0.494	0.666	0.600	0.753	0.691	0.794	0.738	0.819	0.767	0.837	0.789
0.5	0.315	0.472	0.413	0.590	0.530	0.691	0.625	0.740	0.678	0.772	0.713	0.794	0.738
0.7	0.270	0.411	0.359	0.534	0.470	0.644	0.577	0.700	0.634	0.735	0.672	0.760	0.700
0.9	0.236	0.366	0.320	0.490	0.430	0.606	0.539	0.666	0.599	0.705	0.639	0.732	0.669

Table 4.1: Comparison of Channel Capacities Under Packet Switc ing and Message Switching for the Slotted M_0-persistent CSMA Scheme

Corollarv 4.2:

When aG^* is small

$$G^* \leq \frac{(1 - S^*L) - \sqrt{(1-S^*L)^2 - 4aS^*}}{2a} \tag{4.11}$$

under message switching.

Proof:

From (4.10), by neglecting $O(g^{*3})$, we obtain

$$S^* \geq \frac{aG^*(1 - aG^*)}{a + LaG^*}$$

Multiplying both sides by $(a + LaG^*)$, after simplification, we get

$$aG^{*2} - (1 - S^*L)\,G^* + S^* \geq 0$$

which is equivalent to

$$G^* \geq \frac{(1 - S^*L) + \sqrt{(1-S^*L)^2 - 4aS^*}}{2a}$$

or

$$G^* \leq \frac{(1 - S^*L) - \sqrt{(1-S^*L)^2 - 4aS^*}}{2a}$$

From Fig. 4.2 it is apparent that

$$G^* \leq \frac{(1 - S^*L) - \sqrt{(1-S^*L)^2 - 4aS^*}}{2a}$$

is the correct solution

Furthermore, let G be the offered packet traffic under packet switching, and S be the corresponding throughput. By similar argument, we can get the asmpytotic behavior of throughput under packet switching when aG is small.

Theorem 4.3:

When aG is small

$$S = \frac{aG(1-aG) + O((aG)^3)}{aG + \frac{a^2G^2}{2} + O((aG)^3)}$$

under packet switching

Corollary 4.3:

When aG is small

$$G \approx \frac{(1-S) - \sqrt{(1-S)^2 - 4a(1-\frac{S}{2})S}}{2a\,(1-S/2)} \tag{4.12}$$

under packet switching.

Let us compare the average number of retransmissions per successful message under message switching with that per successful packet under packet switching. The following theorem proves that before the channel throughput under packet switching gets close to saturation, the average number of retransmissions per successful message is smaller than that per successful packet under packet switching for the slotted M_0-persistent CSMA scheme.

Theorem 4.4 :

For any arbitrary distribution of message length, before the channel throughput S under packet switching gets close to saturation, i.e., when aG and $\frac{a}{(1-S)^2}$ is small, if the mean number of packets contained in each message is larger than 2/(2-S), then the average number of retransmissions per successful message under message switching is less than that per successful packet under packet switching for slotted M_0-persistent CSMA.

Proof:

Using the binominal theorem

$$(1 - x)^{\frac{1}{2}} = \sum_k \binom{\frac{1}{2}}{k} (-x)^k$$
$$= 1 - \frac{1}{2} x - \frac{1}{8} x^2 + 0(x^3)$$

we get

$$\sqrt{(1-S)^2 - 4aS(1-S/2)} = (1-S) - \frac{2aS\,(1-\frac{S}{2})}{1-S} - \frac{2a^2S^2(1-\frac{S}{2})^2}{(1-S)^3} + 0(\frac{a^3}{(1-S)^5}) \tag{4.13}$$

$$\sqrt{(1-S)^2 - 4aS/L} = (1-S) - \frac{2aS}{L(1-S)} - \frac{2a^2S^2}{(1-S)^3L^2} + 0(\frac{a^3}{(1-S)^3}) \tag{4.14}$$

By assumption S = S*L, hence

$$\frac{G}{S} - \frac{G^*}{S^*} = \frac{G}{S} - \frac{G^*L}{S}$$

Combining (4.11), (4.12), (4.13) and (4.14) with the above equation, we get

$$\frac{G}{S} - \frac{G^*}{S^*} \geq \frac{aS}{(1-S)^3} (1 - \frac{S}{2} - \frac{1}{L})$$

Using the fact that $L > \frac{2}{2S-1}$ i.e. $\frac{1}{L} < 1 - \frac{S}{2}$, we get

$$\frac{G}{S} - \frac{G^*}{S^*} \geq 0$$

which is equivalent to

$$(\frac{G}{S} - 1) - (\frac{G^*}{S^*} - 1) > 0$$

If we assume when a message is scheduled for retransmission upon arrival it behaves as if it could transmit and learned about its rescheduling only α units of time after the end of its virtual transmission as before, then the transmission delays under message switching and packet switching are given by delay under message switching: $(\frac{G^*}{S^*} - 1)(L + R') + L + a + d$ (4.15)

delay under packet switching: $L(\frac{G}{S} - 1)(1 + R') + L + a + d + (L-1)(2a+\alpha+d)$ (4.16)

$$\geq (\frac{G}{S} - 1)(L + LR') + L + a + d$$

where $R' = \delta + 2a + \alpha + d > 1$

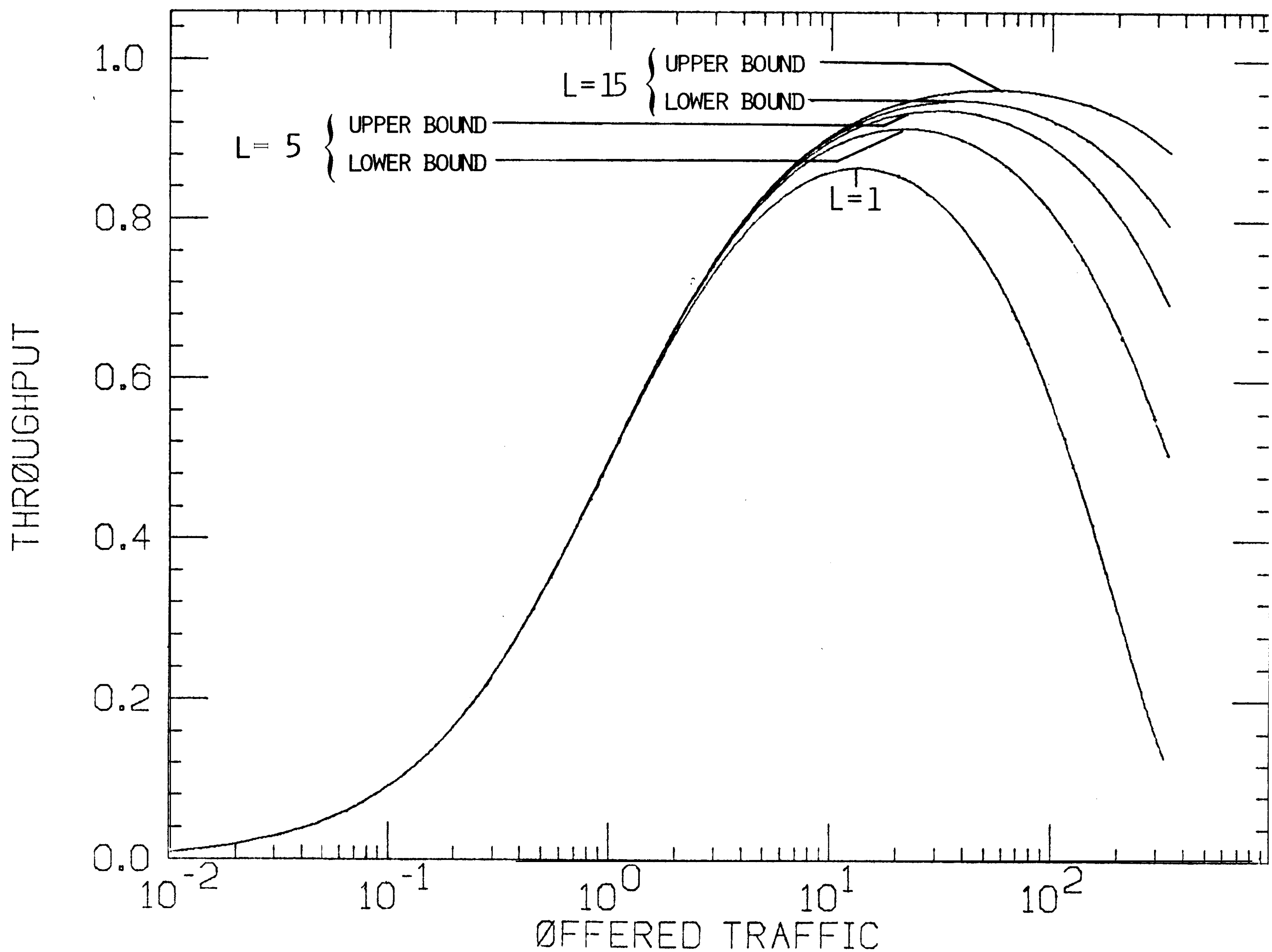
and d is the pretransmission delay, respectively. Even if $(\frac{G^*}{S^*} - 1)$ is larger than $(\frac{G}{S} - 1)$, the transmission delay under message switching can still be less than that under packet switching, if

$$(\frac{G^*}{S^*} - 1) \leq (\frac{L + LR'}{L + R'})(\frac{G}{S} - 1).$$

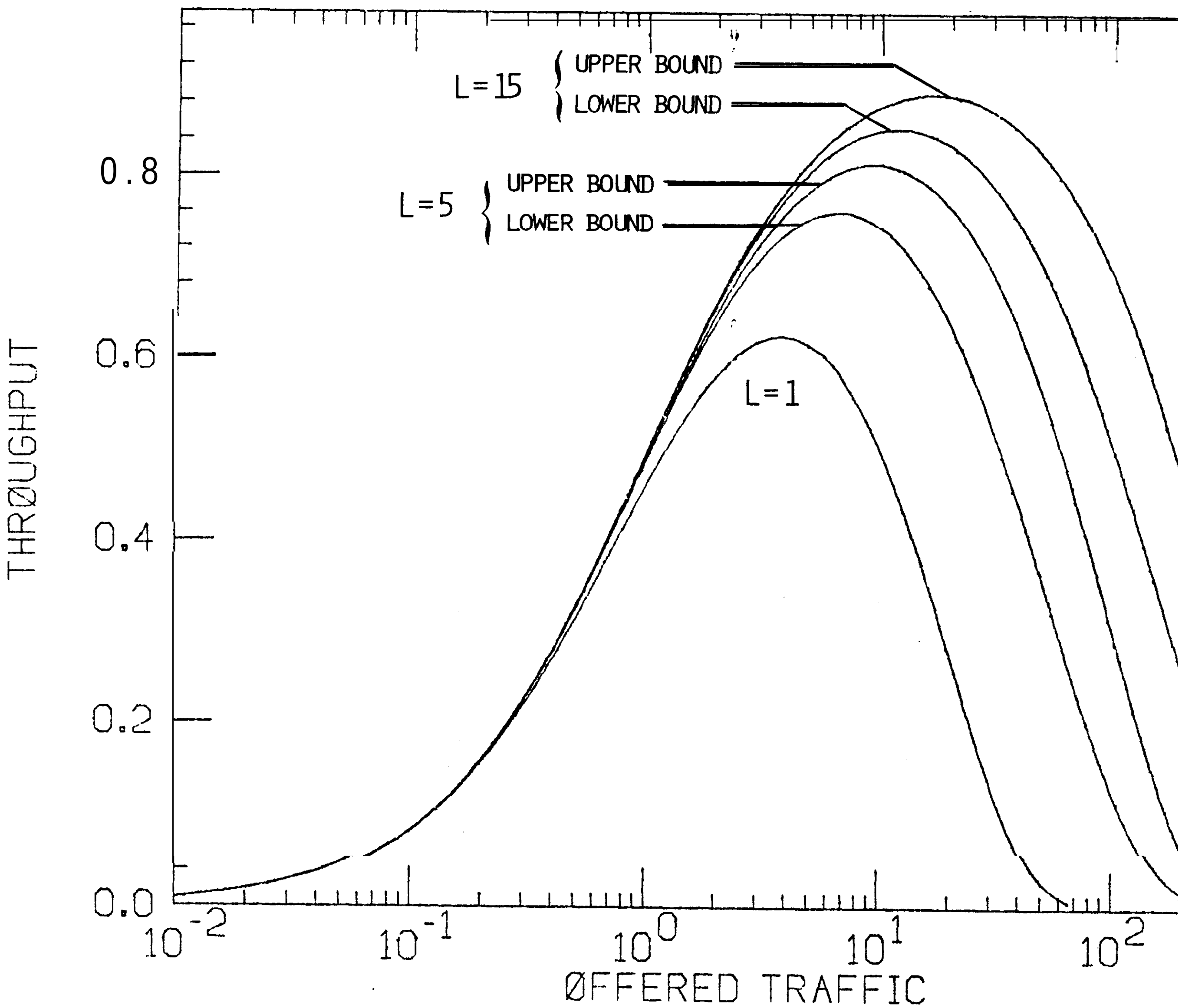
For example, when $L = R' = 5$, as long as $(\frac{G^*}{S^*} - 1) \leq 3(\frac{G}{S} - 1)$, message switching will lead to smaller transmission delay. The condition being proved, i.e. $(\frac{G^*}{S^*} - 1) < (\frac{G}{S} - 1)$ is a much stronger condition than is needed to assert that message switching leads to smaller transmission delay.

In Fig. 4.3a and 4.3b, we plot the upper bound and lower bound on throughput, LS^*, of the slotted M_0-persistent CSMA scheme under message switching versus offered traffic, LG^*, for L = 15 and 5 when a = 0.01 and 0.1, respectively. The throughput under packet switching, i.e., when L=1, is also shown. The closeness of the upper bound and lower bound on throughput under message switching is apparent. Hence, we can say that for the slotted M_0-persistent CSMA scheme, the performance under message switching is mainly determined by the mean number of packets contained in each message and is not sensitive to the exact distribution of the number of packets contained in each message. Furthermore, when the utilization of the channel is low, the value of $(\frac{G^*}{S^*} - 1)$, the number of retransmissions per successful message under message switching, and the value of $(\frac{G}{S} - 1)$, the number of retransmissions per successful packet under packet switching, are quite close to each other. Nevertheless, theorem 4.4 tells us that $(\frac{G^*}{S^*} - 1)$ is slightly smaller than

FIG. 4.3.A SLØTTED M_0 PERS. CSMA (A=0.01)



F G 4 3 B SLØTTED M_0 PERS CSMA (A O



$(\frac{G}{S}-1)$. As the utilization of the channel further increases, $(\frac{G^*}{S^*}-1)$ grows much slower than $(\frac{G}{S}-1)$, and the reduction in the number of retransmissions under message switching is apparent.

After analyzing the performance of the slotted MC-persistent CSMA scheme, we now proceed to study the performance of the unslotted M_0-persistent CSMA scheme under message switching. The following theorem gives the lower bound and upper bound on message throughput. Again, the upper bound can be achieved when every message contains exactly L packets.

Theorem 4.5:

For a given offered message traffic G*, the bounds on message throughput under message switching for the unslotted MC-persistent CSMA scheme is given by

$$\frac{G^* e^{-aG^*}}{2aG^*+LG^*(1+aG^*)+e^{-aG^*}} \leq S^* \leq \frac{G^* e^{-aG^*}}{2aG^*+LG^*+e^{-aG^*}} \tag{4.17}$$

Proof:

Consider the time axis in Fig. 4.4 and let t be the arrival time of a packet which starts a new busy period. The busy period and idle period are defined as in theorem 4.1. As in packet switching, messages arriving during [t,t+a] will sense the channel idle and proceed to transmit.

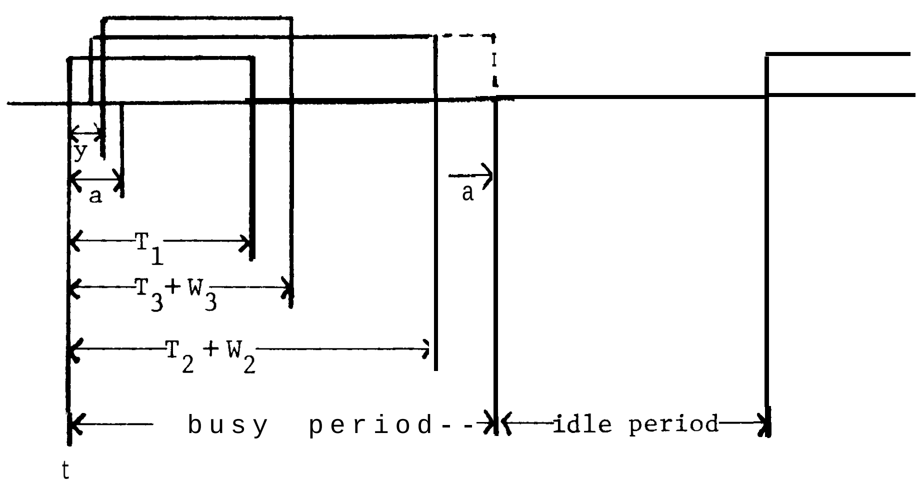


Fig. 4.4: Unslotted M_0-persistent CSMA Under Message Switching: Busy and idle periods

If no other transmissions occur during $[t,t+a]$, the first message will be successful. Let $t + Y$ be the time of occurence of the last message that arrives between t and $t + a$. The case $Y = 0$ corresponds to the situation where the first arrival is the only arrival during $[t,t+a]$ and occurs with probability e^{-aG^*}. Furthermore, when $Y = y > 0$, let $p_n(y)$ be the probability that exactly n messages arrive in $(t,t+y)$, not counting the first and last arrivals at t and $t+y$, respectively, and $B_n(y)$ denote the length of the corresponding busy period. Then

$$B_n(y) = \text{Max}\,(T_1,\ T_2 + W_2,\ \ldots,\ T_{n+1} + W_{n+1},\ T_{n+2} + y) + a, \text{ for } y > 0$$

where T_i denotes the length of the i-th message arriving during $[t,t+y]$ (including the messages arriving at t and $t+y$) and $t+W_i$ denotes the arrival time of the i-th message with $W_1 = 0$ and $W_{n+2} = y$.
All the T_i's are assumed to be independently and identically distributed. Clearly, the W_i's are dependent.
Obviously,

$$a + T_{n+2} + y \leq B_n(y) \leq a + \sum_{i=1}^{n+2} T_i + y, \text{ for } y > 0$$

Let $\overline{B}_n(y)$ be the average length of $B_n(y)$ for a given n and y and recall L is the average number of packets in each message. Then, taking expection of the above inequality, we get

$$a + L + y \leq \overline{B}_n(y) \leq a + (n+2)\,L + y$$

We now proceed to evaluate bounds on the average length of a busy period, by first removing the condition on the number of arrivals during $(t,t+y)$. Let $\overline{B}(y)$ be average length of the busy period conditioning on the last arrival before $t+a$ occurs at $Y = y$. Since the arrival process is a Poisson process,

$$p_n(y) = \frac{(y\,G^*)^n e^{-yG^*}}{n!} \quad , \text{ for } y > 0$$

Using the fact that

$$\overline{B}(y) = \sum_{n=0}^{\infty} p_n(y)\,\overline{B}_n(y) \quad , \text{ for } y > 0$$

we get

$$a + L + y \le \overline{B}(y) \le a + 2L + yG^{*}L + y \qquad , \text{ for } y > 0 \tag{4.18}$$

Furthermore,

$$\overline{B}(0) = a + L \tag{4.19}$$

Finally, we remove the condition on y. The distribution function of y is given by (3.17) with G replaced by G^{*}, i.e. $F_Y(y) = e^{-(a-y)G^{*}}$

Let $\overline{B}$ be the average length of a busy period

$$\overline{B} = \int_0^a \overline{B}(y)\, dF_Y(y)$$

$$= e^{-aG^{*}}\overline{B}(0) + \int_0^a \overline{B}(y)\; G^{*} e^{-(a-y)G^{*}} dy$$

Combining with (4.18) and (4.19), after simplification, we get

$$a + L + \overline{Y} \le \overline{B} \le a + L + \overline{Y} + aLG^{*} \tag{4.20}$$

where $\overline{Y}$ is the mean length of Y and is given by

$$\overline{Y} = a - \frac{1}{G^{*}}(1-e^{-aG^{*}}) \tag{4.21}$$

Since the arrival process is Poisson, the mean idle period is given by

$$\overline{I} = \frac{1}{G^{*}} \tag{4.22}$$

Furthermore, let $\overline{U}^{*}$ be the probability of success, then

$$\overline{U}^{*} = e^{-aG^{*}} \tag{4.23}$$

As before, the average message throughput is given by

$$S^{*} = \frac{\overline{U}^{*}}{\overline{B} + \overline{I}} \tag{4.24}$$

Finally, combining (4.20), (4.21), (4.22), (4.23) and (4.24) together, we obtain (4.17).

Now let us compare the performance of message switching with that of packet switching under unslotted M_0-persistent CSMA. In Table 4.2, we tabulate the channel capacity under packet switching and the upper bound and lower bound on channel capacity under message switching for various values of a and L. The lower bound on channel capacity under message switching is larger than the channel capacity under packet switching in all cases. The lower bound and upper bound are again quite close to each other. Furthermore, comparing the results in Table 4.1 and 4.2, we find that under message switching, the lower bound on the channel capacity of slotted M_0-persistent CSMA is never less than the upper bound on the channel capacity of unslotted CSMA for any a and L shown in the tables. That is to say, the channel capacity of the slotted version is never less than that of the unslotted version under message switching. This is also true for packet switching as can be seen from Table 4.1 and 4.2.

Finally, let us compare the average number of retransmissions under message switching with that under packet switching for the unslotted M_0-persistent CSMA scheme. Again, before the channel starts to get saturated under packet switching, we can prove that the average number of retransmissions per successful message under message switching is less than that per successful packet under packet switching. The following theorem establishes this fact.

<u>Theorem 4.6:</u>

For any arbitrary distribution of message length, before the channel throughput S under packet switching gets close to saturation, i.e. when aG is small, the mean number of retransmissions per successful message under message switching is less than that per successful packet under packet switching in unslotted M_0-persistent CSMA.

a	packet switching	Message Switching											
		L=2		L=4		L=8		L=12		L=16		L=20	
		upper bound	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound	lower bound
0.001	0.938	0.956	0.939	0.969	0.957	0.978	0.969	0.982	0.975	0.984	0.978	0.986	0.980
0.005	0.866	0.904	0.871	0.931	0.906	0.951	0.932	0.960	0.944	0.965	0.952	0.969	0.957
0.01	0.815	0.866	0.824	0.904	0.871	0.931	0.906	0.944	0.923	0.951	0.932	0.956	0.939
0.03	0.699	0.778	0.719	0.838	0.790	0.883	0.845	0.904	0.871	0.916	0.887	0.925	0.898
0.05	0.628	0.721	0.658	0.795	0.739	0.851	0.806	0.877	0.837	0.893	0.857	0.904	0.871
0.07	0.575	0.678	0.612	0.762	0.702	0.826	0.775	0.856	0.811	0.874	0.834	0.887	0.849
0.1	0.515	0.628	0.561	0.721	0.658	0.795	0.739	0.830	0.780	0.851	0.806	0.866	0.824
0.3	0.320	0.444	0.388	0.564	0.499	0.669	0.602	0.721	0.658	0.754	0.693	0.778	0.719
0.5	0.236	0.352	0.308	0.476	0.417	0.593	0.527	0.654	0.587	0.693	0.628	0.721	0.657
0.7	0.188	0.294	0.259	0.416	0.364	0.538	0.474	0.604	0.537	0.647	0.580	0.678	0.612
0.9	0.156	0.253	0.224	0.371	0.324	0.495	0.434	0.564	0.499	0.610	0.543	0.643	0.576

Table 4.2: Comparison of Channel Capacities Under Packet Switching and Message Switching for the Unslotted M_0-persistent CSMA Scheme

Proof:

Using the fact that

$$e^x = 1 + x + 0(x^2)$$

the throughput equation under packet switching (2.5) reduces to

$$S = \frac{G(1-aG+0(a^2))}{G(1+2a)+(1-aG+0(a^2))} \tag{4.25}$$

where S and G are the packet throughput and offered traffic, respectively.

After simplification and neglecting the $0(a^2)$ terms, we find G satisfies the following quadratic equations

$$aG^2 - (1-S-aS)\,G + S \approx 0$$

Taking the smaller root, we get

$$G \sim \frac{(1-S-Sa) - \sqrt{(1-S-Sa)^2-4aS}}{2a} \tag{4.26}$$

Similarly, the throughput equation for message switching (4.16) reduces to

$$S^* \geq \frac{G^*(1-aG^*+0(a^2))}{2aG^* + LG^*(1+aG^*)+(1-aG^*+0(a^2))} \tag{4.27}$$

where S^* and G^* are the message throughput and offered message traffic, respectively.

After simplification and neglecting the $0(a^2)$ terms, we find G^* statisfies the following inequality

$$(a + aLS^*)\,G^{*2} - (1 - LS^* - aS^*) + S^* \geq 0 \tag{4.28}$$

Using a similar argument as in lemma 3.2, we get

$$G^* \leq \frac{(1-LS^*- aS^*) - \sqrt{(1-LS^*-aS^*)^2 - 4S^*(a+aLS^*)}}{2(a + S^*aL)} \tag{4.29}$$

The root, y, of $AX^2 - BX + C = 0$ given by

$$y = \frac{B - \sqrt{B^2 - 4AC}}{2A}$$

can be expanded by the formula

$$(1 - Z)^{\frac{1}{2}} = 1 - \frac{1}{2} Z + O(Z^2) \qquad \text{when } Z \text{ is small,}$$

After simplification, we get

$$y = \frac{C}{B} + O(C^2)$$

Applying the above result to (4.26) and (4.29), we get

$$G \approx \frac{S}{1-aS-S} \tag{4.30}$$

and

$$G^* \leq \frac{S^*}{1-LS^* - aS^*} \tag{4.31}$$

By assumption $S = S^*L$, from (4.30) and (4.31), we have

$$\left(\frac{G}{S} - 1\right) - \left(\frac{G^*}{S^*} - 1\right) \geq \frac{aS(1-1/L)}{(1-(a+1)S)(1-(a+L)S^*)} \tag{4.32}$$

The average transmission period for a successful message is (a + L) under message switching. Hence, even if every message is successfully transmitted at the first transmission, we must have

$$(a + L)\, S^* \leq 1 \tag{4.33}$$

under steady state. Since conflict is inevitable, we must have

$$(a + L)\, S^* < 1$$

under steady state.

Similarly, for packet switching, we must have

$$(a + 1)\, S < 1 \tag{4.34}$$

under steady state.

Since L > 1

i.e. $1-1/L > 0$ (4.35)

Using (4.33), (4.34) and (4.35), we can conclude from (4.32) that

$$\left(\frac{G}{S} - 1\right) - \left(\frac{G^*}{S^*} - 1\right) \geq 0$$

In Fig. 4.5a and 4.5b, we plot the upper bound and lower bound on the throughput of the unslotted M_0-persistent CSMA scheme under message switching versus offered traffic for L = 15 and 5 when a = 0.01 and 0.1, respectively. The throughput of unslotted M_0-persistent CSMA under packet switching, i.e. when L = 1, is also shown. The upper bound and lower bound are very close to each other as in the slotted case. Hence, we can say that for unslotted MC-persistent CSMA, the performance under message switching is again mainly determined by the mean number of packets contained in each message. Furthermere, when the utilization of the channel is low, the number of retransmissions per successful message under message swithcing and that per successful packet under packet switching are quite close to each other. Nevertheless, theorem 4.6 proves that the number of retransmissions under message switching is slightly smaller. As the utilization of the channel further increases, the reduction in the number of retransmissions under message switching is apparent.

Now let us look at a specific example to gain some feeling on the performance improvement under message switching. Here we use slotted M_0-persistent CSMA as the random access scheme. Assume that there are Poisson arrivals of both single packet and multipacket messages at each station. The message arrival rate is S^* with a fraction h of single packets and the remainder of multipackets. All multipacket messages consist of exactly eight packets. Hence the average number of packets contained in

FIG. 4.5.A UNSLØTTED M_0 PERS. CSMA (A=0.01)

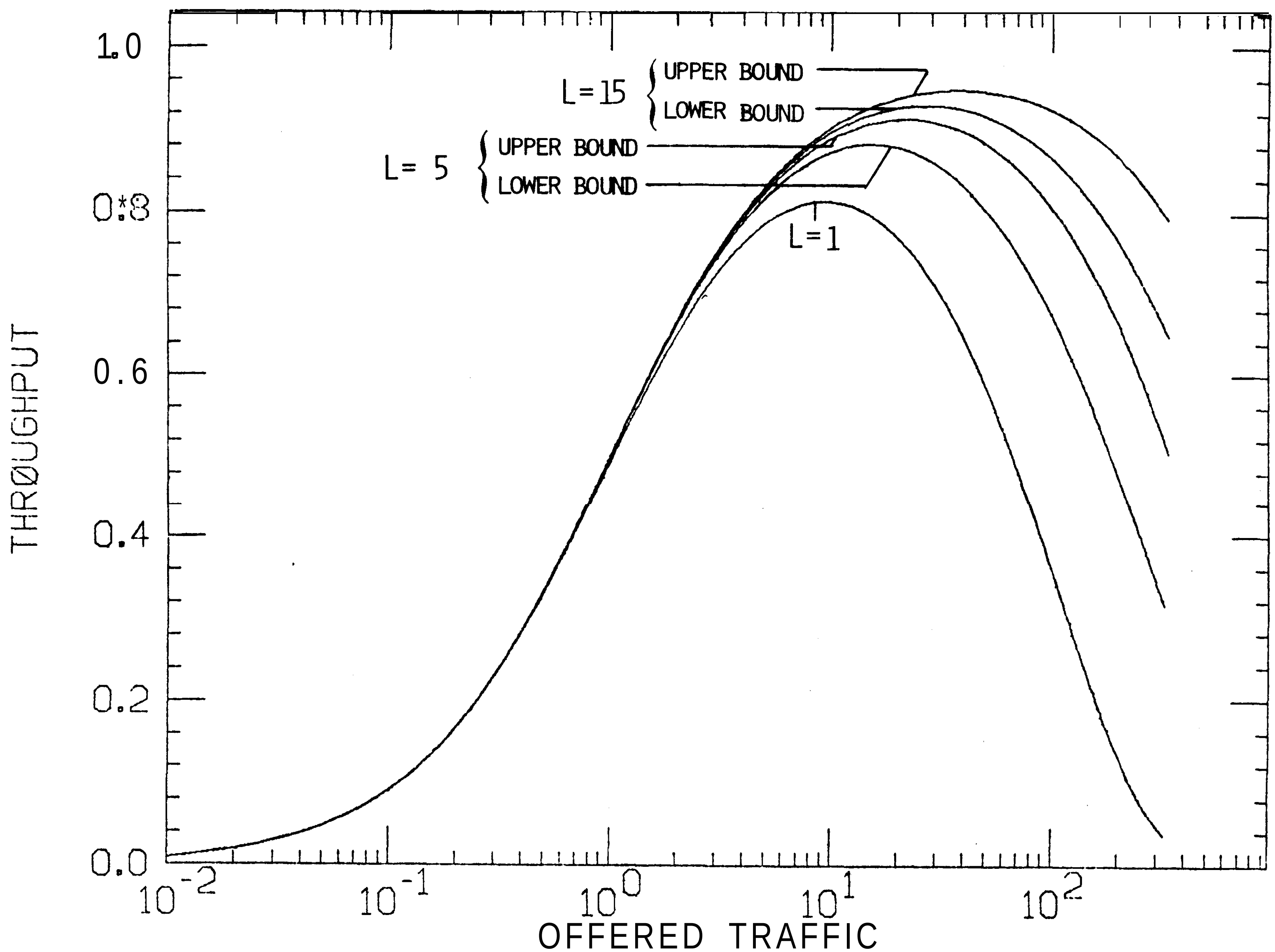
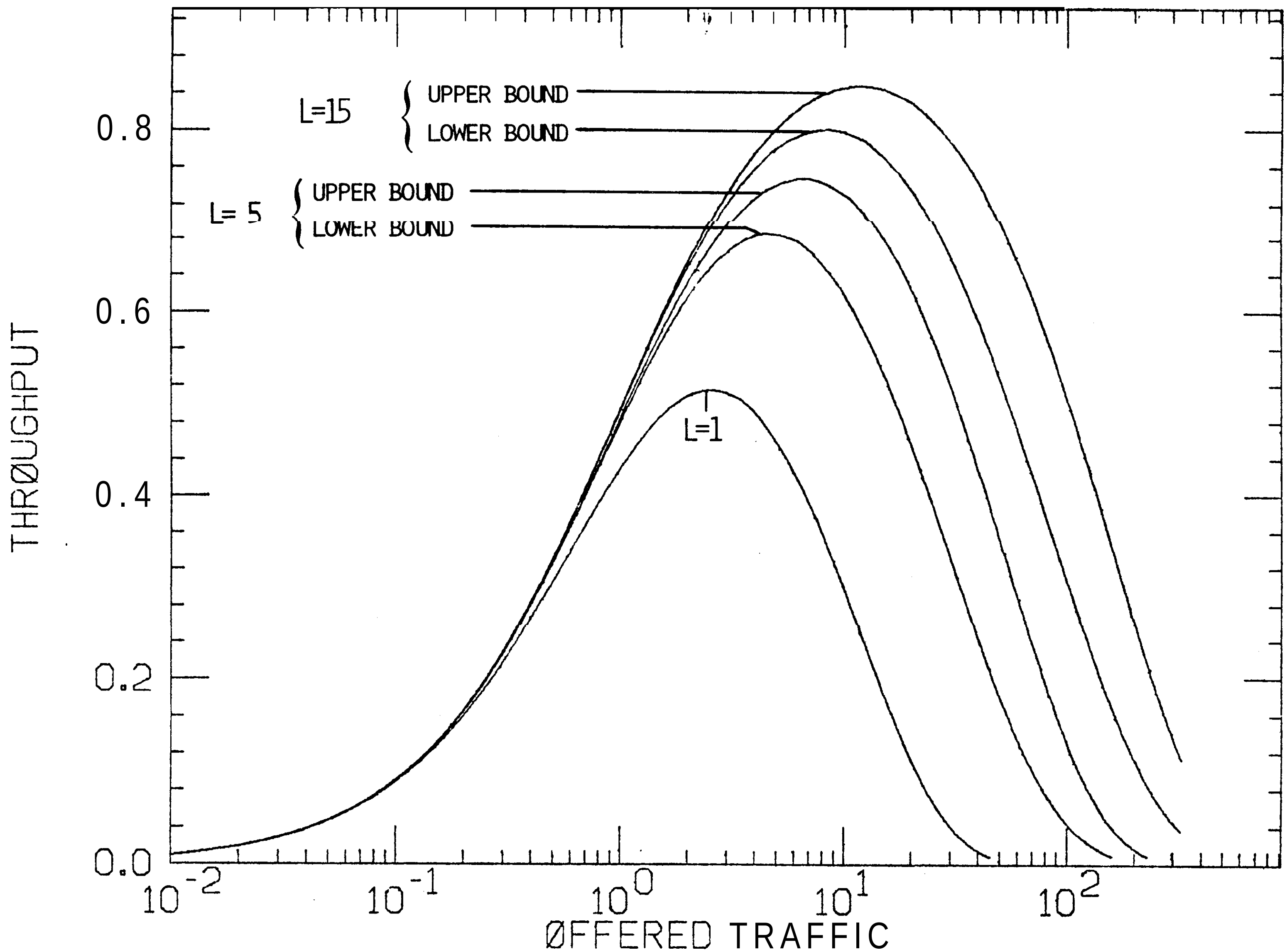


FIG. 4.5.B UNSLØTTED M_0 PERS. CSMA (A=0.1)



each message is given by

$$L = 8 - 7h$$

For this distribution, we can obtain the average length of a busy period. The distribution function of a busy period, B_n, conditioning on n arrivals at the beginning of a busy period is given by

$$F_n(k) = \begin{cases} 0 & \text{for } k < 1 + a \\ h^n & \text{for } 1 + a \le k < 8 + a \\ 1 & \text{for } k \ge 8 + a \end{cases}$$

Furthermore, its mean is given by

$$\bar{B}_n = 8 - 7h^n + a$$

Since the average length of a busy period is given by

$$\bar{B} = \sum_{n=1}^{\infty} \bar{B}_n \frac{(aG^*)^n e^{-aG^*}}{n!(1-e^{-aG^*})}$$

After simplificaton, we get

$$\bar{B} = 8 + a - \frac{7(e^{(h-1)aG^*} - e^{-aG^*})}{1 - e^{-aG^*}}$$

Combining (4.7), (4.8), (4.9) and the above equation together, we get

$$S^* = \frac{aG^* e^{-aG^*}}{a + 8 - 7e^{aG^*(h-1)} - e^{-aG^*}}$$

When a is small, using the asymptotic expansion $e^x = 1 + x + \frac{x^2}{2} + 0(x^3)$ we get

$$S^* = \frac{aG^* - (aG^*)^2 + 0(aG^*)^3}{a - (7h-8)aG^* - \frac{1}{2}(7(h-1)^2+1)a^2G^{*2} + 0((aG^*)^3)}$$

Solving the quadratic equation of G*, and then taking the smaller root, we obtain

$$G^* = \frac{(1-(8-7h)S^*) - \sqrt{[1-(8-7h)S^*]^2 - 4(1-[7(h-1)^2+1]\frac{S^*}{2})aS^*}}{2(1-[7(h-1)^2+1]S^*/2)a}$$

Recall under packet switching, from (3.13) the offered packet traffic is given by

$$G = \frac{(1-S) - \sqrt{(1-S)^2 - 4a(1-{}^{S}/2)S}}{2a(1-{}^{S}/2)}$$

where $S = (8-7h)S^*$

For the case $h = \frac{1}{2}$, a = 0.01, we tabulate the average number of retransmissions per successful message under message switching, $(\frac{G^*}{S^*} - 1)$, and that per successful packet under packet switching, $(\frac{G}{S} - 1)$, in Table 4-3 for slotted M_0-persistent CSMA scheme. The upper bound of $(\frac{G^*}{S^*} - 1)$ given in corollary 4.2 is also shown in Table 4-3 and is clearly very tight.

Finally, to check the validity of the assumptions made before, we release assumptions 1 and 3, i.e. the message independence assumption and the assumption that interevent times among the newly arrivals and retransmissions are independent and exponentially distributed, and conduct simulations for the slotted M_0-persistent CSMA scheme. The number of packets contained in each message is assumed to have the same distribution as that in the previous example, i.e. half of the messages is single packet messages, and the other half is eight packet messages. From Table 4.4 we can see that the simulation results and the analytic results obtained under the restricted assumptions are very close to each other as expected.

After analyzing both the slotted and unslotted versions of the M_0-persistent CSMA scheme, let us proceed to consider the difficulties encountered in the analysis of M_p-persistent CSMA under message switching when $p \neq 0$. In Fig. 4.6, we present a typical diagram of idle and busy periods of slotted M_p-persistent CSMA. Each busy period may consist

S^*	G^*/S^*-1	upper bound of G^*/S^*-1	G/S-l
0.02	0.099237	0.099258	0.100067
0.04	0.220239	0.220255	0.222495
0.06	0.371284	0.371451	0.375921
0.08	0.565249	0.565590	0.573930
0.10	0.823418	0.824288	0.839650
0.12	1.184342	1.186421	1.216000
0.14	1.725416	1.730935	1.793737
0.16	2.630175	2.647486	2.810375
0.18	4.477025	4.555560	5.255877

where S^* is the message arrival rate

(G^*/S^*-1) is the average number of retranmissions per successful message under message switching

(G/S-l) is the average number of retransmission per successful packet under packet swithcing

Table 4.3: Comparison of Average Number of Retransmissions Under Packet and Message Switchings for Slotted M_0-persistent CSMA

S^*	G*(by simulation)	G*(analytic result under message independence assumption)
0.02	0.0220	0.0220
0.04	0.0486	0.0488
0.06	0.0816	0.0823
0.08	0.130	0.125

Table 4.4: Comparison of Analytic Result and Simulation Result on Offered Channel Traffic for Slotted M_0- persistnet CSMA

of one or more transmission periods. The problem is that the lengths of the consecutive transmission periods are dependent. The longer the transmission period is, the more accumulated messages are lilely to occur at the start of the next transmission period, hence the longer duration the next transmission period tends to be when messages may consist of different number of packets. Nevertheless, we will study the slotted M_1-persistent CSMA scheme to show the change in performance as p goes to the other extreme by simulations. Again the distribution function of the number of packets contained in each message is assumed to be

$$F(k) = \begin{cases} 0 & \text{for } k < 1 \\ h & \text{for } 1 \leq k < M \\ 1 & \text{for } k \geq M \end{cases}$$

Furthermore, let $h = \frac{1}{2}$ and $a = 0.01$. Figure 4.7 shows the simulation results, indicated by "X" and "#", on the average number of transmissions per successful message under message switching when L = 5 and 16, i.e. M = 9 and 31, respectively. The average number of transmissions per packet under packet switching, i.e. when L = 1, is also plotted in Figure 4.7 for comparison.

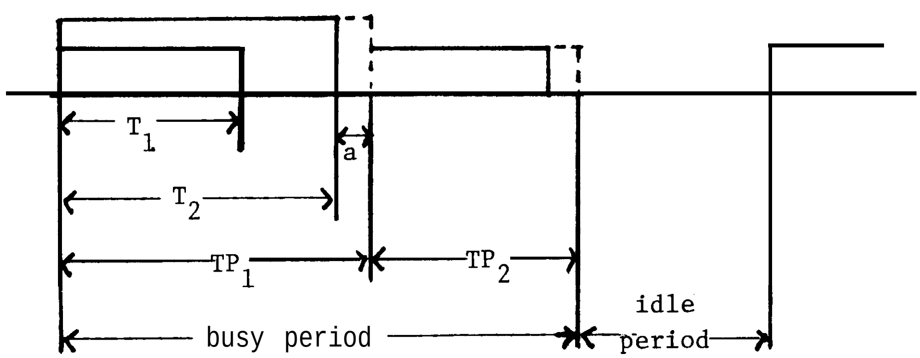


Figure 4.6: Busy and Idle Periods of Slotted M_p-persistent CSMA

FIG.4.7 SLØTTED M_1 PERS. CSMA (VAR. L ≤ N G

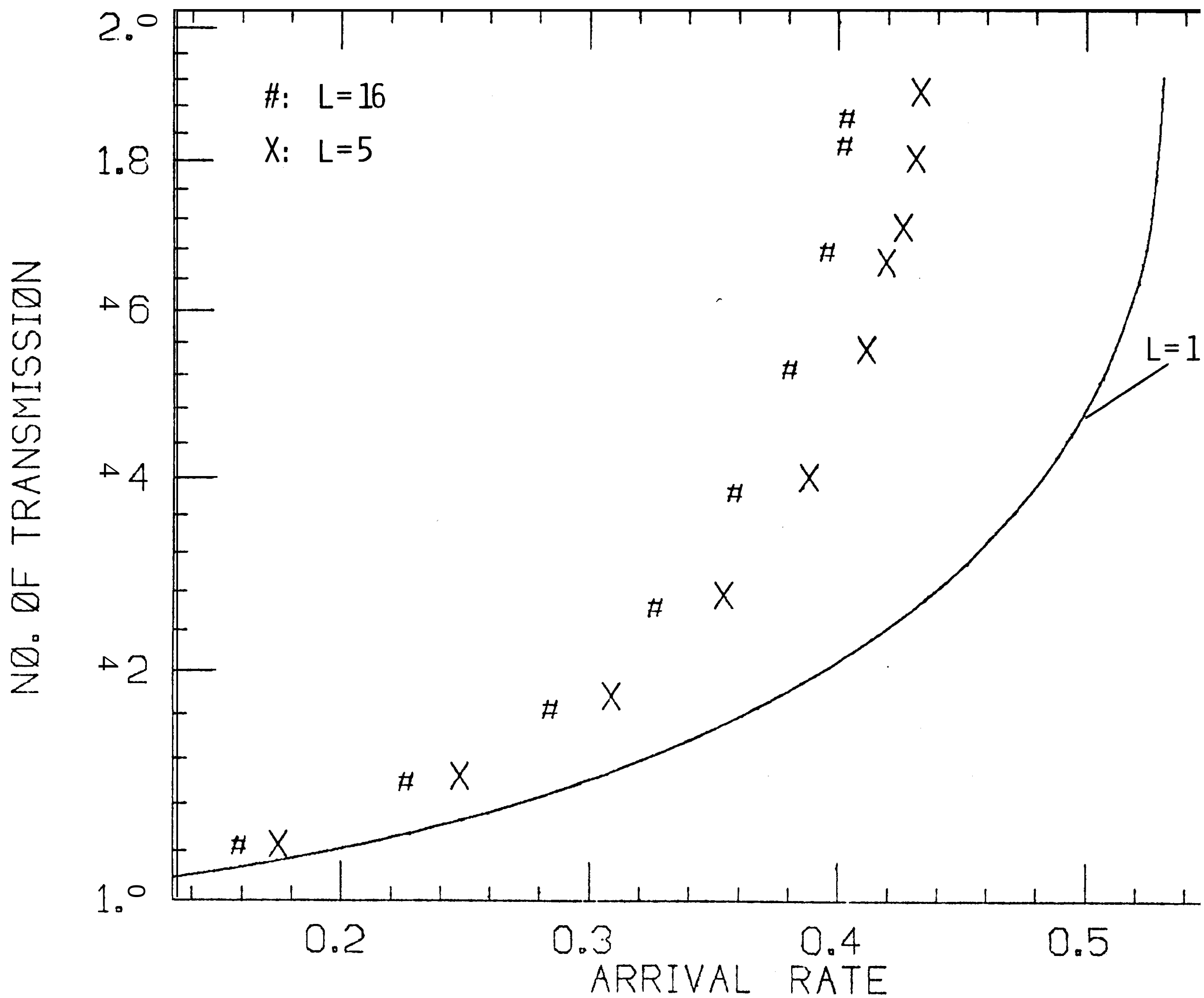
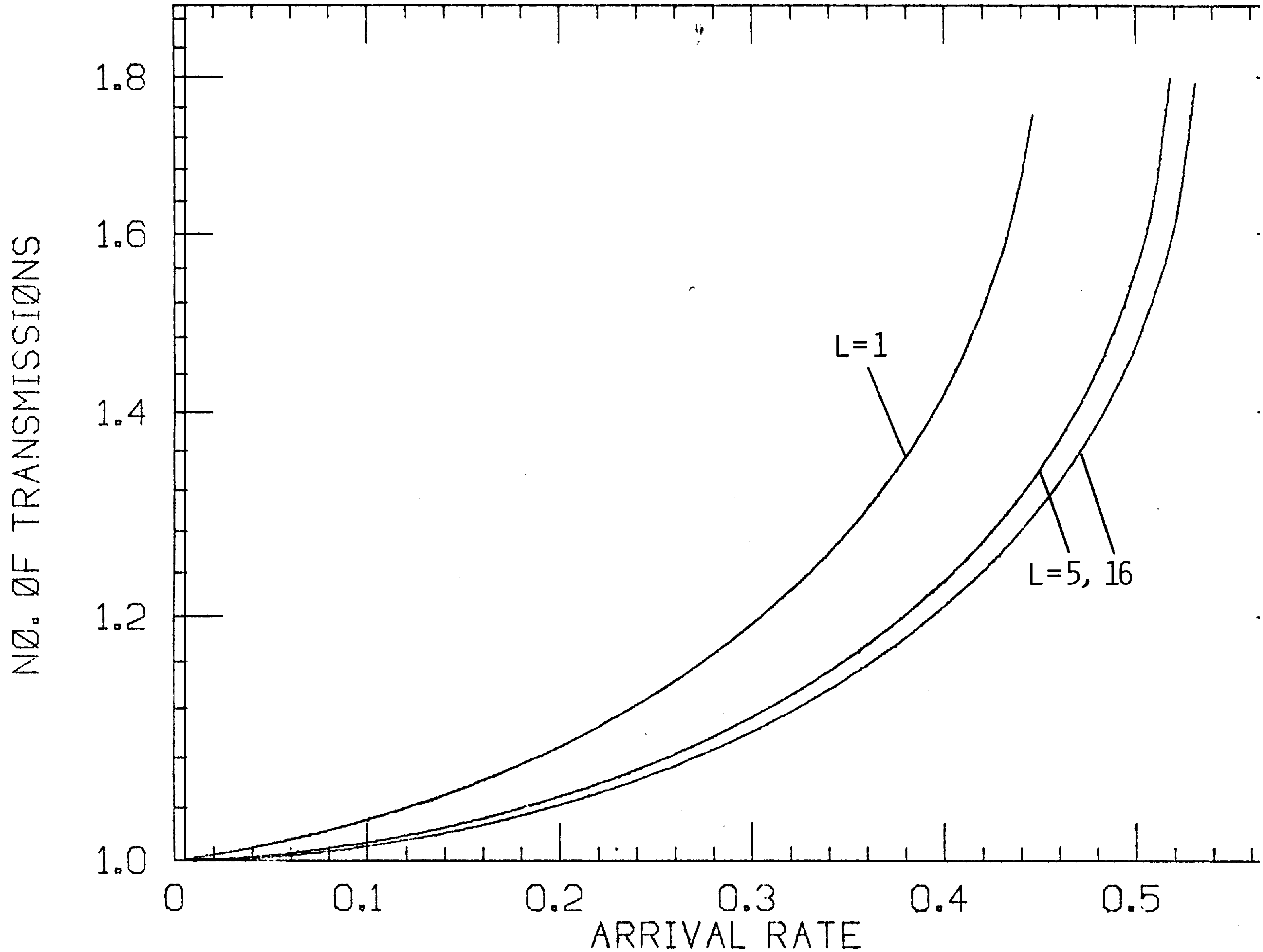


FIG. 1.8 SLØT. M_1 PERS. CSMA (FIXED LENGTH)



As we can see that under message switching not only the channel capacity becomes smaller, but also the number of transmissions per successful message is larger than that per successful packet under packet switching for both cases. Although the transmission delay under message switching may still be smaller than that under packet switching when the traffic intensity is low, the unstableness or low channel capacity under message switching makes packet switching more favorable in both cases. In Figure 4.8, we plot the number of tranmsissions per successful transmission versus the arrival rate when the number of packets contained in each message is fixed with length 5 and 16 which are the same as the means in the previous plot, respectively, for a = 0.1. The number of transmissions per successful packet under packet switching is again plotted for comparison. As we can see that when the coefficient of variation of message length is zero, i.e. all messages have fixed length, the performance under message switching is in fact better than that under packet switching in both throughput and delay. As the coefficient of variation increases, the performance under message switching may degrade and gradually become inferior to that under packet switching. That is to say on contrary to the M_0-persistent CSMA scheme, the performance of the M_1-persistent CSMA scheme under message switching can not be determined by the mean message length alone. The distribution of the number of packets contained in each message, or the degree of variation around the mean will have strong effect on the performance under message switching when the parameter p of M_p-persistent CSMA goes to 1.

Based on the above analysis, we make the following remark. If the channel is highly utilized and our primary concern is to achieve larger channel capacity or improve the channel stability, then message switching with M_0-persistent CSMA or optimum M_p-persistent CSMA seems to be desirable. On the other hand, if the channel is not highly utilized, and the primary concern is to reduce the number of transmissions per successful transmission or the transmission delay, then packet switching with M_1-persistent CSMA, or the other CSMA schemes such as optimum M_p-persistent CSMA, and optimum p-persistent CSMA which also provide better stability property, may be desirable, especially if the receiving station is part of a store and forward terrestrial network as we shall see in section 5.

4.C The ALOHA Scheme

In this subsection, we examine the performance of the ALOHA scheme under message switching. The slotted version is first considered. In the slotted ALOHA scheme, messages are restricted to transmit at the beginning of a slot whose length is equal to the transmission time of a single packet. Let S_i denote the message arrival rate of messages with length i and G_i denote the corresponding offered message traffic. Furthermore, recall S^* and G^* is the overall message arrival rate and offered message traffic and F(k) is the distribution function of the number of packets contained in each message. A message of length k beginning to transmit at time t which is the start of a time slot will be successful, if

(1) no new or retransmission messages of length n occurs during the interval (t-n, t+k-1) for $1 \leq n \leq \infty$ and $n \neq k$

(2) no other messages of length k except the one cited above occurs during the interval (t-k, t+k-1).

Both events are independent. The probability of occurrence of the first event is equal to $\prod_{\substack{n=1\\n\neq k}}^{\infty} e^{-(n+k-1)G_n}$ and that of the second event is equal to $e^{-(2k-1)G_k}$. Hence, under steady state

$$S_k = G_k e^{-(2k-1)G_k} \prod_{\substack{n=1\\n\neq k}}^{\infty} e^{-(n+k-1)G_n}$$

$$= G_k\, e^{-\sum_{n=1}^{\infty}(n+k-1)G_n}$$

Using the fact that $G^* = \sum_{n=1}^{\infty} G_n$, after simplification we get

$$S_k = G_k\, e^{-(k-1)G^*}\, e^{-\sum_{n=1}^{\infty} nG_n} \tag{4.36}$$

Under packet switching, the primary reason why slotted ALOHA can achieve twice the channel capacity of pure ALOHA is that collisions of packets occur only under complete overlapping of packets in the slotted ALOHA scheme. That is to say in slotted ALOHA the vulnerable interval of any packet is the length of the packet under packet swithcing. Now, under message switching the length of the vulnerable interval of any message,instead of being equal to the length of the message, is equal to the sum of its own length plus i-l with respect to messages with length i. Hence the longer the message is, the harder the transmission will be. Long messages are discriminated in this case. When the arrival rate increases, G_n/S_n increases more rapidly for large n. Since long messages also have a greater chance to conflict other messages, the fast growing offered traffic of long messages will make the system more susceptible to saturation than before. It shouldn't be a surprise that the channel capacity of slotted ALOHA under message switching is less than e^{-1}, the channel capacity of slotted ALOHA under packet switching, if some messages contain more than one packet. The following theorem proves this fact.

Theorem 4.7:

The channel capacity of slotted ALOHA under message switching is less than e^{-1}, if some messages contain more than one packet.

Proof:

Since

$$S' = \sum_{k=1}^{\infty} kS_k$$

Substituting (4.36) into the above equation, we get

$$S' = e^{-\sum_{n=1}^{\infty} nG_n} \sum_{k=1}^{\infty} kG_k e^{-(k-1)G^*}$$

Since $e^{-(k-1)G^*} \leq 1$ and the equality only holds when k = 1, we get

$$S' < \left(\sum_{k=1}^{\infty} kG_k\right) e^{-\sum_{n=1}^{\infty} nG_n}$$

Due to the simple fact that the maximum value of $f(x) = xe^{-x}$ is e^{-1} which occurs at x=1 we conclude that

$$S' < e^{-1}$$

Hence, the channel capacity, the maximum achievable value of S', is less than e^{-1}.

For unslotted ALOHA, Ferguson [24][25] has studied its performance under message switching using finite source model where message length can have arbitrary distribution. The anslysis based on the message independent assumption shows that the throughput under packet switching is larger than that under message switching. Under our infinite source model we can prove that the same result holds for arbitrary distribution of the number of packets contained in each message.

By similar argument as in the slotted version, we obtain

$$S_k = G_k e^{-kG^*} e^{-\sum_{n=1}^{\infty} nG_n} \tag{4.37}$$

Theorem 4.8:

The channel capacity of pure ALOHA under message switching is less than or equal to $\frac{1}{2}e^{-1}$. (The equality only holds for the case where each message contains the same number of packets.)

Proof:

$$S' = \sum_{k=1}^{\infty} kS_k$$

Substituting (4.37) into the above equation, we get

$$S' = \left(\sum_{k=1}^{\infty} kG_k e^{-kG^*}\right) e^{-\sum_{n=1}^{\infty} nG_n} \tag{4.38}$$

Let

$$q_{k,0} = \frac{G_k}{\sum_{i=1}^{\infty} G_i} \tag{4.39}$$

then $\{q_{k,0}\}$ represents a discrete probability distribution.

Furthermore, let

$$G' = \sum_{k=1}^{\infty} kG_k$$

and

$$h = \sum_{k=1}^{\infty} k\, q_{k,0}$$

After simple manipulation, we get

$$h = \frac{G'}{G^*} \tag{4.40}$$

Since $G' \leq 1$ under steady state, we get

$$G^* \leq \frac{1}{h}$$

Equation (4.38) can be rewritten in the following form

$$S' = G^* e^{-G'} \sum_{k=1}^{\infty} k\, q_{k,0}\, e^{-kG^*} \tag{4.41}$$

If all the probability mass is concentrated at h,

$$\sum_{k=1}^{\infty} k\, q_{k,0}\, e^{-kG^*} = h e^{-hG^*}$$

Otherwise, there will be some mass in both regions $[1,h)$ and (h,∞).

Let us move the amount of mass y at $m(>h)$ to h and also the amount of massx at $n(<h)$ to h for some m,n such that

$$(h-n)x = (m-h)y$$

and

$$x \leq q_{n,0}$$
$$y \leq q_{m,0}$$

We refer to this new probability distribution after perturbation as $\{q_{k,1}\}$. Notice $\{q_{k,1}\}$ and $\{q_{k,0}\}$ has the same mean, h.

Let

$$A = \{h\} \cup \{k \mid k > 0,\ k \text{ is an integer})$$

and

$$f(y,G^*) = \sum_{k\varepsilon A} k\, q_{k,1} e^{-kG^*} - \sum_{k=1}^{\infty} k\, q_{k,0} e^{-kG^*}$$

Clearly,

$$f(y,G^*) = h\left(1+\frac{m-h}{h-n}\right) ye^{-hG^*} - mye^{-mG^*} - \frac{m-h}{h-n} nye^{-nG^*}$$

and

$$f(y,0) = 0$$

$$\frac{\partial}{\partial G^*} f(y,G^*)\Big|_{G^*=0} = y(m-h)(m-n) > 0, \qquad \text{for } y > 0.$$

The curve of $f(y,G^*)$ has the form shown in Fig. 4.9 for any $y > 0$.

$$f(y,\tfrac{1}{h}) = hy\left(\left(1 + \frac{m-h}{h-n}\right)e^{-1} - \frac{m}{h}\, e^{-\frac{m}{h}} - \frac{m-h}{h-n}\,\frac{n}{h}\, e^{-\frac{n}{h}}\right)$$

Using the fact that $xe^{-x} < e^{-1}$ for $x \neq 1$, we get

$$f(y,\tfrac{1}{h}) > hy\left(\left(1 + \frac{m-h}{h-n}\right)e^{-1} - e^{-1} - \frac{m-h}{h-n}\, e^{-1}\right) = 0$$

Hence

$$f(y,G^*) > 0 \text{ for } y > 0,\ G^* < \frac{1}{h}$$

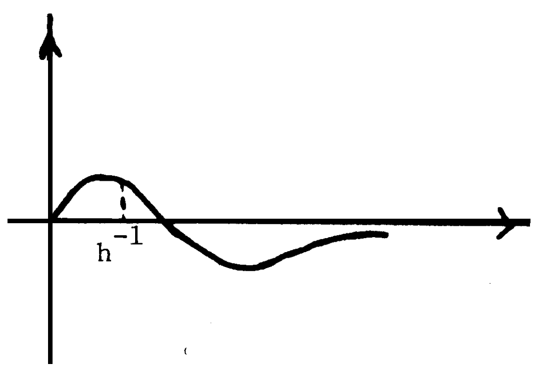


Figure 4.9: $f(y,G^*)$ Versus G^*

Repeating the same procedure, we can move all the probability mass not at h to h. Let $\{q_{k,i}\}$ be the discrete probability after the i-th iteration. Then $\sum_{k\varepsilon A} k\, q_{k,i} e^{-kG^*}$ is a monotonically increasing function of i and the final limit is he^{-hG^*}. That is to say

$$\sum_{k=1}^{\infty} k\, q_{k,0} e^{-kG^*} < he^{-hG^*}$$

Combining the above inequality with (4.41) and (4.40), we get

$$S' \leq hG^* e^{-2hG^*}$$

Since the maximum value of xe^{-2x} is $\frac{1}{2}e^{-1}$ which occurs at $x = \frac{1}{2}$, we get

$$S' \leq \tfrac{1}{2}e^{-1}$$

Using message independent assumption, we can obtain optimistic lower bound on channel capacity in simple form for both slotted and unslotted ALOHA. The throughput equation reduces to the following form under message independent assumption

Slotted ALOHA $$S' = G^* e^{-G^*(L-1)} \sum_{k=1}^{\infty} kP_k e^{-kG^*}$$

Unslotted ALOHA $$S' = G^* e^{-G^*L} \sum_{k=1}^{\infty} kP_k e^{-kG^*}$$

Let C be the squared coefficient of variation of the number of packets contained in each message. The following theorem gives the lower bounds on channel capacities of both slotted and unslotted versions, **respectively** As we shall see the lower bounds are almost inversely proportional to C. Furthermore, under message independent assumptions, we can prove the upper bounds on channel capacities given in the previous two theorems are still valid by similar argument.

Theorem 4.9:

Using message independent asumption, the lower bounds on channel capacity under message switching for the ALOHA scheme are

Slotted ALOHA $$\max_{G^*} S' \geq \frac{e^{-1}}{C+2-1/L}$$

Unslotted ALOHA $$\max_{G^*} S' \geq \frac{e^{-1}}{C+2}$$

respectively.

Proof:

We only prove the unslotted case. The slotted case can be proved by similar argument.

$$S' = G^* e^{-G^*L} \sum_{k=1}^{\infty} kP_k e^{-kG^*}$$

$$= LG^* e^{-G^*L} \sum_{k=1}^{\infty} \frac{kP_k}{L} e^{-kG^*}$$

Let

$$f(x) = e^{-xG^*}$$

Since f(x) is a convex function, i.e. $\frac{d^2f(x)}{dx^2} \geq 0$, we can prove by induction

$$\sum_{k=1}^{m} \alpha_k f(x_k) \geq f(\sum_{k=1}^{m} \alpha_k x_k)$$

for all $\alpha_k > \sum_{k=1}^{m} \alpha_k = 1$

Letting α_k equal to kP_k/L and x_k equal to k, we get

$$\sum_{k=1}^{\infty} \frac{kP_k}{L} e^{-kG^*} \geq e^{-\sum_{k=1}^{\infty} k^2P_kG^*/L}$$

$$= e^{-(C+1)G^*L}$$

Hence

$$S' \geq LG^* e^{-(C+2)G^*L}$$

$$\max_{G^*} S' \geq \frac{e^{-1}}{C+2}$$

4.D The Dynamic Reservation Scheme [13]

In the dynamic reservation scheme considered here, each user first makes a request for service on the channel when it has a message packet ready for transmission. After the request is accepted by the central station, the message will be scheduled for transmission. The central station maintains a queue of requests and informs the user of its position in the queue. Although the conflicts between messages have been avoided, the conflicts between requests are inevitable. Those random access schemes mentioned in the previous section, e.g. slotted or pure ALOHA, various CSMA schemes, can be used to multiplex requests on the channel. In order to prevent the conflicts between request and message packets, the channel is either time divided or frequency divided between the two types of packets.

The reserved ALOHA scheme [4] is a typical example of time division scheme. The channel has two different states, ALOHA and RESERVED. On start up and every time thereafter when the reservation queue becomes empty, the channel is in the ALOHA state. In this state, all slots are small and the ALOHA scheme is used for request transmissions. The first successful reservation causes the RESERVED state to begin. However, after every M RESERVED slots, one slot is subdivided into V small ALOHA slots. Before a data packet is'transmitted, the user transmits a reservation in a randomly selected one of the V small slots in the next ALOHA group. Upon seeing the reservation, each user adds the number of slots requested to a count which records the number of slots currently reserved. Thus, there is a common queue for all users and by broadcasting reservations they can claim space on the queue.

The frequency division scheme proposed by Kleinrock and Tobagi [13] is called split channel reservation multiple access (SRMA). Two versions of SRMA have been considered, i.e. the RAM and RM schemes. In the request answer-to-request message (RAM) scheme, the available bandwidth is divided into three channels: one used to transmit requests, the second used to transmit answers to requests, the third used for the messages themselves The request channel will operate under random access mode. When a user has a message ready for transmission, it sends on the request channel a request packet containing information about the address of the user, and in the case of multipacket messages, the length of the message. At the correct reception of the request packet, the scheduling station will compute the time required to serve the backlog on the message channel and then transmit back to the users, on the answer-to-request channel, an answer packet containing the address of the answered user and the time at which it should start transmission. In the RM scheme, the total available bandwidth is divided into only two channels: the request channel and the message channel. Again, the request channel will operate under random access mode. When a user has a message ready for transmission, it sends on the request channel a request packet. When correctly received by the scheduling station, the request joins the request queue. The scheduling station may adopt any priority schelduling algorithm. When the message channel becomes available, an answer packet containing the address of the next user is transmitted by the station on the message channel. After receiving the answer packet, the user starts to transmit its message on the message channel. If a user does not hear the answer after a certain amount of time, it will assume the previous request to be unsuccessful and retransmit the request packet. Since the time between receiving and answering of a correctly received request is equal to the queueing delay of the request packet in the request queue of the scheduling station, it is a random variable. Clearly the user may undertake some additional transmissions of a request after it is correctly received. The shorter the time out period is, the larger the traffic is on the request channel and hence, the smaller the probability of success. On the other hand, the longer the time out period is, the smaller the traffic is on the request channel but the longer the delay between retransmissions.

In dynamic reservation scheme, priority scheduling may be employed either to reduce the mean queueing delay in the central station or give higher priority to a certain class of messages. As pointed out earlier, the RM scheme might be incorporated with any priority scheduling. One priority scheduling which is referred to as shortest processing time (SPT) scheduling [23] achieves the minimum average waiting time among all the nonpreemptive scheduling algorithm. Under the shortest processing time discipline, the scheduling station always selects the shortest message to transmit after the message channel is available. Notice this scheduling algorithm is not a feasible CPU scheduling algorithm since the processing time of each job is not known beforehand. It is usually used to obtain a-lower bound on waiting' time for evaluating the performance of other practical sucheduling algorithms. Under the RM scheme it is indeed feasible, since the message length is contained in the request packet. The priority scheduling may also be extended to other reservation schemes under ground radio channel if we can ensure the high reliability of the answer packet, i.e. when received correctly by any user to be received correctly by all users. Robert [4] suggests one way to do this by properly endcoding the reservation. The strategy uses the standard packet sum-check hardware, and sends three independently sumchecked copies of the reservation data. The high reliability is assured even if the channel error rate is high. With highly reliable answer packets the priority scheduling can be implemented in the following way. Assume the answer packet contains not only the address of the transmission user but also the length and the priority of the message. (Under the SPT scheduling, the priority of the message is in fact the length of the message and can be omitted in the answer packet.) After hearing the answer packet, each user having a pending message will delay its message transmission time automatically by an amount equal to the message length specified in the answer packet if the priority specified in the answer packet is higher than that of its pending message. The user matching the address specified in the answer packet will start to transmit the message at the schedule time if it does not hear anv answer packets of messages with higher prioritfes before it starts to transmit its message.

Now let us examine the available analytic results on various reservation schemes. For the RAM scheme [13], the total delay can be decomposed into two parts. The first part is the time required for a request packet to be successfully received at the central station. The delay depends upon the random access scheme used to reserve a request and is given in section 2. The second part is the time between reception of the request packet at central station and the end of the message transmission. If the scheduling discipline at the central station is first come first served, the queueing delay is exactly that of the M/G/l system. If the scheduling discipline is SPT, the average queueing delay is given in [23]. The maximum bandwidth utilization is determined by the fact that the throughput of the request channel does not exceed its capacity and the utilization of the message channel does not exceed one. The analytic result for the RM scheme is hard to obtain but simulation results have been obtained in [13] which shows that the performance of the RM scheme is comparable and even slightly superior to the RAM scheme. In [4], the performance of reserved ALOHA scheme has been analyzed. The transmission delay can be decomposed into reservation delay, central queueing delay and propagation delay and evaluated separately.

In any reservation scheme, the channel capacity under message switching is larger than that under packet switching. Since under packet switching, the request rate in the request channel is L times larger than that under message switching, more bandwidth needs to be allocated to the request channel. Let us consider the transmission delay when the channel is not highly utilized. For the RAM scheme, packet switching may lead to smaller transmission delay especially when FCFS discipline is adopted at the central station. If the message channel has the same bandwidth under message switching and packet switching, the mean queueing delay at the central station under message switching will be at least L times larger than that under packet switching. This fact can be easily derived by comparing the mean queueing time under M/G/l and M/D/l systems. On the other hand, the mean reservation delay under packet switching is larger than that under message switching and the mean delay due to interpacket gaps under packet switching will be (L-l) times the mean reservation delay. Under packet switching, usually the bandwidth allocated to message channel is smaller than that under message switching, so the contention in the request channel can be alleviated and the total delay may become smaller. Nevertheless, priority scheduling can be used in the central station to reduce the queueing delay under message switching. For the RM scheme, if the request for next packet transmission is not issued until the request for the previous packet is acknowledged, packet switching will lead to larger transmission delay. For communications via satellite channels, the propagation delay is very large. Hence, message switching will lead to smaller transmission delay in reserved ALOHA scheme. Generally speaking, message switching is more favorable in dynamic reservation schemes.

5. EXAME'LE

In this section, we will analyze the performance of the computer communication netwrok shown in Fig. 5.1. The terrestrial network has the same topology as the communication network, CIGALE, within CYCLADES [20], which is a general purpose computer network being installed in France. The performance measure under investigation is the mean transmission delay of the messages from the group of terminals indicated in the figure to various stations in the network. The message transmissions between the terminals and station A are via radio channel and the message transmissions between stations in the network are via terrestrial links. Furthermore, we assume the access scheme employed in transmitting the messages from the terminals to station A is slotted M_0-persistent CSMA scheme. All the terristrial links are assumed to be full duplex. The numbers on the terrestrial links represent servers and their queues. Thus 3 refers to the server which transfers messages from node C to node A and 2 refers to the server which transfers messages in the opposite direction. Traffic moving in the two oppisite directions along the same link is assumed to be noninterfering. Each station receives external traffic which forms a Poisson process. We also assume that each message arriving from outside to each station has equal probabilities of having any of the other 4 stations as its final destination. The routing algorithm of the networks is assumed to be fixed and will be described later. All the above assumptions about the terrestrial network have been adopted by Gelenbe [21] in modeling a similar network under packet switching.

Let C_i be the channel capacity, the number of packets that can be transmitted per second, of link i. The channel capacity of each link is indicated in Table 5.1a. The fixed routing algorithm is summarized in Table 5.1b. The routes which are not shown in Table 5.1b are the links which directly connect the source stations and destination stations.

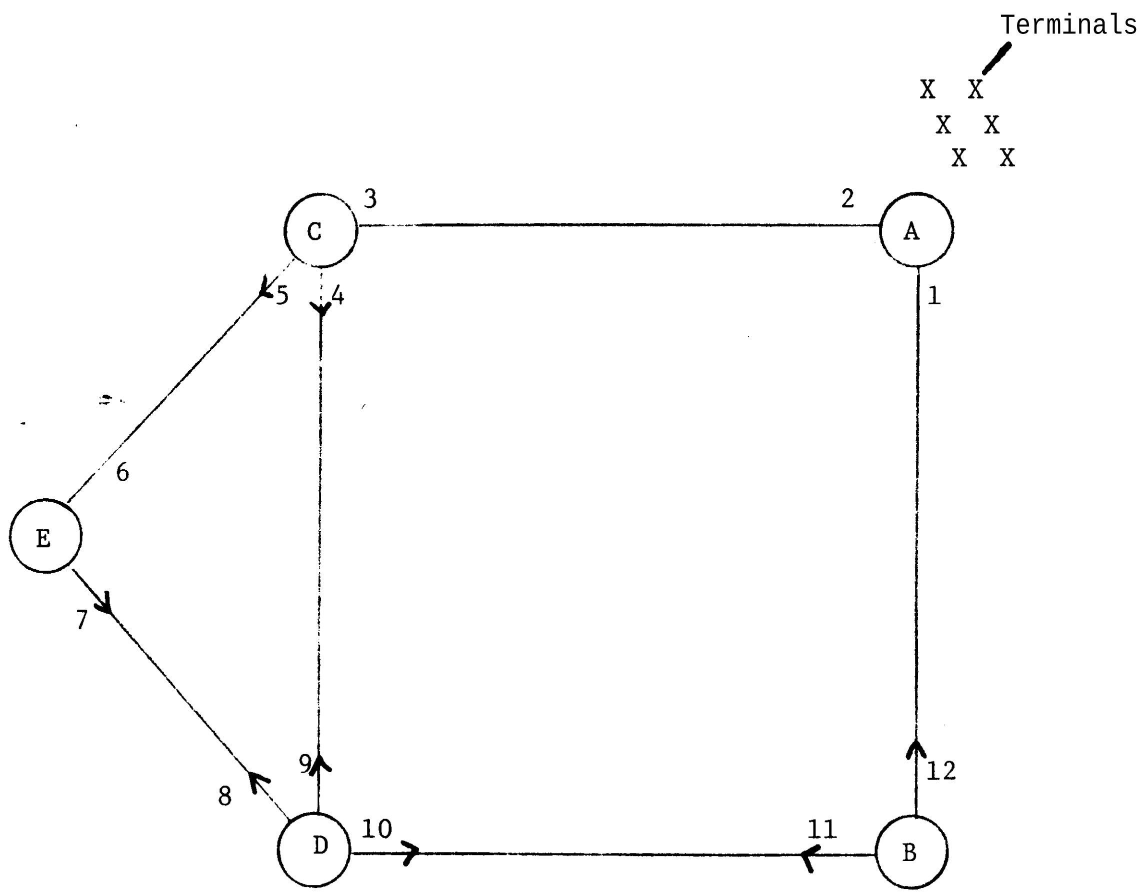


Figure 5.1: NETWORK TOPOLOGY

Link	C_i (packet/sec)
1, 12	50
2, 3	80
4, 9	70
5, 6	45
7, 8	50
10, 11	70

Table 5.1a: Channel Capacity of Each Link in the Terrestrial Network

Source Stations	Destination Stations	Route
A	D	2, 4
A	E	2, 5
B	C	12, 2
B	E	11, 8
C	B	4, 10
D	A	9, 3
E	A	6, 3
E	B	7, 10

Table 5.1.b: Routing Table

We first consider the mean transmissions delay under message switching. In this case, the transmission delays from terminals to destination nodes consist of two parts. The first part of the delay denoted by T_1 is the mean transmission delay from the terminal to station A via radio channel. The second part of the delay denoted by T_2 is the mean transmission delay from station A to the destination station via the terrestrial network. For example, let the number of packets contained in each message have geometric distribution with mean five. The message arrival rate to each node is indicated in Table 5.1.c. Since the mean pretransmission delay under slotted M_0-persistent CSMA scheme is $\frac{a}{2}$, the mean transmission delay in (4.15) becomes

$$T_1 = (\frac{G^*}{S^*} - 1)(L + R') + \frac{3a}{2} + L$$

where

$$R' = L + \frac{5}{2}a + \delta + \alpha$$

and

L = mean number of packets contained in each message

node	message arrival rate	(equivalent) packet arrival rate
A	12	60
B	16	80
C	16	80
D	16	80
E	16	80

Table 5.1c: External Arrival Rate (per second)

By theorem 4.1, we can evaluate the upper bound and lower bound of G^*/S^*. From the previous analysis, we know the upper bound and lower bound of S^* is quite close to each other. Hence we will use the average value of the upper bound and lower bound as an approximattion of G/S. Let us assume that a = 0.05, δ = 0.5, and α = 5. The approximate value of G^*/S^* is 2.58 and the upper bound and lower bound of G^*/S^* is 2.577 and 2.589, respectively, when S^* = 0.12. Furthermore, let us assume that the transmission time of a packet via the radio channel is 10 msec, then from the formula just mentioned above we obtain that T_1 is equal to 218.6 msec.

To evaluate the delay T_2 in the terrestrial network, we have two alternatives. The first one is to approximate the geometric distribution of message length by an exponential distribution with the same mean. The problem now becomes analytically tractable. By the Jackson theorem [I], we can obtain the mean queue length in each queue and then by Little's formula, we can get the mean response time. The second method utilizes the fact that the terrestrial network is congested and uses the diffusion approximation technique [22] to evaluate the mean queue length in each queue and again by Little's formula to obtain the mean response time. In Table 5.2a we tabulate the delay T_2 to each station in the network under both approximations. The simulation results which consist of not only the point estimations but also the half widths of the 95% confidence intervals are also included in the same table. The diffusion approximations are very close to the simulation results and the exponential approximations have about 10% errors.

Next, we consider the mean delay under packet switching. Here we assume that at terminals messages are transmitted packet by packet. After the transmission of the first packet, the terminal waits for the acknowledgement for station A. If the packet is successfully transmitted, the terminal starts to schedule the transmission of the second packet of the message, if any. Otherwise, the terminal schedules the retransmission of the first packet according to the retransmission delay distribution. A similar remark holds for the other packets of the message. Hence, the mean total delay can be decomposed into two parts. The first part, $\tilde{T}_1$, is the mean time in between a message is ready on a terminal and the last packet of the message has been accepted by station A. The second part, $\tilde{T}_2$, is the mean time required for the last packet to reach its destination station via the terrestrial network.

From (4.16) and the fact $d = \frac{a}{2}$, we get

$$\tilde{T}_1 = L(\frac{G}{S} - 1)(1 + R') + (\frac{5}{2}a + \alpha)(L - 1) + L + \frac{3}{2}a$$

where

$$R' = \frac{5}{2}a + \alpha + \delta$$

Under the same assumption on the parameters' values as before, we obtain that $\tilde{T}_1$ is equal to 713.3 msec. Again, $\tilde{T}_2$ can be evaluated using the diffusion approximation. Since we assume that each packet has fixed length, using an exponential distribution of the same mean to approximate the distribution of message length does not seem to be a very sound approach. Note that the squared coefficient of variation of an exponential distribution is 1 but that of a constant distribution is 0. In Table 5.2b, we compare the simulation results obtained under diffusion approximation and exponential approximation. As expected, the results under diffusion approximation are very close to the simulation results and the results under exponential approximation are very inaccurate in this case. The simulation results consist of both the point estimations and the half widths of the 95% confidence intervals as before.

Comparing the relative performance of packet switching and message switching, it appears that message switching has smaller delay via the radio

	mean message delay time T2 (msec)		
	exponential approximation	diffusion approximation	simulation
T_2^{AB}	138.6	142.9	142.9 (exact)
T_2^{AC}	333.3	303.6	299.7 $\pm$ 5.6
T_2^{AD}	666.7	609.6	604.8 $\pm$ 13.0
T_2^{AE}	833.4	762.1	760.5 $\pm$ 16.5

Table 5.2a: T_2 for each destination node when node A is the source node (message switching)

	mean packet delay time $\tilde{T}_2$ (msec)		
	exponential approximation	diffusion approximation	simulation
$\tilde{T}_2^{AB}$	28.57	24.29	24.29 (exact:
$\tilde{T}_2^{AC}$	66.7	36.92	36.89 $\pm$ 1.45
$\tilde{T}_2^{AD}$	133.3	76.31	76.27 $\pm$ 2.56
$\tilde{T}_2^{AE}$	166.7	95.49	94.84 $\pm$ 3.26

Table 5.2b: $\tilde{T}_2$ for each destination node when node A is the source node (packet switching)

channel and larger delay via the terrestrial network. A better approach in this case seems to be a mixture of the two, i.e. to use message switching for transmissions between terminals and station A via radio channel, then chop each message into packets after it is received by station A and use packet switching for transmission via terrestrial links. The mean total transmission delay in this case will be $T_1 + \tilde{T}_2 + (L - 1) T_3$ where T_3 is the mean transmission time of a packet over the first link in the route and $(L - 1) T_3$ is equal to 57.5 msec in this case.

If we keep the topology of the network unchanged and reduce the service rate of each channel and external arrival rate to each station by one half, the utilization of each link in the terrestrial network is unchanged but the utilization of the radio channel will be reduced by one half. Now the radio channel is only under low utilization and the probability of conflict is greatly reduced. Hence, the peroformance under message switching and packet switching via radio channels becomes closer. The delay T_1 under message switching is 97.5 msec and the delay $\tilde{T}_1$ under packet switching is 180.8 msec. The delay over the terrestrial network is tabulated in Table 5.3 and 5.4 for T_2 under message switching and $\tilde{T}_2$ under packet switching, respectively.

As pointed out before, the mean total transmission delay of the mixed strategy is equal to $T_1 + \tilde{T}_2 + (L-1) T_3$. Now the value of $(L - 1) T_3$ is 115 msec. Hence, the performance of packet switching is in fact better than-that of the mixed strategy in this case. That is to say if the utilization of the channel is low, packet switching should be used through out the transmission. In fact, we can use the slotted M_1-persistent CSMA scheme to reduce mean transmission delay from terminals down to 89.4 msec under packet switching. The transmission delay can be further reduced by using a smaller α, mean retransmission delay, as long as we are certain that

	mean message delay time T_2 (msec)		
	exponential approximation	diffusion approximation	simulation
T_2^{AB}	277.1	285.7	285.7 (exact)
T_2^{AC}	666.7	607.2	599.3 $\pm$ 11.3
T_2^{AD}	1333	1219	1210 $\pm$ 26
T_2^{AE}	1667	1524	1521 $\pm$ 33

Table 5.3: T_2 for each destination node when node A is the source node (message switching)

	mean packet delay time T_2 (msec)		
	exponential approximation	diffusion approximation	simulation
$\tilde{T}_2^{AB}$	57.14	48.57	48.57 (exact)
$\tilde{T}_2^{AC}$	133.3	73.83	73.78 $\pm$ 2.90
$\tilde{T}_2^{AD}$	266.7	152.6	152.5 $\pm$ 5.1
$\tilde{T}_2^{AE}$	333.3	191.0	189.7 $\pm$ 6.5

Table 5.4: $\tilde{T}_2$ for each destination node when node A is the source node (packet switching)

the channel is in low utilization.

With the insight from performance analysis, it is quite clear what is the appropriate strategy to be taken in order to improve the performance of the network under a given load, e.g. mixed strategy should be used when the radio channel is highly utilized and packet switching should be used when the radio channel is not highly utilized. As pointed out earlier, one of the nice impacts of performance analysis is that it often leads to better control strategy.

6. CONCLUSION

Computer communication networks have increased in utility in recent years. One way to communicate is via multiaccess broadcast channels. Both packet switching and message switching can be employed to transmit information. A new class of random access schemes referred to as the M_p-persistent CSMA scheme is proposed. The M_p-persistent CSMA scheme incorporates the nonpersistent CSMA scheme and the l-persistent CSMA scheme, both slotted and unslotted versions, as its special cases with p = 0 and 1, respectively. It is similar to p-persistent CSMA in the sense that they both try to reduce the interference due to terminals sensing the channel busy by approximately (1 – p), when a is small. Both slotted and unslotted versions of M_p-persistent CSMA lead to closed form expressions for throughput equations under packet switching and make the determination of the optimum p to operate an easy task. Under packet switching, the unslotted version of optimum M_p-persistent CSMA achieves larger channel capacity and smaller transmission delay than the currently available unslotted CSMA schemes and the slotted version achieves larger channel capacity than the optimum p-persistent CSMA scheme.

Furthermore, the performance of various random access schemes is examined and compared with that under packet switching. We first analyze the performance of M_0-persistent CSMA under message switching and obtain tight upper bound and lower bound on throughput without any specific assumption on the distribution of the number of packets contained in each message. In both slotted and unslotted versions of M_0-persistent CSMA, the performance under message switching is superior to that under packet switching in the sense that not only the channel capacity is larger but also the average number of retransmissions per successful message under message switching is smaller than that per successful packet under packet switching. In dynamic reservation schemes, message switching leads to larger channel capacity. A unique feature of dynamic reservation schemes is that priority scheduling can often be employed to give short messages higher priorities and reduce the overall transmission delay under message switching. However, in both slotted and unslotted versions of the ALOHA scheme, the channel capacity is reduced when message switching is used instead of packet switching. It is interesting to note that the lower bound of the channel capacity under message switching is almost inversely proportional to the squared coefficient of variation of message length. The reduction in channel capacity under message switching may also happen in the M_p-persistent CSMA scheme as p deviates from 0 to 1 for certain distributions of message length. Hence, the performance under message &itching via random access channels may be superior to or inferior to that under packet switching depending upon the random access scheme being used and the distributions of message length for certain random access schemes-- in this case, usually a large coefficient of variation of message length implies large degradation in channel capacity. Nevertheless, message switching can increase the channel cpaacity of radio channels if appropriate CSMA schemes have been chosen. If the terminal access networks of a store and forward computer communication network communicate via random access radio channels, a mixed strategy which uses message switching via radio channels and packet switching via terrestrial links of the network will lead to better performance when the radio channels are highly utilized.

REFERENCES

1. Kleinrock, L., "Queueing Systems, Vol. 1 Theory, Vol. 2 Computer Applications", New York, Wiley Interscience, 1975.

2. Kleinrock, L., "Communication Nets", Dover, 1972.

3. Kleinrock, L. and Lam, S.S., "Packet-switching in a Slotted Satellite Channel", AFIPS 1973, Vol. 42, pp. 703-710.

4. Roberts, L. G., 'Dynamic Allocation of Satellite Capacity Through Packet Reservations", AFIPS 1973, Vol. 42, pp. 711-716.

5. Roberts, L. G., "Data by Packet", IEEE Spectrum, Vol. 11, Feb. 1974, pp. 46-51.

6. Abramson, N., "The ALOHA System--Another Alternative for Computer Communications", AFIPS 1970, Vol. 37, pp. 281-285.

7. Abramson, N. and Kuo, F., "Computer Communication Networks", Prentice Hall, 1973.

8. Kobayashi, H. and Konheim, A. G., "Queueing Models for Computer Communication System Analysis", IBM Res. Rep. RC5922, 1976.

9. Kleinrock, L., "Analytic and Simulation Methods in Computer Network Design", AFIPS 1970, Vol. 36, pp. 569-579.

10. Abramson, N., "Packet Switching with Satellites", AFIPS 1973, Vol. 42, pp. 695-702.

11. Kleinrock, L. and Tobagi, F. A., "Packet Switching in Radio Channels: Part I--Carrier Sense Multiple Access Modes and Their Throughput Delay Characteristic", IEEE Trans. Commun., Corn-23, Dec. 1975, pp. 1400-1416.

12. Tobagi, F. A. and Kleinrock, L., "Pakcet Switching in Radio Channels: Part II--The Hidden Terminal Problem in Carrier Sense Multiple Access and the Busy Tone Solution', IEEE Trans. Commun., Com-23, Dec. 1975, pp. 1417-1433.

13. Tobagi, F. A. and Kleinrock, L., "Packet Switching in Radio Channels, Part III--The Polling and Dynamic Split-channel Reservation Multiple Access", IEEE Trans. Commun., Com.-24, Aug. 1976, pp. 832-844.

14. Lam, S. S. and Kleinrock, L., "Packet Switching in a Multiaccess Broadcast Channel: Dynamic Control Procedures", IEEE Trans. Commun., Corn.-23, Sept. 1975, pp. 891-904

15. Kleinrock, L. and Lam, S. S., "Packet Switching in a Multiaccess Broadcast Channel: Performance Evaluation", IEEE Trans. Commun., Corn.-23, April 1975, pp. 410-422.

16. Lam, S. S., 'Packet Switching in a Multi-access Broadcast Channel With Application to Satellite Communications in a Computer Network", Rep. UCLA-ENG 7429, 1974.

17. Lam, S. S., "Store and Forward Buffer Requirements in a Packet Switching Network', IEEE Trans. Commun., Corn.-24, April 1976.

18. Kleinrock, L., "Resource Allocation in Computer Systems and Computer Communication Networks", Inf. Proc. 74, North-Holland, 1974.

19. Data Channels, Vol. 1, No. 3 (Verona, New Jersey), Nov. 1973.

20. Pouzin, L., 'CIGALE, the Packet-switching Machine of the Cyclades Computer Network", Inf. Proc. 74, North-Holland, 1974.

21. Gelenbe, E. and Pujolle, G., "The Behavior of a Single Queue in a General Queueing Network", Acta Informatica 7, 1976.

22. Yu, P. s., "On Accuracy Improvement and Applicability Conditions of Diffusion Approximation With Applications to Modelling of Computer Systems", Stanford Univ. Tech. Rep. 129, 1977.

23. Coffman, E. G. and Denning, P. J., 'Operating Systems Theory", Prentice Hall, 1973.

24. Ferguson, M. J., "A Study of Unslotted ALOHA with Arbitrary Message Lengths", Data Comm., Quebec, Canada, Oct. 1975, pp. 5.20-5.25.

25. Ferguson, M. J., "An Approximate Analysis of Delay for Fixed and Variable Length Packets in an Unslotted ALOHA Channel', TR B76-6 University of Hawaii, 1976.